JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY

1/2011

(Contents Continued on Back Cover)

Editorial Board

Editorial Advisory Board

ISSN 1509-4553 on-line: ISSN 1899-8852

Warsaw 2011

Circulation: 300 copies

Sowa - Druk na życzenie, www.sowadruk.pl, tel. 22 431-81-40

JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY

Preface

Modern military operations are conducted in a dynamic environment usually with unanticipated partners and irregular adversaries. In order to act successfully, they need technical support that gives modularity, flexibility and security in connecting heterogeneous systems of cooperating allies. The efficiency of such operations strongly relies on communications and information systems' (CIS) ability to facilitate decision superiority, the state in which better-informed decisions are made and implemented faster than an adversary can react. A key challenge is to improve the situational awareness and reactivity, i.e., the capability to quickly benefit from new information that changes the operational situation. Usually, this depends on a complete reaction process, involving environmental data, data services, and decision makers, all of them interconnected. However, the broad use of information technology in modern command and control system makes its infrastructure the most valuable asset and the most vulnerable point of attack. Finding effective ways to protect and defend communications and information systems by ensuring their availability, integrity and confidentiality challenges even the most advanced technology.

Many research efforts aimed at elaboration and implementation of innovative communications and information technologies in military systems, and especially at delivering new information assurance and cyber defence capabilities have been undertaken world-wide. Selected results of such activities are presented in this issue of the *Journal of Telecommunications and Information Technology*. It contains 11 carefully selected papers that reflect the current state of the art in selected areas of communications and information technology development with application to the military domain. The papers cover a wide spectrum of questions relevant to information assurance (IA) and cyber defence (CD) provision, service oriented architecture (SOA) implementation in tactical domain as well as an effective use of wireless networks resources.

The first group of contributions consists of 6 papers related to information assurance and cyber defence. A discussion of digital signature scheme implementation in environments with space and bandwidth constraints is a subject of the paper *A New Short Signature Scheme with Random Oracle from Bilinear Pairings* by S. Akleylek *et al*. They propose a new and efficient short signature scheme constructed by bilinear inverse-square Diffie-Hellman problem that does not require any special hash function. The exact security proofs are also explained in the paper. The authors compare the results of the proposed solution with the BLS and ZSS signature schemes. The next paper, *Network Management in Non-classified Data Hiding System Using Master Resident over Hidden Layer* by K. Sawicki and Z. Piotrowski, depicts a practical implementation of a system that takes advantage of hidden data transmission during voice communication leading to information superiority over the adversary. The authors describe a mechanism of master resident, the transmission controller that allows the system's operator to use commands transmitted over hidden layer for remote control of protocol interpreter. The third paper, *Authentication in VoIP Telephony with Use of the Echo Hiding Method* by J. Rachoń, Z. Piotrowski and P. Gajewski, describes an implementation of echo hiding technique for VoIP subscriber identification. The authors present the results of experiments performed in testbedding environment that confirm the efficiency of the proposed solution. A. Flizikowski *et al.* in the paper *The INTERSECTION Framework: Applied Security for Heterogeneous Networks* present an example of security framework. The authors describe various ISO standards addressing telecommunication security management and intrusion detection architecture. They discuss the impact of known network threats on connected networks and propose anomaly detection techniques. The next paper, *Anomaly Detection Framework Based on Matching Pursuit for Network Security Enhancement* by R. Renk and W. Hołubowicz, present a novel framework for recognizing the anomalies in a network traffic based on correlation approach and propose new signal-based procedure for intrusion detection using matching pursuit algorithm. They combine and correlate parameters from different layers that allow detection of 0-day attacks and reduction of false positives. The effectiveness of the proposed approach has been proved in attack and anomaly detection scenarios. The final paper in this group, *Tunneling Activities Detection Using Machine Learning Techniques* by F. Allard *et al.*, describes a statistical analysis of ciphered flows that allows detection of the carried inner protocol. Regarding the deployed security policy, this technology could be added in security tools to detect forbidden protocols usages. In the defence domain, this technology could help preventing information leaks through side channels. The authors present a high-level tunnel detection tool architecture and discuss the results of experiments with a public database containing real data flows.

The next group composed of 3 papers is focused on vital aspects of service oriented architecture implementation in military domain. The first paper in this group, *Success Factors for SOA Implementation in Network Centric Environment* by J. Śliwa and M. Amanowicz, identifies 9 fundamental challenges for the SOA approach that make the benefit for the network enabled capability (NEC) programme and increase the effectiveness of military missions. The authors propose the quick wins solutions that can speed up the process of achieving network-enabled capability in heterogeneous multinational NEC environment. B. Jasiul *et al.* in the paper entitled *Authentication and Authorization of Users and Services in Dynamic Military SOA Environments* discuss the security requirements for a cross-domain information exchange in a federated environment. The authors propose an effective method of secure access to information resources based on web services. A special attention is paid to the authentication and authorization of users and services. The solution presented in the paper was examined in multinational experimentations and military exercises. The last paper in this group, *Web Services Efficiency in Disadvantaged Environment* contributed by J. Śliwa, T. Podlasek and M. Amanowicz presents the experimental results of web services (WS) provision techniques carried out in a test-bedding environment that emulates tactical disruptive network. The authors discuss the advantage of different WS adaptation techniques that allow minimizing the XML message size and JPEG image attachments. The presented results show the efficiency of considered methods that adapt the WS provision scheme to the network's constraints.

The last group of papers deals with the questions of effective use of resources in military wireless networks. T. Ginzler and M. Amanowicz in the paper entitled *Adaptation of the Kademila Routing for Tactical Networks* propose a modification of the widely used Kademlia peer-to-peer system to tactical environment. They show that optimizations in the routing may lead to faster lookups and extend the battery lifetime of mobile nodes, as well as increase the robustness of the network. The final paper in this issue, *Review of Distributed Beamforming* contributed by J. Uher, T. A. Wysocki and B. J. Wysocki, discusses the question of improving the range of communications and saving the precious battery power in wireless sensor networks by cooperative distributed beamforming. The authors present a review of current solutions focused on distributed beamformers implementations. The paper covers the calculation of ideal beamforming weights, practical considerations, such as carrier alignment or smart antennas based on distributed beamformers, and concludes with open research problems.

I would like to take this opportunity to express my thanks to the authors and reviewers for their efforts in the preparation of this issue of the *Journal of Telecommunications and Information Technology*. I trust that the Readers will find the papers dealing with the most recent research results in the area of military information and communications technology both useful and interesting.

Marek Amanowicz
Guest Editor

Paper

A New Short Signature Scheme with Random Oracle from Bilinear Pairings

Sedat Akleylek[a,b], Barış Bülent Kırlar[a,c], Ömer Sever[a], and Zaliha Yüce[a]

[a] *Institute of Applied Mathematics, Middle East Technical University, Ankara, Turkey*
[b] *Department of Computer Engineering, Ondokuz Mayıs University, Samsun, Turkey*
[c] *Department of Mathematics, Süleyman Demirel University, Isparta, Turkey*

Abstract—In this paper, we propose a new and efficient short signature scheme from the bilinear pairings. Our scheme is constructed by bilinear inverse-square Diffie-Hellman problem (BISDHP) and does not require any special hash function. The exact security proofs are also explained in the random Oracle model. We give the implementation and comparison results of our proposed signature scheme with the signature scheme proposed by Boneh, Lynn, Shacham (BLS) and Zhang, Safavi, Susilo (ZSS). Furthermore, we use this signature scheme to construct a ring signature scheme.

Keywords—bilinear Diffie-Hellman problem, bilinear pairings, ring signature, short signature.

1. Introduction

Digital signatures are one of the most important cryptographic primitive for the daily life. Short signatures are needed in environments with space and bandwidth constraints. Up to pairing-based cryptography, the best known shortest signature was obtained by using the digital signature algorithm (DSA) [1] over a finite field $\mathbb{F}_q$. The length of the signature is approximately $2\log q$. On the other hand, when the pairing-based cryptographic protocol is used the length of the signature is about $\rho \log r$, where $\rho = \log q/\log r$ and r is the largest prime divisor of the number of the points in the elliptic curve. For example, if one uses RSA signature 1024 bit modulus, the output of elliptic curve digital signature algorithm (ECDSA) is 320 bit long for the same security level. However, short signature provides the same security level only for 160 bits for the best choice.

In 2001 Boneh, Lynn and Shacham [2] proposed the idea of short signature scheme by using bilinear pairings. This scheme is based on Weil pairing and needs a special hash function [2], [3], [4]. Over the last years, there are various applications of bilinear pairings in short signature schemes to construct new efficient schemes [5], [6], [7]. The main improvement in short signature schemes is the use of cryptographic hash function such as MD5, SHA-1 [7] instead of special hash function called `MapToPoint` hash operation. It is known that short signature scheme with cryptographic hash function is more efficient than others since `MapToPoint` hash operation is still probabilistic.

In this study, we describe a new short signature scheme in a similar setting in ZSS scheme [7]. Our system is based on bilinear inverse-square Diffie-Hellman problem a combination of bilinear inverse Diffie-Hellman problem (BIDHP) and bilinear square Diffie-Hellman problem (BSDHP). The main advantage of our scheme is that it can be used with any cryptographic hash function such as MD5, SHA-1. To give the exact security proofs, we define a new problem called inverse square problem with k traitors ($k-$ISP). Then, the exact security proofs of proposed scheme are also explained in the random Oracle model. We give the comparison of our scheme with the BLS scheme and ZSS scheme. According to the comparison results, our scheme is more efficient than BLS scheme.

Furthermore, based on new proposed signature scheme, we construct a ring signature scheme.

This paper is organized as follows: Some preliminaries about bilinear pairings and some related problems to pairings are given in Section 2. Proposed short signature scheme and its security analysis are explained in Section 3. A construction of ring signature scheme is given in Section 4. We conclude in Section 5.

2. Pairing-Based Cryptography

In this section, we give some facts about bilinear pairings and define some new problems. The proposed short signature scheme uses bilinearity like others.

2.1. Bilinear Pairings

Definition 1: Let G_1 and G_2 be additive cyclic groups of order n. Let G_3 be a multiplicative cyclic group of order n. A bilinear pairing is an efficiently computable map $e: G_1 \times G_2 \longrightarrow G_3$ which satisfies the following additional properties:

1. (bilinearity) For all $P,R \in G_1$ and all $Q,S \in G_2$, we have $e(P+R,Q) = e(P,Q)e(R,Q)$ and $e(P,Q+S) = e(P,Q)e(P,S)$.

2. (non-degeneracy) For all $P \in G_1$, with $P \neq Id_{G_1}$, there is some $Q \in G_2$ such that $e(P,Q) \neq Id_{G_3}$. For all $Q \in G_2$, with $Q \neq Id_{G_2}$, there is some $P \in G_1$ such that $e(P,Q) \neq Id_{G_3}$. When $G_1 = G_2$ and n is prime, $e(P,P)$ is a generator of G_3 for all $P \neq Id_{G_1}$

The following lemma which is related to the properties of bilinear pairings can be easily verified.

Lemma 1: Let $e : G_1 \times G_2 \longrightarrow G_3$ be a bilinear pairing. Let $P \in G_1$ and $Q \in G_2$. Then

1. $e(P, Id_{G_2}) = e(Id_{G_1}, Q) = Id_{G_3}$
2. $e(-P,Q) = e(P,-Q) = e(P,Q)^{-1}$
3. $e(kP,Q) = e(P,kQ) = e(P,Q)^k$ for all $k \in \mathbb{Z}$.
4. $e(kP,lP) = e(P,P)^{kl}$ for all $k,l \in \mathbb{Z}$.

2.2. Some Problems

We consider the following problems in the additive group $(G,+)$ of order n.

- **Discrete logarithm problem (DLP):** For $P,Q \in G$, find $k \in \mathbb{Z}_n^*$ such that $Q = kP$ whenever such k exists.
- **Decisional Diffie-Hellman Problem (DDHP):** For $a,b,c \in \mathbb{Z}_n^*$, given P, aP, bP, cP decide whether $c \equiv ab \pmod{n}$.
- **Computational Diffie-Hellman problem (CDHP):** For $a,b \in \mathbb{Z}_n^*$, given P, aP, bP compute abP.

There are two variations of CDHP:

- **Inverse computational Diffie-Hellman problem (ICDHP):** For $a \in \mathbb{Z}_n^*$, given P, aP, compute $a^{-1}P$.
- **Square computational Diffie-Hellman problem (SCDHP):** For $a \in \mathbb{Z}_n^*$, given P, aP, compute a^2P.

The following theorem shows the relation of these problems that the proof can be found in [8].

Theorem 1: CDHP, ICDHP and SCDHP are polynomial time equivalent.
The security of some applications of bilinear pairings in cryptography relies on the difficulty of bilinear Diffie-Hellman problem (BDHP) which was first stated in [4].

Definition 2: Let G be a finite additive cyclic group of order n with a generator P, let e be a bilinear pairing on G, and let a,b,c be integers. The BDHP is to compute the value of the bilinear pairing $e(abcP,P)$, whenever aP, bP and cP are given.

The well known pairing-based protocols are three-party key exchange in one round protocol proposed by Joux in [9], identity-based encryption scheme by Boneh-Franklin in [4] and short signature scheme by Boneh-Lynn-Shacham in [2] that the security of them depends on the BDHP. There are variants of BDHP:

- **Bilinear inverse Diffie-Hellman problem (BIDHP):** For $a,b \in \mathbb{Z}_n^*$, given P, aP, bP to compute $e(P,P)^{a^{-1}b}$.
- **Bilinear square Diffie-Hellman problem (BSDHP):** For $a,b \in \mathbb{Z}_n^*$, given P, aP, bP to compute $e(P,P)^{a^2b}$.

It is not hard to obtain bilinear inverse-square Diffie-Hellman Problem as a combination of BIDHP and BSDHP:

- **Bilinear inverse-square Diffie-Hellman problem (BISDHP):** For $a,b \in \mathbb{Z}_n^*$, given P, aP, bP to compute $e(P,P)^{a^{-2}b}$.

Theorem 2: BDHP, BIDHP, BSDHP and BISDHP are polynomial time equivalent.

Proof:
BDHP $\Rightarrow$ BIDHP is trivial.
BIDHP $\Rightarrow$ BSDHP:
Given P, aP, bP, set the input of BIDHP as

$$Q = aP,\ Q_1 = P = a^{-1}Q,\ Q_2 = bP = ba^{-1}Q,$$

then BIDHP outputs

$$e(Q_1,Q_2) = e(Q,Q)^{(a^{-1})^{-1}ba^{-1}} = e(aP,aP)^b = e(P,P)^{a^2b}$$

BSDHP $\Rightarrow$ BISDHP:
Given P, a^2P, bP, set the input of BSDHP as

$$Q = a^2P,\ Q_1 = a^{-2}Q = P,\ Q_2 = a^{-2}bQ = bP,$$

then BSDHP outputs

$$e(Q_1,Q_2) = e(Q,Q)^{(a^{-2})^2ba^{-2}} = e(P,P)^{a^{-2}b}$$

BISDHP $\Rightarrow$ BDHP:
Given P, aP, bP, cP, set the input of BSDHP as the triples

$$(P, aP, cP),\ (P, bP, cP),\ (P, aP + bP, cP),$$

then we have $e(P,P)^{a^{-2}c}$, $e(P,P)^{b^{-2}c}$ and $e(P,P)^{(a+b)^{-2}c}$, respectively. Therefore, we obtain

$$e(P,P)^{abc} = \left(\frac{e(P,P)^{a^{-2}c} \cdot e(P,P)^{b^{-2}c}}{e(P,P)^{(a+b)^{-2}c}}\right)^{1/2}.$$

■

3. New Short Signature Scheme From Bilinear Pairings

In this section, we propose our signature scheme, and then explain its security. Moreover, we compare our scheme with BLS and ZSS schemes from the implementation point of view.

3.1. Signature Scheme

A signature scheme consists of four steps: a parameter generation algorithm `ParamGen`, a key generation algorithm `KeyGen`, a signature generation algorithm `Sign` and a signature verification algorithm `Verify`.
We describe the new signature scheme as follows:
Let $(G_1,+)$ and $(G_2,\cdot)$ be cyclic groups of prime order n, $P \in G_1$, $G_1 = <P>$ and $e : G_1 \times G_1 \rightarrow G_2$ be a bilinear map. Let $H : Z_2^{\infty} \rightarrow Z_2^{\lambda}$, where $160 \leq \lambda \leq \log(n)$ be a cryptographic hash function such as SHA1 or MD5. Suppose that $\mathcal{A}$ wants to send a signed message to $\mathcal{B}$.

- `ParamGen`: $\{G_1, G_2, e, n, P, H\}$
- `KeyGen`: $\mathcal{A}$ randomly selects $x \in \mathbb{Z}_n$ and computes $P_{pub1} = x^2P$ and $P_{pub2} = 2xP$. In this structure, P, P_{pub1} and P_{pub2} are the public keys, x is the secret key.
- `Sign`: Given a secret key x and a message m, $\mathcal{A}$ computes the signature, $s = (H(m)+x)^{-2}P$.
- `Verify`: Given the public keys P, P_{pub1} and P_{pub2}, a message m and a signature s, $\mathcal{B}$ verifies the signature if

$$e(H(m)^2P + P_{pub1} + P_{pub2}H(m), s) = e(P,P) \text{ holds.}$$

The verification is done by using bilinearity in the following equations:

$$e((H(m)+x)^2P, (H(m)+x)^{-2}P) =$$
$$e(P,P)^{(H(m)+x)^2(H(m)+x)^{-2}} = e(P,P).$$

3.2. Signature Security

The well-known attacks against signature schemes are without message attack and chosen-message attack. The strongest version of these attacks is an adaptive chosen-message attack. In this scenario, the attacker can ask the signer to sign any message that he/she chooses. He also knows the public key of the signer. Then, he can customize his queries according to the previous message and chosen signature pairs.

The strongest notion of security for signature schemes that is existentially unforgeable under adaptive chosen-message attack was defined by Goldwasser, Micali and Rivest [10]. Here, we use the definition of exact secure signature schemes by Bellare and Rogaway [11] stated as follows:

Definition 3: A signature scheme S, defined by $S = <\texttt{ParamGen}, \texttt{KeyGen}, \texttt{Sign}, \texttt{Verify}>$, is $(t, q_H, q_S, \varepsilon)$-existentially unforgeable under adaptive chosen-message attack if for every probabilistic polynomial time forger algorithm $\mathcal{F}$ running in t processing time, at most q_H queries to the hash oracle and q_S signatures queries, there does not exist a non-negligible probability ε.

A signature scheme S is $(t, q_H, q_S, \varepsilon)$-secure if there is no forger who $(t, q_H, q_S, \varepsilon)$ breaks the scheme.

We introduce a new problem that was called **k-ISP** (inverse square problem with k traitors) to give the security proof of the new signature scheme. This problem is similar to **k-CAA** (collusion attack algorithm with k traitors) that was proposed by Mitsunari, Sakai and Kasahara in [12].

Definition 4: (**k-ISP**) For an integer k, and $x \in \mathbb{Z}_n$, $P \in G_1$, given

$$\{P, xP, H_1, H_2, \cdots, H_k, (H_1+x)^{-2}P, (H_2+x)^{-2}P, \cdots, (H_k + x)^{-2}P\},$$

compute $(H+x)^{-2}P$ for some $H \notin \{H_1, H_2, \cdots, H_k\}$.

k-ISP is (t, ε)-hard if for any t-time adversaries $\mathcal{A}$, we have

$$Pr\left[\begin{array}{r} \mathcal{A}\Big(P, xP, H_1, H_2, \cdots, H_k, (H_1+x)^{-2}P, (H_2+x)^{-2}P, \cdots, \\ (H_k+x)^{-2}P) | x \in \mathbb{Z}_n, P \in G_1, H_1, H_2, \cdots, H_k \in \mathbb{Z}_n\Big) \\ = (H+x)^{-2}P, H \notin \{H_1, H_2, \cdots, H_k\} \end{array}\right] < \varepsilon$$

where ε is negligible.

The following theorem shows that proposed signature scheme is secure against the adaptive chosen-message attack.

Theorem 3: If there exists a $(t, q_H, q_S, \varepsilon)$-forger $\mathcal{F}$ using an adaptive chosen message attack for the signature scheme proposed in Section 3.1, then there exists a (t', ε')-algorithm $\mathcal{A}$ solving $q_S - ISP$, where $t' = t$ and $\varepsilon' \geq (\frac{q_S}{q_H})^{q_S} \cdot \varepsilon$.

Proof: Assume that the output of the hash function is uniformly distributed and the hash oracle will give a correct response for any hash query.

Suppose that a forger $\mathcal{F}$ $(t, q_H, q_S, \varepsilon)$-break the signature scheme using an adaptive chosen message attack. One needs an algorithm $\mathcal{A}$ to solve $q_s - ISP$. In this structure, the challenge is to compute $(H+x)^{-2}P$ for some $H \notin \{H_1, H_2, \cdots, H_k\}$ for given $P \in G_1$, $P_{pub1} = x^2P$, $P_{pub2} = 2xP$, $H_1, H_2, \cdots, H_{q_s} \in Z_n$ and $(H_1+x)^{-2}P$, $(H_2+x)^{-2}P, \cdots, (H_{q_s}+x)^{-2}P$

$\mathcal{A}$ is the signer and answers hash and signing queries by itself. Algorithm is as follows:

Step 1: $\{H_1, H_2, \cdots, H_{q_H}\}$ are the responses of the hash oracle queries for the corresponding messages $\{m_1, m_2, \cdots, m_{q_H}\}$.

Step 2: $\mathcal{F}$ makes a signature oracle query for each H_i for $1 \leq i \leq q_H$. If the hash oracle answers truely, $\mathcal{A}$ returns $(H_i+x)^{-2}P$ to $\mathcal{F}$ as the response. Otherwise, the process stops.

Step 3: $\mathcal{F}$ outputs a message-signature pair (m, S). The hash value of m is some H and $H \notin \{H_1, H_2, \cdots, H_{q_H}\}$. It satisfies:

$$e(x^2P + 2xP + H^2P, S) = e(P,P)$$

So, $S = (H+x)^{-2}P$. $\mathcal{A}$ outputs (H, S) as a solution of challenge.

Since the operations are the same for $\mathcal{A}$ and $\mathcal{F}$, the running time of $\mathcal{A}$ and $\mathcal{F}$ is equal, $t = t'$. The success probability of $\mathcal{A}$ is $\frac{q_S}{q_H}$ is Step 2. $\mathcal{A}$ will not fail with probability $p \geq (\frac{q_S}{q_H})^{q_S}$. Then, the success probability of the algorithm, $\mathcal{A}$ for all steps is $\varepsilon' \geq (\frac{q_S}{q_H})^{q_S} \cdot \varepsilon$. This completes the proof. ■

Note that, one can obtain a good bound if q_S and q_H are very closed.

We now recall k-weak computational Diffie-Hellman problem (**k-wCDHP**) proposed by Mitsunari *et. al* [12].

Definition 5: (**k-wCDHP**) For an integer k, and $x, H \in \mathbb{Z}_n$, $P \in G_1$, given $k+1$ values

$$\{P, (H+x)P, (H+x)^2P, \cdots, (H+x)^kP\},$$

compute $(H+x)^{-1}P$.

We define a new problem that is called $k+1$ inverse exponent problem (**k+1-IEP**) to give a specific evaluation of the security of our proposed signature scheme.

Definition 6: (**k+1-IEP**) For an integer k, and $a \in \mathbb{Z}_n$, $P \in G_1$, given $k+1$ values

$$\{P, aP, a^{-2}P, \cdots, a^{-k}P\},$$

compute $a^{-(k+1)}P$.

Theorem 4: **k-wCDHP** and **k+1-IEP** are polynomial time equivalent.

Proof:
k-wCDHP $\Rightarrow$ **k+1-IEP**:
Given $k+1$ values $P, (H+x)^{-1}P, (H+x)^{-2}P, \cdots, (H+x)^{-k}P$, let $Q = (H+x)^{-k}P$, $tQ = (H+x)^{-(k-1)}P$, and so $t = (H+x)$.
Set the input of **k-wCDHP** to be

$$(H+x)^{-k}P = Q,\ (H+x)^{-(k-1)}P = tQ,\\ (H+x)^{-(k-2)}P = t^2Q,\ \cdots,\\ (H+x)^{-1}P = t^{k-1}Q,\ P = t^kQ.$$

Then, **k-wCDHP** outputs

$$t^{-1}Q = (H+x)^{-1}(H+x)^{-k}P = (H+x)^{-(k+1)}.$$

k+1-IEP $\Rightarrow$ **k-wCDHP**:
Given $k+1$ values $P, (H+x)P, (H+x)^2P, \cdots, (H+x)^kP$, let $Q = (H+x)^kP$, $t^{-1}Q = (H+x)^{(k-1)}P$, and so $t = (H+x)$.
Set the input of **k+1-IEP** to be

$$(H+x)^kP = Q,\ (H+x)^{(k-1)}P = t^{-1}Q,\\ (H+x)^{(k-2)}P = t^{-2}Q,\ \cdots,\\ (H+x)P = t^{-(k-1)}Q,\ P = t^{-k}Q.$$

Then, **k+1-IEP** outputs

$$t^{-(k+1)}Q = (H+x)^{-1}P.$$

■

We note that **k+1-IEP** and **k-wCDHP** are no harder than the CDHP. There is a special case that **k+1-IEP** or **k-wCDHP** can be easily solved :
Given

$$P_0 = P,\ P_1 = (H+x)^{-1}P,\ P_2 = (H+x)^{-2}P,\ \cdots,\\ P_{(k-1)} = (H+x)^{-(k-1)}P,\ P_k = (H+x)^{-k}P,$$

if $P_i = P_j$ for $i \neq j$, this means that $(H+x)^{-i}P \equiv (H+x)^{-j}P \pmod{q}$, so the order of $(H+x)$ in $\mathbb{Z}_q$ is $j-i$. Then,

$$(H+x)^{-1}P = P_{j-i-1} \text{ or } (H+x)^{k+1}P = P_{k+1 \mod (j-i)}.$$

This case gives an attack on our proposed signature scheme. However, because of considering $(H+x)$ as a random element in $\mathbb{Z}_q^*$, we can show that the success probability of this attack is negligible.
Let q be a prime. Then, for any $a \in \mathbb{Z}_q^*$, the order of a, $ord(a)$, is a divisor of $q-1$. Given $k > 1$, assume that the number of element $a \in \mathbb{Z}_q^*$ such that $ord(a) \leq k$ is given by N. Since $\mathbb{Z}_q$ is a field, $N < k^2$ for $k > 1$. Let ρ be the probability that a randomly chosen element in $\mathbb{Z}_q^*$ has order less than k, then

$$\rho = \frac{N}{q} < \frac{k^2}{q}.$$

This gives us an opportunity to give a bound on k, such as, if $q \approx 2^{256}$, we limit $k \leq 2^{64}$, which means that the attacker has at most 2^{64} message-signature pairs. Therefore, using the above attack, the success probability is

$$\frac{(2^{64})^2}{2^{256}} = 2^{-128} \cdot 0.29387 \cdot 10^{-38}.$$

As a result, we have the following corollary.

Corollary 1: Assume that there is no polynomial time algorithm to solve the problem **k+1-IEP** with non-negligible probability, then the proposed signature scheme is secure under the random Oracle model.

3.3. Efficiency

We compare our signature scheme with the BLS scheme and ZSS scheme from the implementation point of view. PO, SM, PA, Squ, Inv, MTP and H denote the pairing operation, scalar multiplication in G_1, point addition in G_1, squaring in $\mathbb{Z}_n$, inversion in $\mathbb{Z}_n$, `MapToPoint` hash operation and hash operation in $\mathbb{Z}_n$, respectively. In the light of above, Table 1 summarizes the result.

Table 1
Comparison of our scheme with the BLS scheme and ZSS scheme

Scheme	BLS	ZSS	Proposed
Key generation	1 SM	1 SM	2 SM
Signing	1 MTP 1 SM	1 H 1 Inv 1 SM	1 H 1 Inv 1 SM 1 Squ
Verification	1 MTP 2 PO	1 H 1 SM 1 PO	1 H 1 SM 1 PO 2 PA 1 Squ

We implemented proposed signature scheme by using Pairing-Based Cryptography (PBC) Library [13] and The GNU Multiple Precision Arithmetic Library (GMP) [14]. Both libraries are installed as default installation. We run Cygwin as Linux simulator for GMP. The performance of signature schemas was measured on an Intel Core Duo

Table 2
Time comparison of our scheme with the BLS scheme and ZSS scheme

Scheme	BLS	ZSS	Proposed
All time including: key generation, signing, verification [s]	0.171000	0.098000	0.101000

1.6 GHz with 2 GB RAM, running Windows XP SP2. We have used standard functions of GMP 4.2.1/PBC 0.4.18 and compiled by GNU C Compiler. It should be noted that computation of pairing is the most time-consuming part in short signature schemes. According to the implementation result given in Table 2, our new scheme is more efficient than BLS scheme since it requires less pairing operation.

4. A Ring Signature Scheme

Ring signature schemes were proposed in [15]. Main purpose of a ring signature is to provide anonymity for the signer, by making it impossible to determine who among the possible signers is the actual one. By this way, the signature provides anonymity for the signer. Ring signature schemes satisfy signer ambiguity and security against an adaptive chosen message attack. A ring signature scheme is defined by:

- **ring signing** $(m, P_1, P_2, \cdots, P_r, x_i)$ produces a ring signature σ for the message m and a ring with r members, given the public keys $P_1, P_2, \cdots, P_r$ together with secret key of the signer x_i.
- **ring verifying** a signature pair (m, σ) includes the public keys of all the ring members i.e. possible signers.

The system parameters are $\{G_1, G_2, e, n, r, P, H\}$ which are defined in Section 3.1.

- **Sign:** Assume that the ith member of the ring sign the message. Let the public keys of the ring members be P_{pub1j} and P_{pub2j}, the secret key of the signer be x_i. Then,

$$\begin{aligned} S_i &= (H(m)+x_i)^{-2}P + (H(m)\sum_{j=1,i\neq j}^{r-1} 2x_jP \\ &+ \sum_{j=1,i\neq j}^{r-1} (x_j^2P + H(m)^2P)) \end{aligned}$$

- **Verify:**

$$\prod_{j=1}^{r} e((H(m)+x_j)^2P, S_i) = e(P,P).$$

Proof:

$$\begin{aligned} & \prod_{j=1}^{r} e((H(m)+x_j)^2P, S_i) \\ =\ & e(\sum_{j=1}^{r}(H(m)+x_j^2)P, S_i) \\ =\ & e(\sum_{j=1}^{r}(H(m)+x_j^2)P, (H(m)+x_i)^{-2}P \\ +\ & (H(m)\sum_{j=1,i\neq j}^{r-1} 2x_jP + \sum_{j=1,i\neq j}^{r-1}(x_j^2P + H(m)^2P)) \\ =\ & e(P,P). \end{aligned}$$

■

The security of the proposed ring signature scheme is similar as given in Section 3.2 since it is based on the signature scheme described in Section 3.1.

5. Conclusion

In this paper, we propose a new short signature scheme not requiring any special hash function. The security of this signature scheme depends on a new problem called bilinear inverse-square Diffie-Hellman problem (BISDHP). It is shown that this problem and BDHP are polynomial time equivalent. We also propose a new complexity assumption called the $k+1$ inverse exponent problem. The exact security proofs are also explained in the random Oracle model. We give the implementation and comparison results of our proposed signature scheme with the BLS and ZSS schemes. According to the implementation results, our new scheme is more efficient than BLS scheme since it requires less pairing operation. Finally, we construct a ring signature scheme based on our proposed scheme.

Acknowledgments

A preliminary version of this work was presented at Information Systems Technology Panel Symposium, 2010 (IST-091/RSY-021). We would like to thank Professor Ersan Akyıldız for his valuable comments. This research is partially supported by ASELSAN A.Ş. with the project 2008.15.00.2.00.02.

References

[1] *Digital Signature Standard*, FIPS PUB 186. National Institute of Standards and Technology, 1994.

[2] D. Boneh, B. Lynn and H. Shacham, "Short signatures from the Weil pairing", in *Advances in Cryptology – Asiacrypt 2001*, Lecture Notes in Computer Science, vol. 2248. Berlin: Springer, 2001, pp. 514–532.

[3] P. S. L. M. Barreto and H. Y. Kim, "Fast hashing onto elliptic curves over fields of characteristic 3", Cryptology ePrint Archive, Report 2001/098 [Online]. Available: http://eprint.iacr.org/2001/098/

[4] D. Boneh and M. Franklin, "Identity-based encryption from the Weil pairing", in *Advances in Cryptology – Crypto 2001*. Lecture Notes in Computer Science, vol. 2139, Berlin: Springer, 2001, pp. 213–229.

[5] D. Boneh and X. Boyen, "Short signatures without random Oracles", in *Advances in Cryptology – Eurocrypt 2004*, Lecture Notes in Computer Science, vol. 3027. Berlin: Springer, 2004, pp. 56–73.

[6] D. Boneh, X. Boyen and H. Shacham, "Short group signatures", in *Advances in Cryptology – Crypto 2004*, Lecture Notes in Computer Science, vol. 3152. Berlin: Springer, 2004, pp. 41–55.

[7] F. Zhang, R. Safavi-Naini and W. Susilo, "An efficient signature scheme from bilinear pairings and its applications", in *Advances in Cryptology – PKC 2004*, Lecture Notes in Computer Science, vol. 2947. Berlin: Springer, 2004, pp. 277–290.

[8] A. R. Sadeghi and M. Steiner, "Assumptions related to discrete logarithms: why subtleties make a real difference", in *Advances in Cryptology – Eurocrypt 2001*, Lecture Notes in Computer Science, vol. 2045. Berlin: Springer, 2001, pp. 243–260.

[9] A. Joux, "A one round protocol for tripartite Diffie-Hellman", in *Advances in Cryptology – ANTS 4*, Lecture Notes in Computer Science, vol. 1838. Berlin: Springer, 2000, pp. 385–394.

[10] S. Goldwasser, S. Micali and R. Rivest, “A digital signature scheme secure against adaptive chosen-message attacks”, *SIAM J. Comput.*, vol. 17, no. 2, pp. 281–308, 1988.

[11] M. Bellare and P. Rogaway, “The exact security of digital signatures – how to sign with RSA and Rabin”, in *Advances in Cryptology – Eurocrypt 1996*, Lecture Notes in Computer Science, vol. 1070. Berlin: Springer, 1996, pp. 399–416.

[12] S. Mitsunari, R. Sakai and M. Kasahara, “A new traitor tracing”, *IEICE Trans. Fundamentals*, vol. E85-A, no. 2, pp. 481–484, 2002.

[13] T*he Pairing-Based Cryptography (PBC) Library* [Online]. Available: http://crypto.stanford.edu/pbc/

[14] *The GNU Multiple Precision Arithmetic Library (GMP)* [Online]. Available: http://gmplib.org/

[15] R. L. Rivest, A. Shamir and Y. Tauman, “How to leak a secret”, in *Advances in Cryptology – Asiacrypt 2001*, Lecture Notes in Computer Science, vol. 2248. Berlin: Springer, 2001, pp. 552–565.

Sedat Akleylek received the B.Sc. degree in mathematics form Ege University, in Turkey, in 2004, the M.Sc. degree and the Ph.D. degree in cryptography both from Middle East Technical University (METU), in Turkey in 2008 and 2010, respectively. He has been with Institute of Applied Mathematics, METU and Department of Computer Engineering, Ondokuz Mayis University, Turkey since 2005. His research interest are in the areas of cryptography, algorithms and architectures for computations in Galois fields, computer algebra and e-learning.

e-mail: akleylek@metu.edu.tr
Institute of Applied Mathematics
Middle East Technical University
06531 Ankara, Turkey

Department of Computer Engineering
Ondokuz Mayıs University
55139 Samsun, Turkey

Barış Bülent Kırlar received his Ph.D. in cryptography in 2010 from the Institute of Applied Mathematics (IAM), Middle East Technical University (METU), Turkey. Since 2004 he has been with IAM, METU and Department of Mathematics, Süleyman Demirel University, Turkey where he is a research assistant. His research interests are number theory, finite fields and cryptography, with emphasis on elliptic curve cryptography.

e-mail: kirlar@metu.edu.tr
Institute of Applied Mathematics
Middle East Technical University
06531 Ankara, Turkey

Department of Mathematics
Süleyman Demirel University
32260 Isparta, Turkey

Ömer Sever received the B.S. degree in computer engineering in Middle East Technical University (METU), Turkey, and the M.Sc. degree in cryptography in METU, Turkey, in 2002 and 2007, respectively. Since 2002, he has been working in Turkish Navy as engineer officer. He is continuing Ph.D. in Cryptography in METU. His research interests are in the areas of computer security, cryptography and algorithms.

e-mail: severomer@yahoo.com
Institute of Applied Mathematics
Middle East Technical University
06531 Ankara, Turkey

Zaliha Yüce received the B.S. degree in computer engineering and the M.Sc. degree in cryptography in Middle East Technical University (METU), Turkey, in 2004 and 2007, respectively. Since 2008, she is working in STM A.Ţ and still Ph.D. student in Cryptography Department of METU. Her research interests are software implementations of pairing-based cryptography protocols.

e-mail: zyuce@stm.com.tr
Institute of Applied Mathematics
Middle East Technical University
06531 Ankara, Turkey

Software engineer
STM A.Ş, 06800 Ankara, Turkey

Paper

Network Management in Non-classified Data Hiding System Using Master Resident over Hidden Layer

Krzysztof Sawicki and Zbigniew Piotrowski

Military University of Technology, Warsaw, Poland

Abstract—The paper presents a practical implementation of the non-classified data hiding system (NDHS) understood as a military platform for information warfare that takes advantage of the hidden data transmission for voice connections in order to gain informational lead over a potential enemy. The NDHS performs here as a botnet network that is managed by the hidden transmission controller referred to as the master resident. Research studies are dedicated to investigation of various connections in heterogeneous links as well as functionalities of such components as hidden protocol bridges and the master resident.

Keywords*—*hidden protocol bridge, hidden protocol interpreter, master resident, non-classified data hiding system, steganography, voice over IP, watermarking.

1. Introduction

Telecom IT networks present an important component of the contemporary telecom world. Safeguarded circulation of information with its continuous supervision pose substantial challenges for contemporary science. From the military point any attack to the national IT network (along with the entire telecom infrastructure) is the first step to initial warfare actions. In case of public IT networks one has to be aware that such networks may be used as reserved communication means between military troops to substitute, when necessary, commonly used communication means operated regularly by the army. Total destruction of public IT networks is extremely difficult as these networks are really extensive and incorporate great number of protecting measures to preserve their integrity when one or several their components are disrupted.

Beside cryptographic methods the safeguarding measures use also techniques of information hiding that are complementary to cryptographic ones.

This paper deals with the mechanism of the master resident (MR), the possibility to remote control network devices using hidden layer and implementation of those concepts. As stated in [1] "*Network Management normally has 4 components: The component that supervises (...); The component that is supervised (...); A protocol that transfers the information between the agent and the server. (...); A list of possible things to manage*". In non-classified data hiding system (NDHS) the component that supervises is the master resident, components that are supervised are hidden protocol interpreters (HPI), simple protocol used to communicate between master resident and hidden protocol interpreters is transmitted over hidden layer and allows to send commands to HPIs thus NDHS is network management mechanism that is designed to work on hidden networks.

The master resident makes it possible to control heterogeneous IT networks using hidden and confidential transmission of commands and data that takes advantage of the digital watermark technology and encoding with use of the variably modified permutation composition (VMPC) key scheduling algorithm (KSA) [2]. That mechanism is a part of the NDHS [3], [4]. NDHS can be classified as the mechanism of electronic defence (ED) [5], [6] and uses information hiding techniques to transmit information between NDHS functional elements (hidden protocol interpreter and master resident) [3], [7] and allows the NDHS operator to remote control every hidden protocol interpreter using commands transmitted in hidden layer. In case of a hostile conflict such a mechanism, if widely applied in public networks, makes it possible to take control over selected components of public IT networks and then continuous operation of the overall telecom systems is conducive to gain an informational lead.

The remaining part of the paper is structured in the following way: the second part presents the historical solutions that have been developed so far related to the issues of making the information hidden whilst the third part describes the mechanism to be implemented. Finally, the fourth part outlines how the intended mechanism was implemented.

2. Related Works

Continuously more and more research efforts of scientific circles are targeted to the application opportunities of hidden transmissions via IT network. The past interest was focused on data transmission itself via individual segments of the network. Some examples of proposed solutions can be found in studies [8]–[11], where stress is put onto hidden transmission exclusively by one type of channels. Some attempts towards application of hidden transmission were undertaken in studies related to: the steganographic router [12] as well as to the system for hidden transmission of information [4]. There are also other solutions that can be used for authentication procedures in networks. For wireless networks to IEEE 802.11 such a mechanism is described in [13]. A similar mechanism is also applied to RF communication within the VHF bandwidth [14], [15].

The mechanism that is outlined in this paper combines various approaches to the issues how to apply hidden transmission to manage IT networks and was purposefully designed thus it would be operated in a heterogeneous environment and use the hidden layer to supervise the network infrastructure which is the unique feature.

3. Mechanism Description

Operation principle of the mechanism is based on use of voice transmission (phone conversation). Transmission of voice signals is carried out with use of various technologies: voice over IP (VoIP), phone calls within public switched telephone networks (PSTN) as well as connections with use of a military VHF radio stations.

The transmitted voice signal is considered as the transport layer for the binary signature of the watermark. Boundary components of the network (take-over points) that switch voice signals to subscribers can be furnished with suitable software HPI [3], [7] that make it possible to carry out actions on those components in remote manner. Remote commands with instructions to initiate required actions can be transmitted in a hidden manner by means of the voice signal.

The research studies employed three options of voice signals: male speech in English (track no. 1), male speech in German (track no. 2) as well as pop music (track no. 3). The watermarked records that had been processed with use of one of the two following methods: orthogonal frequency-division multiplexing (OFDM) [16] with its information capacity $P = 21$ bits as well as the method of drift correction modulation (DCM) [17] with two options of data payload $P = 84$ and $P = 147$ bits. Parameters of the records were the following: sampling frequency $f_S = 48000$ Hz, 16 bits per sample, mono mode.

3.1. Examination of Transmission Channels

In order to carry out the experiments it was necessary to examine individual transmission channels. The test consisted in transmission of a voice signal with an embedded digital watermark with further reception of the signal and recording it at the other end of the network segment. The recorded signal was then delivered to the watermark extractor where the binary signature of the watermark was obtained at the extractor output. In addition the binary signature bit error rate (BSBER – bit error rate of the binary signature embeded in watermark) was determined for the extracted binary signature of the watermark. The examination procedure covered the following types of transmission channels: RF channel of the NDHS-WiFi type where the voice signal was transmitted as VoIP phone call, PSTN telephone line as well as VHF radio channel. The aim of tests the performed tests was to find out whether a hidden transmission is feasible via such channels and to select optimum transmission parameters. Figure 1 presents results for measurements of the bit error rate for the digital signature of transmitted watermarked tracks for signals determined with use of the OFDM method, transmitted via RF channels of the NDHS WiFi type as well as via a PSTN channel. Figure 2 shows corresponding results for

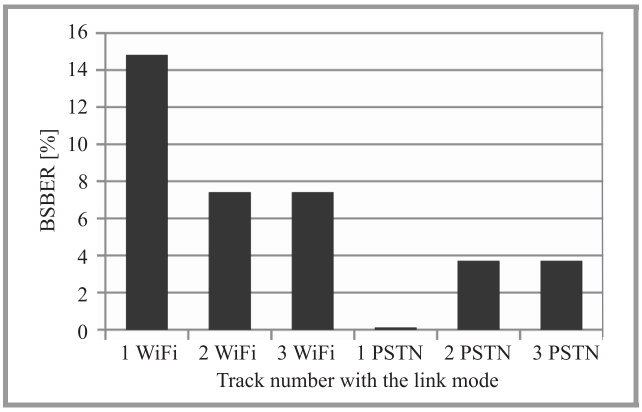


Fig. 1. Measurement results for the binary signature bit error rate when the OFDM method was used.

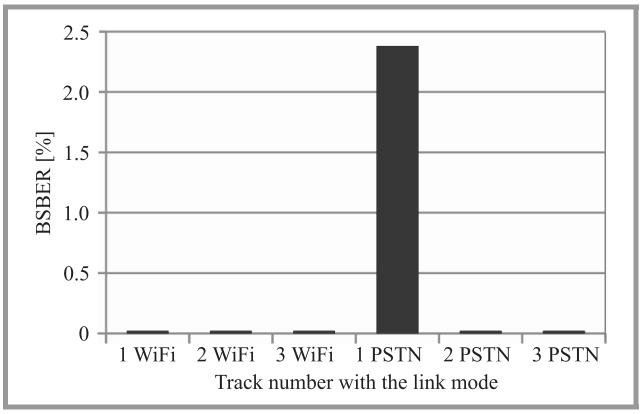


Fig. 2. Measurement results for the bit error rate when the DCM method was used ($P = 84$ bits).

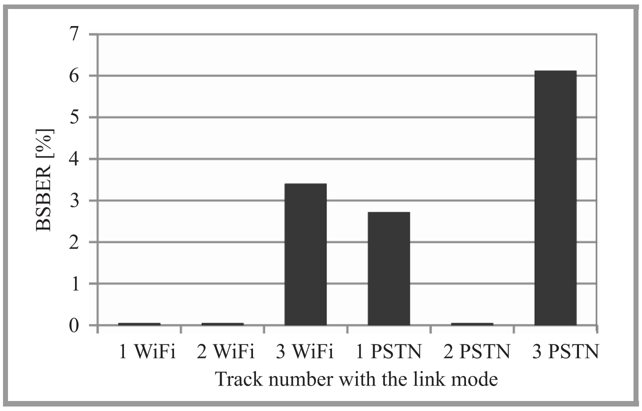


Fig. 3. Measurement results for the binary signature bit error rate when the DCM method was used ($P = 147$ bits).

signals where watermarks had been embedded with use of the DCM method and with the watermark data payload $P = 84$ bits. Similarly, Fig. 3 presents results for the DCM method and information capacity $P = 147$ bits. The analysis covered 30 seconds of a signal recorded at the other end of the examined transmission channel. The test demonstrated that the DCM method is less vulnerable to degradation of voice signals introduced by the IT networks under tests.

3.2. Examination of Bridges

For needs of the experiments the term of hidden protocol bridge (HPB) was introduced. The bridge is a hardware or software unit that is capable to handle the watermark embedded into a voice signal and to forward it between transmission channels of different types. Two bridge types were distinguished, i.e., the hardware bridge HPB-H that is incapable to process hidden information and merely forwards voice signals as well as the software bridge HPB-S that extracts binary signatures from voice signals transmitted via networks and uses corrective code extractors to recover information represented by the signatures. For that purpose two types of extractors can be used: the BCH type in case of the DCM method or the Reed-Solomon (RS) type when the OFDM method is used. Finally, the recovered information is recoded back into the watermark of the voice signal that is forwarded to that second network. In that way information to be forwarded in the hidden layer is refreshed. The completed experiments involved the following corrective codes: the BCH type ($n = 84$, $k = 28$, $t = 7$), where n – length of the code vector, k – length of the information vector and t – correction capacity of the code, another BCH type ($n = 147$, $k = 91$, $t = 7$) as well as the Reed-Solomon (RS) type ($n = 21$, $k = 9$, $t = 3$).

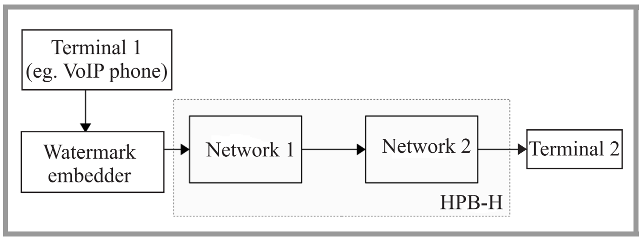


Fig. 4. General diagram of the hardware bridge HPB-H.

Figures 4 and 5 present general diagrams for individual bridge types. The following types of bridges were set up for the experiments: PSTN-NDHS WiFi as well as VHF-NDHS WiFi. Investigation of bridges included verification of their information bit error rates (IBER) for information recovered from the BCH and RS extractors. The calculated error rate served as the criterion to assess whether hidden transmission of commands to boundary appliances is possible or not.

Figure 6 presents measurement results for the bit error rate calculated for the hidden information (a control command) recovered from BCH and RS extractors for the hardware bridge PSTN-NDHS WiFi. One can clearly see that path no. 3 (pop music) is incapable to efficiently transmit control commands via such a bridge. In other cases corrective codes could recover the original form of information. Similar results were obtained for the software bridge (with information refreshment) of the PSTN-NDHS WiFi type. The obtained results can be seen in Fig. 7.

Experiment results for the software bridge VHF-NDHS WiFi are shown in Fig. 8. The graph contains only those paths that were substantially deteriorated so that extraction of the watermark binary signature proved infeasible. The next picture (Fig. 9) presents measurements results for the hardware implementation of the VHF-NDHS WiFi bridge. In such a case the OFDM method proved incapable to extract the watermark binary signature. On the other hand, the second path (male speech in German) with the watermark embedded by means of the DCM method proved sufficiently invulnerable to transmission degradation factors and recovery of the original command was possible.

3.3. Collaboration between Master Resident and Hidden Protocol Interpreter

The master resident is a specific component of the NDHS system as it is intended to supervise operation of individual stations that make up components of a hidden network (botnet) as it is the station where hidden protocol interpreters are installed. The hidden protocol interpreter enhances the forwarding station (router) with the functionality of a network-centric router [12]. The HPI unit analyzes streams of voice packages that pass through the specific router with the aim to detect hidden transmission or embeds a watermark into the forwarded voice signal. The master resident communicates with HPI units operated within the existing network in a hidden or open manner (depending on available possibilities and importance of information transmitted). The master resident makes the decision on the basis of information about data streams forwarded by every specific HPI unit and provides that HPI unit with relevant commands. The commands are always encrypted with use of the symmetric VMPC code. Encryption keys that are used to encode communication sessions between MR and HPI are stored in the MR database (repository) and are unique for each specific HPI unit. After receiving the command the HPI unit undertakes the required action. There are many possible actions, starting from simple recording of the voice signals through modification of these signals (injection of an additional content or extraction of some fragments of the voice package) up to disabling the entire voice signal or forwarding the latter to another subscriber. In extreme cases the entire components of the network infrastructure where the specific HPI is installed can be eliminated or some functions attributable to these components can be disabled.

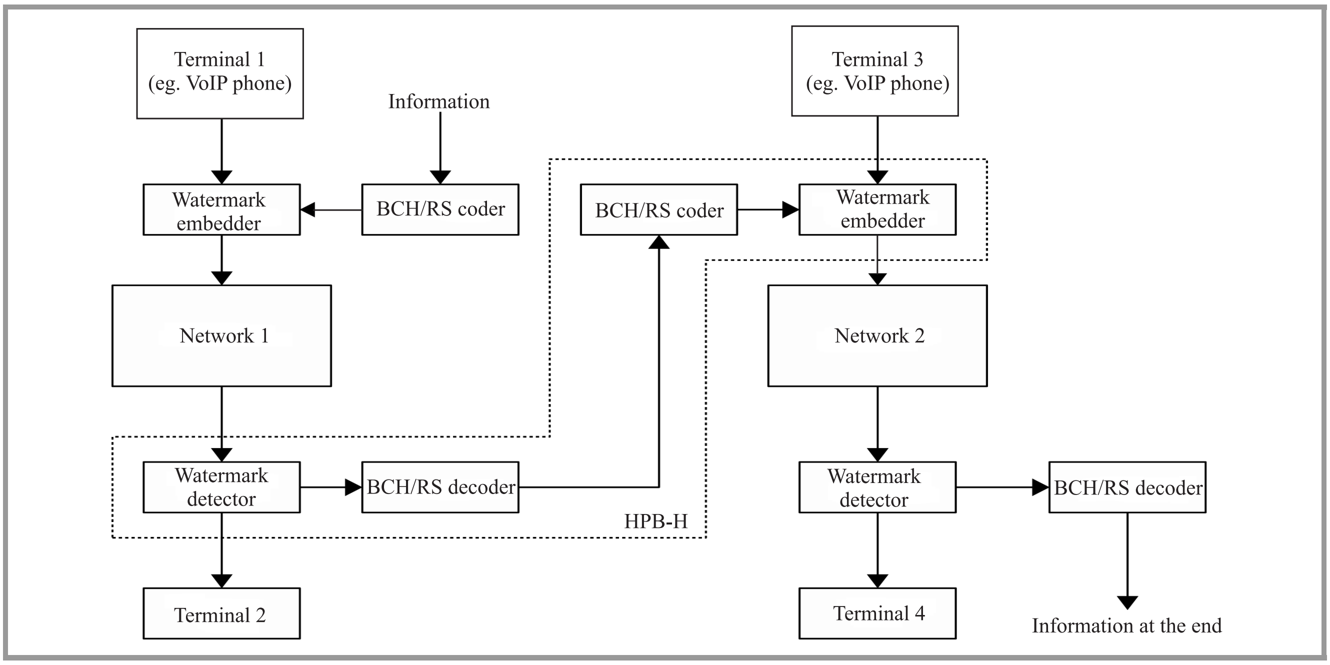


Fig. 5. General diagram of the software bridge HPB-S.

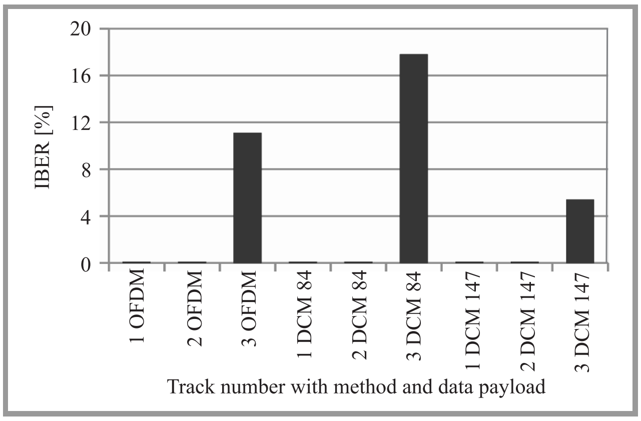


Fig. 6. Results for measurements of information bit error rates (IBER) for information recovered from the PSTN-NDHS WiFi bridge implemented as HPB-H.

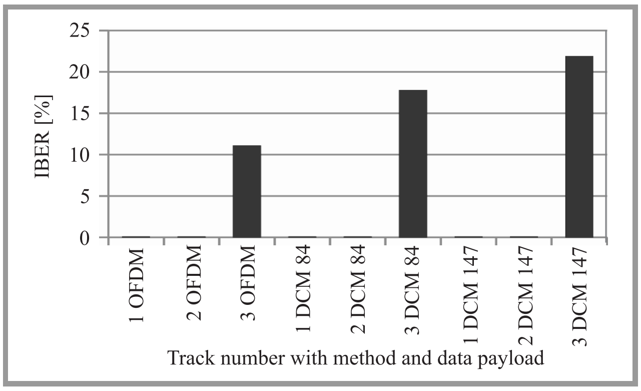


Fig. 7. Results for measurements of information bit error rates (IBER) for information recovered from the software bridge PSTN-NDHS WiFi.

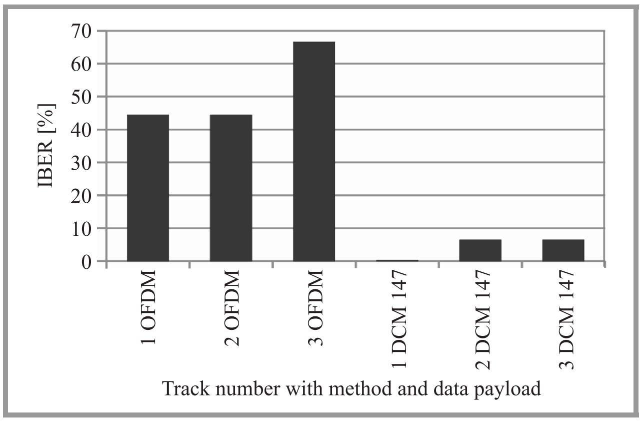


Fig. 8. Results for measurements of information bit error rates (IBER) for information recovered from the software bridge VHF-NDHS WiFi.

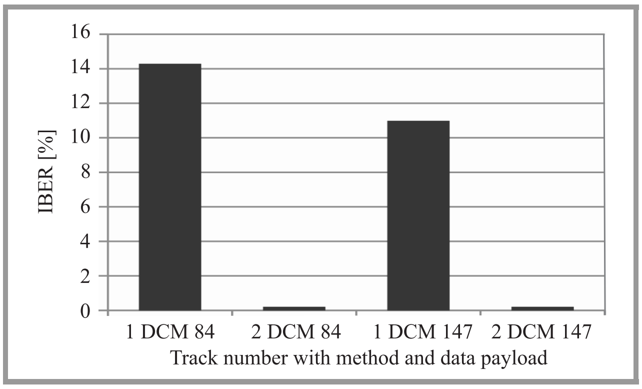


Fig. 9. Results for measurements of information bit error rates (IBER) for information recovered from the hardware bridge VHF-NDHS WiFi.

4. Implementation of the Mechanism

The described mechanism was implemented in practice on the basis of the GNU/Linux operating system. HPI units were developed as kernel modules whilst the MR represented an independent application developed in C programming language.

To confirm operability of the mechanism two hardware bridges were set up, namely VHF-NDHS WiFi as well as PSTN-VHF (in that way a heterogeneous network was developed) where two HPI were operated. The network layout with the established bridges is shown in Fig. 10.

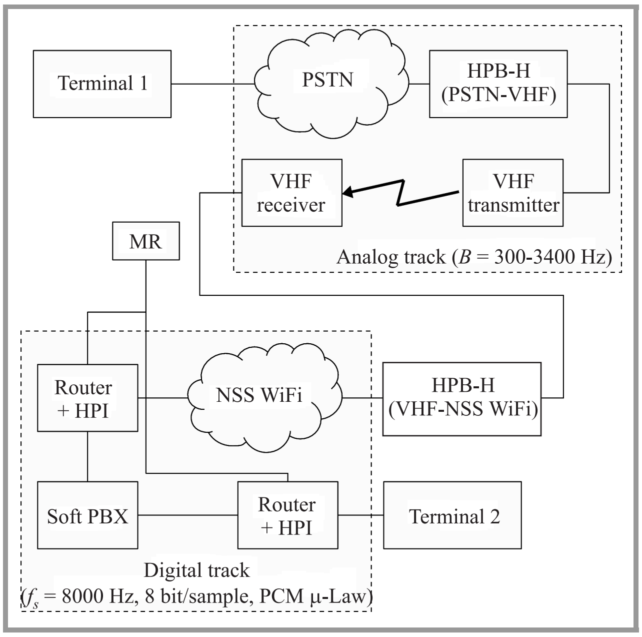


Fig. 10. Diagram of examined heterogeneous network (PSTN-VHF-NDHS WiFi).

The network was used to transmit voice packages with embedded watermarks equivalent to encrypted commands. The record no. 2 was selected for experiments (male speech in German) with the watermarks embedded with use of the DCM method (where the both information capacities were applied). The command was encrypted and then encoded with use of the BCH code. HPI units, after having the command decoded and decrypted, were forced to perform the specific operation, namely to have the voice message “conversation monitored, conversation monitored” introduced into the voice package transmitted between terminal 1 and terminal 2. The completed experiments confirmed the assumption that IT networks can be managed with use of hidden and confidential information transmitted via heterogeneous environment with use of HPI units, the master resident and HPB.

5. Conclusion

Nowadays none of the voice telephone systems (VoIP, PSTN, GSM) fails to be invulnerable to hidden transmission, which enables wide application of hidden transmission systems similar to the described one. Therefore there is the potential risk that the unauthorized hidden transmission shall be used to control components of network infrastructures. Consequently, development of mechanisms intended to safeguard the network against such an unintended use is an urgent must.

Acknowledgements

This paper has been co-financed from science funds granted within the years 2008–2010 as a research project of the Ministry of Science and Higher Education of the Republic of Poland no. 0018/B/T00/2008/34.

References

[1] http://sodaphish.com/files/ebks/try2innovate.com/downloads/E-books/Networking/Network%20Management.pdf

[2] B. Żółtak, “VMPC – One way function and stream cipher”, in *Proc. 11th Int. Worksh. Fast Software Encryption 2004 FSE’2004*, Delhi, India, 2004.

[3] Z. Piotrowski, “The national network-centric system and its components in the age of information warfare”, in *Safety and Security Engineering III*, M. Guarascio, C. A. Brebbia, And F. Garzia, Eds. Southampton, Boston, WIT Press 2009, pp. 301–309.

[4] Z. Piotrowski, “Effective method of the watermark embedding and decoding in the broadcast audio signal band”, Ph.D. thesis, Military University of Technology, Warsaw, Poland, 2005.

[5] *NATO Glossary of Abbreviations Used in NATO Documents and Publications*, NATO Standardization Agency, 2010.

[6] *NATO MC 0064/10 – NATO Electronic Warfare Policy*, NATO Electronic Warfare Advisory Committee NEWAC, 2010.

[7] K. Wodecki, “Hidden protocol interpreter and its main features in the watermarking battle net”, in *Proc. 2nd AFCEA Europe Student Conf.*, Brussels, Belgium, 2009.

[8] Z. Piotrowski, K. Sawicki, M. Bednarczyk, and P. Gajewski, “New hidden and secure data transmission method proposal for military IEEE 802.11 networks”, in *Proc. Sixth Int. Conf. Intel. Inf. Hid. – Multim. Sign. Proces. 2010 IIHMSP 2010*, Darmstadt, Germany (submitted for publication).

[9] K. Szczypiorski, “HICCUPS: Hidden communication system for corrupted networks”, in *Proc. 10th Int. Multi-Conf. Adv. Comp. Sys. ACS 2003*, Międzyzdroje, Poland, 2003.

[10] W. Mazurczyk and K. Szczypiorski, “Steganography in handling oversized IP packets”, in *Proc. Int. Conf. Multime. Inf. Netwo. Sec. MINES 2009*, Wuhan, Hubei, China, 2009, vol. I, pp. 559–564.

[11] L. Frikha, Z. Trabelsi, and W. El-Hajj, “Implementation of a covert channel in the 802.11 header”, in *Proc. Int. Wir. Commun. Mob. Comput. Conf. IWCMC 2008*, Crete Island, Greece, 2008, pp. 594–599.

[12] K. Szczypiorski, I. Margasiński, and W. Mazurczyk, “Steganographic routing in multi agent system environment”, *J. Inf. Assur. Sec.*, vol. 2, iss. 3, pp. 153–154, 2007.

[13] T. E. Calhoun, R. Newman, and R. Beyah, *Authentication in 802.11 LANs Using a Covert Side Channel*, Atlanta, Georgia State University, 2009.

[14] Z. Piotrowski and P. Gajewski, “Novel method for watermarking system operating on the HF and VHF radio links”, in *Computational Methods and Experimental Measurements XIII*, C. A. Brebbia and G. M. Carlomagno, Eds. Southampton, Boston, WIT Press 2007, pp. 791–800.

[15] Z. Piotrowski, L. Zagoździński, P. Gajewski, and L. Nowosielski, “Handset with hidden authorization function”, in *Proc. Eur. DSP Educ. Res. Symp. EDERS 2008*, Tel Aviv, Israel, 2008, pp. 201–205.

[16] P. Gajewski, J. Łopatka and Z. Piotrowski, “A new method of frequency offset correction using coherent averaging”, *J. Telecommun. Inf. Technol.*, vol. 1, pp. 142–146, 2005.

[17] Z. Piotrowski, “Drift correction modulation scheme for digital audio watermarking”, in *Proc. Int. Conf. Multime. Inf. Netw. Secur. MINES 2010*, Nanjing, China (submitted for publication).

Krzysztof Sawicki received his M.Sc. degree in communications from the Military University of Technology (MUT Warsaw) in 2009. Currently he is a Ph.D. student at the Electronics Faculty at MUT. He is interested in information hiding technology especially wireless networks steganography.

e-mail: Krzysztof.Sawicki@wat.edu.pl
Military University of Technology
Kaliskiego st 2
01-489 Warsaw, Poland

Zbigniew Piotrowski received the M.Sc. and Ph.D. degrees in communications from the Military University of Technology (MUT), Warsaw, in 1996, and 2005 (with honours), respectively. At present he is a DSP engineer in the Telecommunication Institute (EF MUT). His main areas of interest are speech and audio processing, telecommunication systems engineering and information hiding technology.

e-mail: Zbigniew.Piotrowski@wat.edu.pl
Military University of Technology
Kaliskiego st 2
01-489 Warsaw, Poland

Paper

Authentication in VoIP Telephony with Use of the Echo Hiding Method

Jakub Rachoń, Zbigniew Piotrowski, and Piotr Gajewski

Military University of Technology, Warsaw, Poland

Abstract—The paper describes the method intended to authenticate identity of a VoIP subscriber with use of the data hiding technique that is specifically implemented by means of the echo hiding method. The scope includes presentation of experimental results related to transmission of information via a hidden channel with use of the SIP/SDP signalling protocol as well as results of subjective assessment on quality of a signal with an embedded watermark.

Keywords—authentication, BS 1116-1 test, digital watermarking, echo hiding.

1. Introduction

Nowadays, when the VoIP technology is being rapidly developed, the problem of subscriber authentication is appearing as the more and more important problem. The existing threats to safety of telephone connections [1], [2] have led to the need to seek for alternative methods robust against intentional attacks. Encryption of dedicated phone calls is only a partial solution of the problem as most of PSTN calls that are currently made are incapable to cope with confidentiality of connections. There appeared an attempt to resolve the problem with use of the electronic appliance called personal trusted terminal (PTT) [3] that bases on an objective (numerical) verification of radio subscribers and that is also applicable to VoIP, GSM and PSTN networks. The concept associated with objective verification of telephone subscribers on telecom lines with use of the information hiding technique can be also applied to hidden authentication of subscribers for telephone and radio links [4]. The present study deals with the method of watermarking by means of the echo hiding technique. Alongside, results of studies on robustness of the watermark to variable conditions attributable to wide area networks (WAN) are presented. In addition, the emulation method for WANs link path conditions is also described, where the emulation is carried out with use of the VMware Player software and the Debian operating system that is derived from the Linux kernel. The mentioned software is free of charge and commonly available from Internet.

2. Echo Hiding Method

2.1. Embedder Design

The echo hiding method consists in filtering of the original signal where one of two filters with the finite impulse response is used. The two filters differ with the following parameters: response delay, rate of the response fading as well as number of delays that produce the echo (kernels). The specific filter is selected pursuant to the bit value of the transmitted watermark. The encryption method with use of the echo hiding technique is described with more details in [5] as well as in [6] and [7]. The applied algorithm implements the module of the detector that is capable to recognize sounding vowels and consonants, which makes it possible to incorporate the watermark to selected frames of the acoustic signal. For that purpose the average magnitude difference function (AMDF) is used. It is the method that for the first time was explained in [8]. Consequently, when information on structure of the applied detector of sounding consonants and vowels is unavailable it is infeasible to correctly detect the watermark on the receiver side.

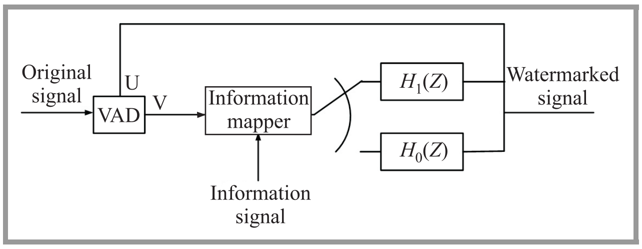


Fig. 1. Design of the watermark embedder.

In that way transmission of the watermark becomes hidden and adopts features of a confidential transmission as it is the case when the process of trials and errors is suitable to detect only selected fragments of the transmitted signal. Taking account for the fact that the information represented by the watermark is used only once for the specific phone connection session, it becomes much more difficult to fake identity of the subscriber. Design of the watermark embedder is shown in Fig. 1.

2.2. Design of the Watermark Embedder

Operation principle of the embedder consists in computation of the auto-cepstrum function intended to detect delay of the echo. It is the method that was described for the first time in [9]. The algorithms takes advantage of a voice activity detector (VAD) for sounding vowels and consonants as it is reasonable to find out only those fragments (frames) of the signal, where the watermark can be embedded. The same VAD is also used on the receiver side. The watermark extractor demands for more computation power than the embedder due to the reason that for each signal frame it must find out the auto-cepstrum function that is defined by the following equation:

$$f_d = \left\{ \mathrm{IFFT} \left[\log \left(\mathrm{FFT} \left(u_w \left(n \right) \right) \right)^2 \right] \right\}^2 , \qquad (1)$$

where:
u_w – amplitude of subsequent samples within the frame,
FFT – fast Fourier transform,
IFFT – inversed fast Fourier transform

2.3. The Real-Time Mode Process of Watermark Embedding and Extraction

Owing to the fact that the algorithm for watermark embedding is relatively uncomplicated, it is feasible to implement it in the real-time mode to system platforms with low computation capacities. It makes possible to use the algorithms in such portable devices as PDA, mobile phones or other appliances, where, e.g., the Java virtual machine is installed. Due to more demanding requirements of the extractor, the stream of received bits must split into two paths, where the first part is forwarded to the D/A converter and then delivered to the loud-speaker (handset), whereas the second part arrives to the watermark extractor. Therefore, it is possible to embed and extract watermarks with no compromise to the voice quality.

3. Measurement Test Bed

In order to measure robustness of the watermark to variable conditions typical for WANs links, two virtual machines created within the VMware Player [9] were used. VMware Player is the software application that makes it possible to assign a part of hardware computer resources to establish an isolated architecture that enables to run any operating system. Furthermore, the virtual machines are capable to communicate by means of the IP protocol. To emulate WAN environment one machine called router was used to launch the packets forwarding service (IP forwarding) and then the tool called traffic control was applied to manage traffic of outgoing packets (Fig. 2). The router forwards packets to the virtual telephone exchanger PBX Asterisk. The traffic control tool allows emulation of the following parameters:

- packet delay,
- packet loss,
- packet repeating,
- packet damage by random bit swapping,
- packet reordering.

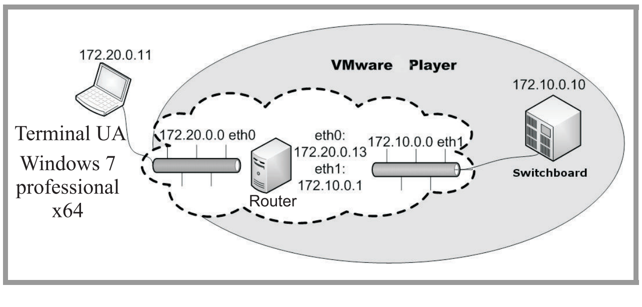


Fig. 2. WAN emulation.

The foregoing configuration of the test bed made it possible to carry out the following experiment. The stream of speech signal was redirected to the subscriber's voice mailbox, whereas the stream of RTP packets was controlled by the router, where traffic of outgoing packets was delayed in accordance with the Gaussian distribution with the constant mean value of 100 ms and increasing standard deviation across the *jitter* experiment. The acoustic signal transmitted with use of the RTP protocol was encoded by means of the G.711 μ-Law codec. The experiments were carried out in the following way:

1. The binary signature of the watermark was embedded by means of the echo hiding algorithm with use of the Matlab environment.
2. The WAVE file containing the examined soundtrack was reproduced as a sound source for the SIPCLI software. It is the software tool that is used to establish connections when the SIP protocol is applied.
3. The recorded voice message stored on the voice mailbox was transferred to the local disk.
4. The file with the recorded message was decoded by the watermark extractor within the Matlab environment.

The connection between the client and the machine, where the VMware Player environment was launched, was established with use of the Ethernet cable UTP cat. 5 with the length of 1.2 m. The connection was handled by network cards operating according to the IEEE 802.3u standard (100Base-TX Fast Ethernet).

4. Measurement Results

Figure 3 presents comparison between quality of the watermark extraction depending on the type of the finite impulse response (FIR) filter applied to embed the echo. The experiments were carried out within a closed loop for the male English speech. The signal was sampled with the frequency of 8 kHz and 16-bit resolution. The d0 parameter stands for the echo delay expressed as a number of samples for the bit with the low logic level (0), whereas the d1 parameter is meant for the echo delay expressed as a number of samples for the bit with the high logic level (1). The echo fading factor is 0.4 for the both cases. The values

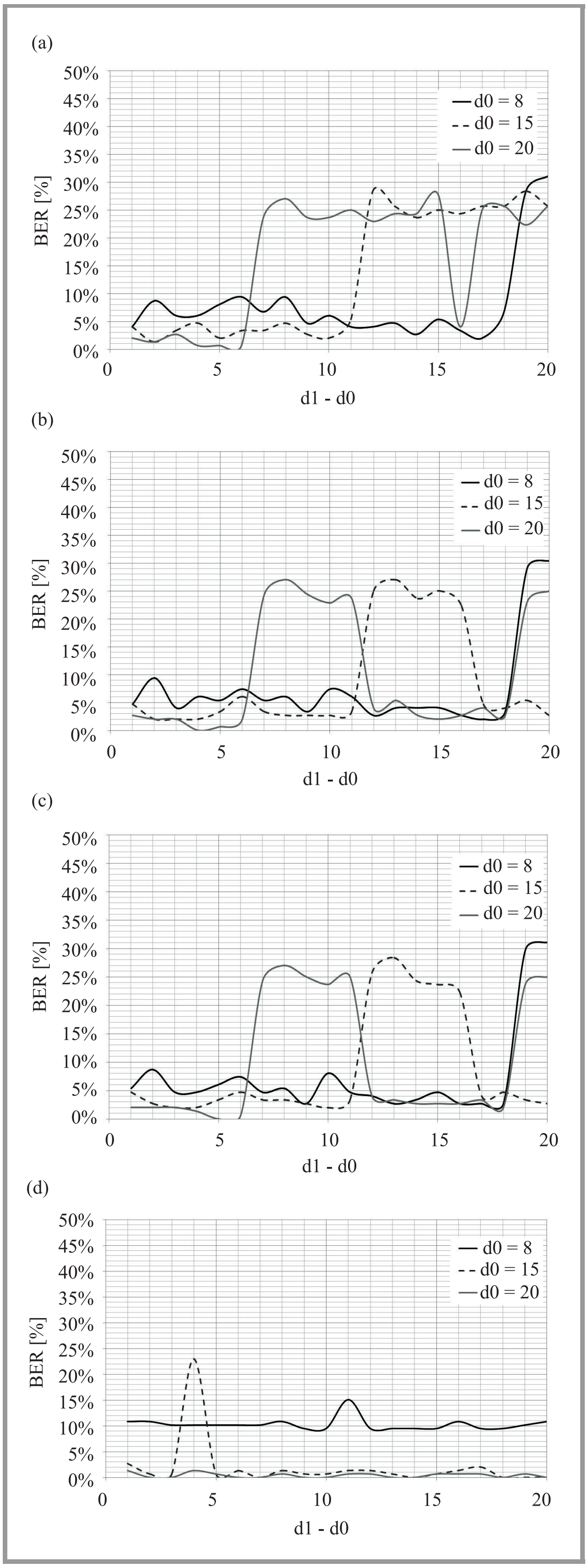


Fig. 3. Efficiency of detection as a function of echo delay and number of echo kernels: (a) = 1, (b) = 3, (c) = 5 for the English speech signal and for signature with 32 [b] payload (d) bilateral echo methode.

of d0 were selected on the basis of the experimental study described in [10]. The presented graphs serve as the proof that any increase in number of echo kernels is not enough to substantially improve the detection efficiency. Only the filter with the preceding or delayed kernel (also known as bilateral kernel) significantly improves extraction quality. On the other hand, the watermark quality estiamtion tests demonstrate that such a solution considerably deteriorates the original signal and leads to distortions that make the watermark easily hearable. To reach a compromise between satisfying results of the watermark extraction and the watermark inaudibility, the following parameters of the watermark embedder were selected for the male English speech:

- d0 = 15 samples,
- d1–d0 = 5 samples,
- number of echo kernels = 1,
- echo fading factor = 0,4.

To guarantee correct detection of the signature secured by means of the error detecting and correcting code BCH, the elementary error rate must be below 10%. As one can see in Fig. 4 detection/extraction of the watermark is feasible when the standard deviation of packet delays ranges within the interval of 9 ms. It imposes the requirement to guarantee relative steady parameters of the line that was established for the needs of the voice connection to the SIP protocol. The average standard deviation for the packet delays is equivalent to the average value of the parameter that is referred to as *jitter*. For digital transmissions such as transmission with use of the RTP protocol the watermark is correctly extracted at the receiver side, even in case of conversion from the PCM format 16 bits per a sample to the G711 μ-Law 8 bits per a sample and the reverse conversion to the PCM format 16 bits/sample at the side of the PBX Asterisk switchboard. The approximate transmission watermark data payload is 7 bit/s. The following paragraph comprises statistical information on parameters of the RTP channel. The parameters have been determined with use of the RTP stream analysis software application incorporated

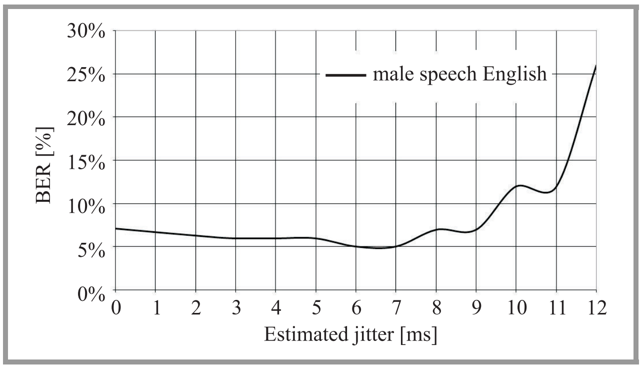


Fig. 4. Robustness of the watermark against jitter in the RTP channel, $fs = 8$ kHz

into the Wireshark package on the basis of data packets captured at the side of the PBX telephone switchboard.

```
Max delta = 66,16 ms at packet no. 1746
Max jitter = 15,23 ms.  Mean jitter = 10,19 ms.
Total RTP packets = 1150 (expected 1150)
Lost RTP packets = 0 (0,00%)
Sequence errors = 120
Duration 23 s
(-386 ms clock drift,
corresponding to 7868 Hz (-1,65%)
```

The above statistics have been found out for the connection with the emulated standard deviation of 9 ms. It turns out that for such circumstances the drift of phase angle amounting to –1,65% occurs. Therefore, the proposed method is also insensitive to the effect of phase angle drift, which is very common on telecom links.

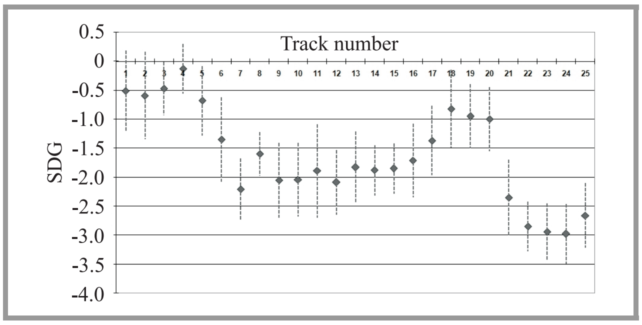


Fig. 5. Subjective fidelity assessment for the signal with the embedded watermark. The results are provided for the confidence interval of 95% and were determined with use of the ITU-R BS 1116-1 test. Details about sound tracks used in this test can be found in Table 1.

To subjectively assess fidelity of signals with embedded watermarks the test defined by the standard ITU R BS.1116-1 was carried out. The test was completed with participation of 10 listeners. Test results were subject to statistic computations with the confidence interval of 95% and then shown in Fig. 5. The SDG values above –1 mean that the watermark is not hearable. The obtained results enable to come out with the following conclusions:

- Authentication of SIP subscribers with use of watermarks is possible for networks where QoS is applied.
- Watermark signal is robust against variable delays of packets transmitted within the network.
- Increase both: echo signal numbers, as well as the echo scaling factor, makes the watermark better hearable on the background of the original signal.
- The subjective assessment how much embedded watermarks distort the original speech signal may vary with the language of conversation and gender of speakers.

The experimental results serve as the confirmation that authentication of subscribers who use VoIP telephony is possible when watermarks are embedded by means of the echo method and the original signal at the receiver side may remains unknown. It was also demonstrated that the proposed method is robust against conversion with use of the G.711 μ-Law codec. In addition, the experiments provided the proof that the method is robust against the phase angle drift (signal jitter) that commonly occurs in telecom channels. Therefore, the 'blind' extraction of watermarks is possible, i.e., the original signal at the receiver side is not necessary to correctly extract the binary signature that is represented by the embedded watermark.

Table 1
Description of sound tracks that were used to assess quality of the watermarking technique

Track number	Type of speech signal	Echo decay factor	Number of echo signals	BER [%]
1	French female	0.4	1	6.56
2	English female	0.4	1	5.71
3	English male	0.4	1	0.00
4	German male	0.4	1	5.56
5	English male	0.4	1	5.00
6	French female	0.8	1	1.64
7	English female	0.8	1	0.00
8	English male	0.8	1	0.00
9	German male	0.8	1	0.00
10	English male	0.8	1	1.67
11	French female	0.6	2	0.00
12	English female	0.6	2	0.00
13	English male	0.6	2	1.35
14	German male	0.6	2	2.78
15	English male	0.6	2	0.00
16	French female	0.4	3	4.92
17	English female	0.4	3	4.29
18	English male	0.4	3	1.35
19	German male	0.4	3	5.56
20	English male	0.4	3	6.67
21	French female	0.4	bilateral echo	3.33
22	English female	0.4	bilateral echo	20.29
23	English male	0.4	bilateral echo	1.35
24	German male	0.4	bilateral echo	1.39
25	English male	0.4	bilateral echo	0.00

5. Recommendations

The studies on implementation of watermark embedding with use of the echo hiding method serve as the evidence that the RTP channels set up for connections to both the H.323 and SIP protocols are suitable for transmission with watermarks embedded into voice signals.

References

[1] Z. Piotrowski and P. Gajewski, "Voice spoofing as an impersonation attack and the way of protection", *J. Inf. Assur. Secur.*, vol. 2, iss. 3, pp. 223–225, 2007.

[2] C. Roberts, "Voice over IP security", Center for Critical Infrastructure Protection, Wellington, New Zealand, March 2005.

[3] Z. Piotrowski, L. Zagoździński, P. Gajewski, and L. Nowosielski, "Handset with hidden authorization function", in *Proc. Eur. DSP Educ. Res. Symp. EDERS 2008*, Texas Instruments, pp. 201–205, 2008.

[4] Z. Piotrowski and P. Gajewski, "Novel method for watermarking system operating on the HF and VHF radio links", in *Computational Methods and Experimental Measurements XIII, CMEM XIII*, C. A. Brebbia and G. M. Carlomagnowit, Eds. Southampton, Boston: Wit Press, 2007, pp. 791–800.

[5] D. Gruhl, A. Lu, and W. Bender, "Echo hidding", Massachusetts Institute of Technology Media Laboratory, 1996, pp. 295–311.

[6] S. A. Chou and S. F. Hsieh, "An echo-hiding watermarking technique based on bilateral symmetric time spread kernel", in *Proc. IEEE ICASP 2006*, Toulouse, France, 2006.

[7] H. J. Kim "Audio watermarking techniques", in *Proc. Pacific Rim Workshop on Digital Steganography*, Kitakyushu, Japan, 2003.

[8] M. J. Ross, H. L. Shaffer, A. Cohen, R. Freudberg and H. J. Manley, "Average magnitude difference function pitch extractor", *IEEE Trans. Acoust., Speech and Sig. Proces.*, vol. 22, no. 5, pp. 353–362, 1974.

[9] B. P. Bogert, M. J. R. Healy, and J. W. Tukey, "The quefrency alanysis of time series for echoes: cepstrum, pseudo autocovariance, cross-cepstrum and saphe cracking", in *Proc. Symp. Time Series Anal.*, M. Rosenblatt, Ed., Chapter 15. New York: Wiley, 1963, pp. 209–243.

[10] Li Li, Ya-Qi Song, "Experimental research on parameter selection of echo hiding in voice", in *Proc. 8th Int. Conf. Machine Learn. Cybernet. ICMLC 2009* , Baoding, China, 2009, p. 2423–2426.

Jakub Rachoń received the M.Sc. in telecommunication systems from the Military University of Technology (MUT), Warsaw, in 2010. He spent one year at Ghent University in Belgium as an exchange student at faculty of engineering. His main area of interest are IT security, digital signal processing, mobile applications designing, intellectual property management and technology transfer processes.

e-mail: jakub.rachon@gmail.com
Military University of Technology
Kaliskiego st 2
01-489 Warsaw, Poland

Piotr Z. Gajewski received the M.Sc., and D.Sc. degrees from Military University of Technology (MUT) Warsaw, Poland in 1970, and 2001, respectively, both in telecommunication engineering. Since 1970 he has been working at Electronic Faculty of Military University of Technology (EF MUT) as a scientist and lecturer in communications systems (radios, cellular, microcellular), signal processing, adaptive techniques in communication and communications and information systems interoperability. He was an Associate Professor at Telecommunication System Institute of EF MUT from 1980 to 1990. From 1990 to 1993 he was Deputy Dean of EF MUT. Currently he is the Director of Telecommunication Institute of EF MUT. He is an author (co-author) of over 80 journal publications and conference papers as well as four monographs. He is a member of the IEEE Vehicular Technology and Communications Societies. He is also a founder member of the Polish Chapter of Armed Forces Communications and Electronics Association.

e-mail: Piotr.Gajewski@wat.edu.pl
Military University of Technology
Kaliskiego st 2
01-489 Warsaw, Poland

Zbigniew Piotrowski – for biography, see this issue, p. 16.

Paper

The INTERSECTION Framework: Applied Security for Heterogeneous Networks

Adam Flizikowski[a], Mateusz Majewski[b], Maria Hołubowicz[b], Zbigniew Kowalczyk[c], and Simon Pietro Romano[d]

[a]Istitute of Telecomunications, University of Technology and Life Science, Bydgoszcz, Poland
[b] ITTI Ltd., Poznań, Poland
[b]Polska Telefonia Cyfrowa, Warsaw, Poland
[d] Computer Science Department, Universita' di Napoli Federico II, Napoli, Italy

Abstract—Inherent heterogeneity of the networks increases risk factor and new security threats emerge due to the variety of network types and their vulnerabilities. This paper presents an example of applied security framework – the INTERSECTION. By referring to the ISO/IEC security standards and to the FP7 INTERSECTION project results, authors underline that in the processes of managing and planning security, investigating technology and business governance should be at least as important as formalizing the need for decisions on security cooperation between operators. INTERSECTION provides security mechanisms and introduces capability possible only with a management solution that is at a higher level than that of any of the connected systems alone.

Keywords*—*IDMEF, IDS, IPFIX, security framework.

1. Introduction

Information technology industries as well as telecommunication operators are seeking efficient and comprehensive security solutions. This crucial task not only aims at providing protection against malicious or sometimes inadvertent attacks – it must also address the business requirements for confidentiality, integrity, availability, non-repudiation, accountability, authenticity, and reliability of information and services. Still, an information system is as secure as the weakest element of the system. In many cases network security consists of building blocks provided by vendors specializing in a single aspect of security. This is why interoperability should be considered. Basic interoperability could be achieved by deploying standard protocols for data exchange between security components. Intelligence of the system could be further enhanced with implementation of a management component capable of aggregating such information and able to link otherwise unrelated events into a big picture view. Such a system may provide comprehensive and efficient security defense even in the case of zero-day exploits. Furthermore, the heterogeneity of networks should be taken into consideration, as it may add new vulnerabilities or open otherwise independent networks to new threats.

At the same time another perspective of the same situation can be observed – there is a value in having access to additional monitoring data for correlation in a security framework. Turning adverse situation of supporting various and complex connections between networks into an advantage of high level managed security solutions capable of preventing complex attacks from spreading into multiple networks and geographic areas may be especially interesting to telecom service provides. This task is the aim of the European research project INTERSECTION. Additionally the project focuses on developing new anomaly detection algorithms that can be used with the traffic correlation engine to predict the network behavior and prevent malicious users from accessing the network, stealing information or disrupting a service. By detecting zero-day exploits and automated remediation the security level is further improved.

This paper is divided into sections organized as follows: Section 2 is a summary of the related work in the field of security frameworks. Section 3 describes the various ISO standards addressing telecommunication security management and intrusion detection framework architecture. Section 4 describes the impact of known network threats (like viruses) on companies network and some information about anomaly detection techniques. Section 5 describes the idea of INTERSECTION and protocols used in framework. Section 6 describes the plausible test scenarios for demonstrating the INTERSECTION capabilities. Section 7 introduces idea of converged security. Section 8 then presents security as a service concept. We conclude in Section 9.

2. Related Work

In order to align the described INTERSECTION framework with current state of the art research authors have reviewed most related papers. The areas covered by analyzed papers span from describing technical solutions for improving security: [1], [2], through business perspective: [3], [4], [5] finally to evaluation criteria. In [3] author describes ten aspects that should be taken into account when planning information security. It is interesting to note that infrastructure, tools and supporting mechanisms are the last items on the list of important factors to include. According to author, even more important then security mechanisms is the need for corporate governance responsibility (security is a business issue and not technical issue) as well as enforcement of information security compliance and monitoring.

According to the autor, the latter are absolutely essential. Framework for unified network security management is presented in [1]. This paper defines architecture of a unified security management system for security framework for converged networks. The framework is based on the following principles: coordination of heterogeneous detection tools performing vulnerability and multistage attack analysis visualization and delivering strategic responses across network boundaries. The architecture of the security framework consists of 3 layers: scanning, modeling and application. Scanning layer is monitoring traffic data from different types of network; it analyzes the data by using vulnerability information and database in order to provide security assessment. The modeling tier provides a functional representation of weaknesses found on networks in the form of requirements and impact. The application tier provides a view of the security features of the network to help identifying potential threats to an enterprise. It provides analytical and correlation tools which can be visualized to provide administrators with information that allows to take effective decisions against security threats.

Similarly Onwubiko *et al.* in [2] propose integrated security framework. The framework defines four types of components: sensor components that contribute evidence about security related events, analysis components that implement autonomous software agents capable of synthesizing evidence, an abstract "security space" through which components communicate and finally response components that implement countermeasures. Response components can be configured to incorporate human decision-making in protecting networks. The logical components of the framework are realized on physical network nodes. A physical network node may realize one or more logical components and may interact with one or more security spaces. The above framework follows the generic model for intrusion detection presented in [6].

Hunter in [4] presents the Tivoli case to create an integrated framework approach and the problems found when the company had to interoperate with other management products not embraced by the framework. He underlines that integration is required and that there is a need for standards and protocols that allow different vendors to inter-operate rather than having dedicated integration frameworks. In addition, the idea of autonomic-management is presented, even though the preliminary stage is to identify potential security threats in advance and to alert security managers so that proactive action can be taken. The longer term objective of the Tivoli case is to provide self healing security management and to fix problems automatically.

On the other hand authors in [5] show that although conventional security solutions have been implemented as standalone systems, designed for solving very specific regional problems it is feasible to create integrated security infrastructure with capabilities for dynamic and automatic interaction between heterogeneous security devices. Presented solution combines firewall, intrusion prevention system (IPS), vulnerability scanners and honeypot technologies to assure a security infrastructure. Each component collaborates with the others in order to choose the best action and to launch adequate countermeasures. Exchange of security events between individual security components allows automatic corrective action without user intervention, while keeping the ability to adapt to an evolving environment. Another possibility for improving security level within large organizations is outsourcing.

Author in [7] state that security falls within the area that does not lend itself well to outsourcing because it is too closely tied to the running of the business. Moreover, Gartner suggests that outsourcing security is not appropriate for everyone and has developed decision framework to determine whether in-house or outsourced security is more appropriate [7]. The typical scope of security outsourcing extends to: monitoring security architecture, continuous configuration of security infrastructure, prevention and recovery of incidents. According to the author the major benefit of outsourcing is achieved when the scope of threats is much larger than a company (operator) can provide in its own right. Even if a company has resources to continuously monitor all the events being generated it can only correlate those events happening within its own perimeter.

3. Security Management Standards

The International Organization for Standardization offers suite of standards responsible for providing detailed guidance on the security aspects of the management, operation and use of information system networks, and their interconnections. Security requirements have been gathered in the ISO/IEC series of standards addressing the following areas:

- (ISO/IEC 18028-1) establishes network security requirements and introduce possible control areas and the specific technical areas,
- (ISO/IEC 18028-2) defines a standard security architecture,
- (ISO/IEC 18043) defines the methods for selecting, deployment and operations of intrusion detection system,
- (ISO/IEC 7498-2) the security issues that have to be address within a security system.

Identification and analysis of the communication related factors that should be taken into account to establish network security are the scope of the ISO/IEC 18028-1 stan-

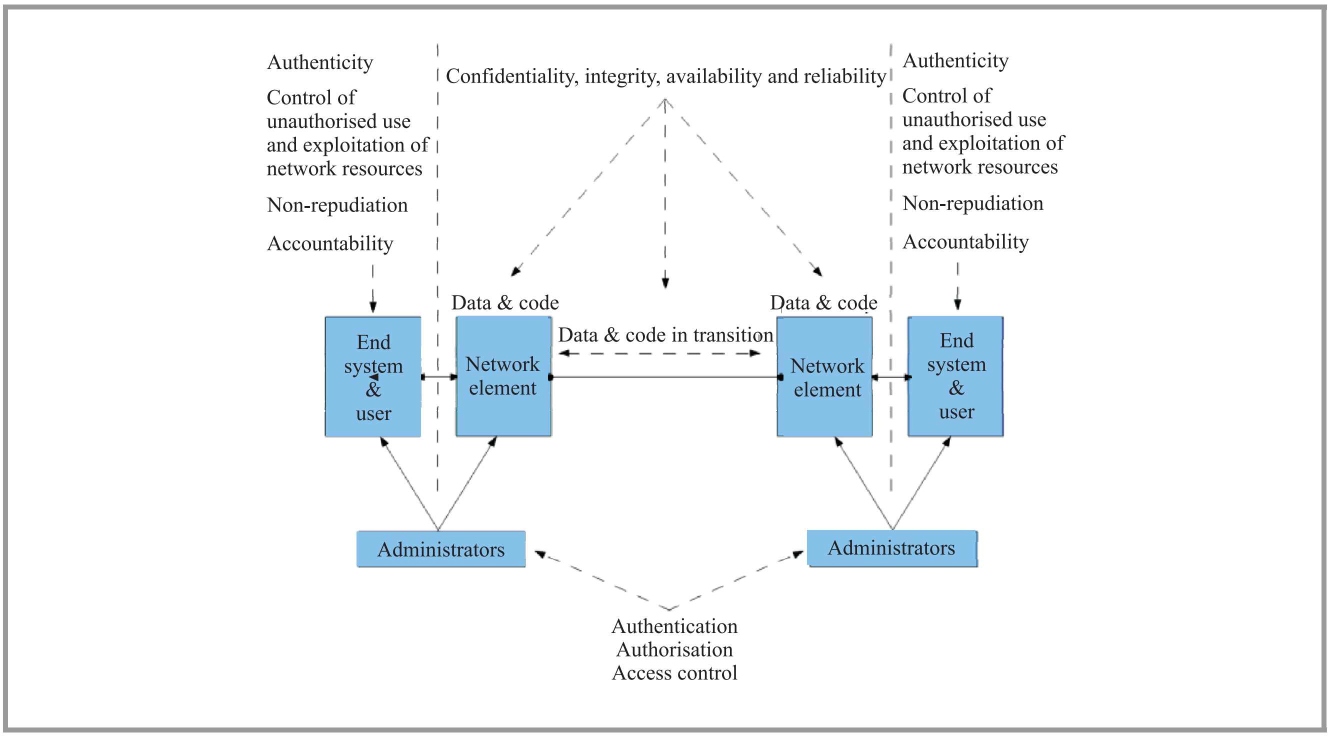


Fig. 1. Conceptual model of network security risk areas [8].

dard. These factors and the corresponding areas of risk are depicted in the Fig. 1.

The results of security risks assessment of a network connection depend on the type and number of networks communicating (e.g., WAN, WLAN, broadband, radio). Selected key risk factors for each type of network are shown in Table 1. When referring to Table 1 one should distinguish between threats and key risk factors. WLAN will certainly be vulnerable to DoS attacks but the impact of such is more severe in WAN or broadband. The same applies for wireless networks. Although radio networks share the same primary security risk with WLAN, there are more prone to disruption due to the possibilities of jamming the system and affecting a considerably greater number of users. Columns in the Table 1 represents the key risks related to particular network whereas speaking about connection that uses for instance WLAN and WAN one should intersect risk factors from both networks. Each risk factor represents certain threat to the system.

Table 1
Key risk factors according to connection type [8]

Risk	WAN	WLAN	Radio	Broadband
Intrusion	+			
DoS	+		+	+
Eavesdropping		+	+	
Unauthorized access		+		
Misconfiguration		+		+
Flawed WEP or TKIP		+		
Session hijacking			+	
Propagation of malicious code				+
UL/DL of unauthorized access				+

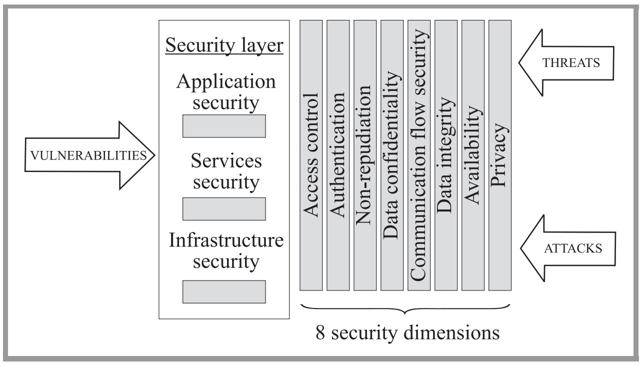


Fig. 2. Security conceptual architecture [10].

According to the ISO 7498-2:1989 specification [9] various threats may be grouped and categorized as follows:

- destruction of information and/or resources (I),
- corruption or modification of information (II),
- theft, removal of loss of information and other resources (III),
- disclosure of information (IV),
- interruption of services (V).

Particular threats should be addressed by defining a set of principles that describe a security structure for the end-to-end security solution. According to ISO/IEC 18028-2 the most generic security framework aimed at combating broad range of threats can rely on the eight-dimensional model as presented in Fig. 2. The figure depicts the concept of protecting a network by defining security dimensions at each security plane of each security layer to provide comprehensive security solutions. Thus according to [10], to be resilient, an end-to-end security solution must address the spectrum of depictured areas and dimensions. Protection elements have to be placed throughout the network to protect the company from malicious attacks. The target coverage of threats by the well established security dimensions in an organization is presented in Table 2.

Table 2
Threats and security dimension relation [10]

Security dimension	Security threat				
	I	II	III	IV	V
Access control	Y	Y	Y	Y	
Authentication			Y	Y	
Non-repudiation	Y	Y	Y	Y	Y
Data confidentiality			Y$^{(*)}$	Y$^{(*)}$	
Comm. flow security			Y	Y	
Data integrity	Y$^{(*)}$	Y$^{(*)}$			
Avaliability	Y$^{(*)}$				Y$^{(*)}$
Privacy				Y	
$^{(*)}$ feasible with IDS.					

Network security is achieved by addressing a specific group of threats (column name refers to the numbering in the threat list above) with a security component or system that provides functionalities described by given dimension (row). When mitigating particular risk with the proper countermeasure a certain level of security is achieved – which can further be extended by applying more mature solutions and robust security components. A practical way to enhance the security level is to introduce an intrusion detection system (IDS) in the network. IDS will, by definition, cover certain threats in the context of eight-dimensional security model (Table 2). According to [11] generic IDS should address the authentication, integrity, confidentiality and availability dimensions as indicated by Table 2. It is worth noticing that by addressing only four out of eight security dimensions IDS can cover a complete spectrum of security threats. A generic model for IDS defined by [11] is presented in Fig. 3.

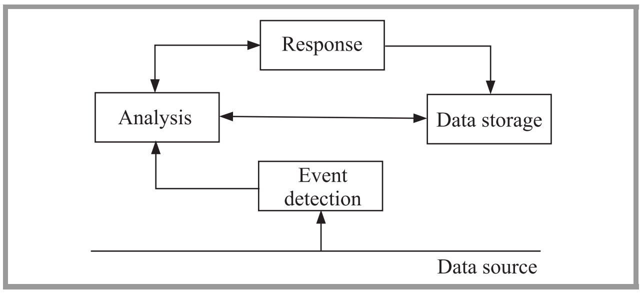


Fig. 3. Generic model of intrusion detection [6].

The event detection module will gather data scattered around the network; this will include information about interfaces, traffic, active users and system logs. Data correlation will take place inside the analysis block, where patterns of properly functioning network will be defined. All data is stored in a data storage module. If an IDS works in anomaly detection mode the system can compute the traffic profiles for normal behavior and compare it to ongoing traffic to determine possibility of an attack. Once the attack is detected the IDS can in turn scan set of available countermeasures and with a presence of a response module – reconfigure the network devices or interfaces to slow down the attack, thus providing enough time for the system administrator to trace the intrusion source. A secure network may contain single IDS as well as multiple IDSes spread through the network. Hierarchical architecture is proposed in [11] for multiple IDS management as shown in the Fig. 4.

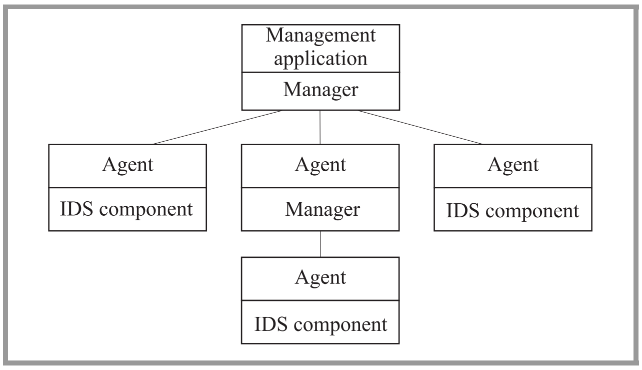


Fig. 4. An intrusion detection management model [6].

The more data is gathered from agents for analysis the more reliable decision can be made by manager and an ongoing attack may be detected in less time. Thus according to [6] it could be beneficial for operators to share data on intrusion information and interconnect their IDS. The ISO/IEC 18043 advises such solution but also points out that operators are not willing to give their knowledge of intrusions that have affected their IT systems to the public, as it could reveal their business operations. This is even more important when we take zero day exploits under consideration. Well known worms like Witty or Slammer [12] have caused tremendous financial losses to many companies around the world. Would the worms be more destructive and target mostly critical infrastructure networks their impact could be far more severe. The next section provides use-case rationale for developing applied security infrastructure that is capable of aggregating and linking otherwise unrelated events into a big picture view to increase protection level.

4. Rationale for INTERSECTION

On the 25th January 2003 a virus called Slammer (sometimes also Sapphire) started infecting hosts by exploiting a buffer-overflow security hole in computers connected to Internet that were running the Microsoft SQL server and Microsoft SQL server desktop engine (MSDE) 2000 [12]. Once a host was infected the worm started scanning random IP addresses to spread further. Figure 5 presents the number of packets send by Slammer from infected locations during the first 12 hours after activation. Because Slammers behavior was highly anomalous (e.g., regarding amplified traffic envelope) it could be detected by a method called network telescope [12]. Success in suppresing the virus was achieved by analyzing intrusion detection system logs gathered from attacked companies and history of events collected by NMS systems.

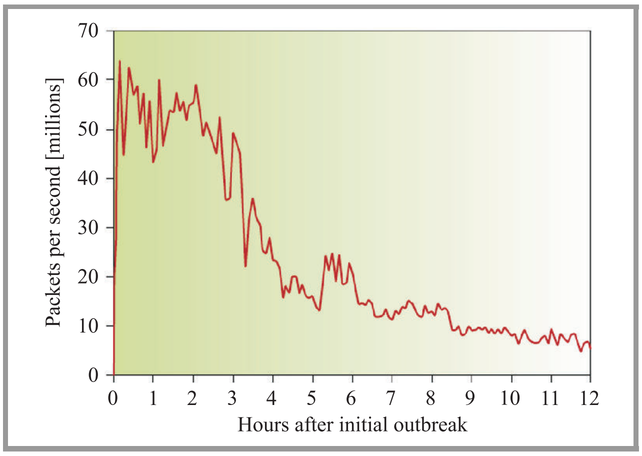


Fig. 5. The response to Slammer during the 12 hours after its release [12].

Would the infected systems been able to exchange information between different companies IDS the worm could have had respectively smaller impact and could have been suppressed from spreading worldwide. The distributed IDS could sufficiently increase the security of network operator infrastructures engaged in a supporting communication with malicious traffic in-band. The data for analysis could be spread through the network so if one operator would face the attack another one could benefit from his experience by exchanging information about pattern of anomalous (attacked) traffic between IDS. The INTERSECTION framework among other goals aims at proposing new anomaly detection algorithms as well as investigation of known algorithms [13] and providing a security framework that interconnects different network operators, which in turn allows exchanging traffic flow information between them. This could lead to enhancing the current security solutions by the factor proportional to the synergic effect of information exchange between operators.

5. A View on INTERSECTION

The aim of the INTERECTION project is to come up with specifications of an integrated framework for security and resiliency in complex and heterogeneous communication networks. Three objectives have been identified during the architecture process:

- to define what data must be shared among security systems of critical infrastructures and to specify hierarchy of communication and rules for data access,
- to design an integrated framework for securing networked systems,
- to specify appropriate protocols enabling communication between security systems in order to assure interoperability in an inter-domain environment.

Figure 6 presents a general overview of the proposed INTERSECTION framework. The INTERSECTION framework includes the following components for: monitoring, detection, reaction, remediation, visualization, and topology discovery. Integration of these components and exchange of information between modules collecting data from heterogeneous environments leads to improved network protection and security for participating systems. Monitoring, detection, reaction, and remediation components cooperate in real time and in automated fashion. These modules gather data from probes and network elements, analyze it, detect intrusions and anomalies, and select the most suitable reaction (e.g., reconfiguration of network components).

Remediation module is responsible for taking appropriate action in order to prevent similar attacks in the future. A network must operate at least one remediation point (e.g., at a gateway) in order to effect remedies, but may operate several if appropriate (e.g., one per border router) or additional that actually exist in neighbor networks (provided co-operation of those networks). The INTERSECTION framework also includes offline functions aiming at using data coming from the network, or provided by the real-time elements, to help the human operator in analyzing the network state and to evaluate configuration changes implemented by remediation module. The offline functions include topology discovery, visualization and anomaly detection. It is also important to highlight the relevant protocols used for data exchange within INTERSECTION. These include: IDEMF and IPFIX. A short description of each protocol is provided in the next two subsections.

5.1. IDEMF

IDMEF is an XML domain specific language (DSL) for intrusion detection systems. Its purpose is to provide an homogeneous environment to improve the network security.

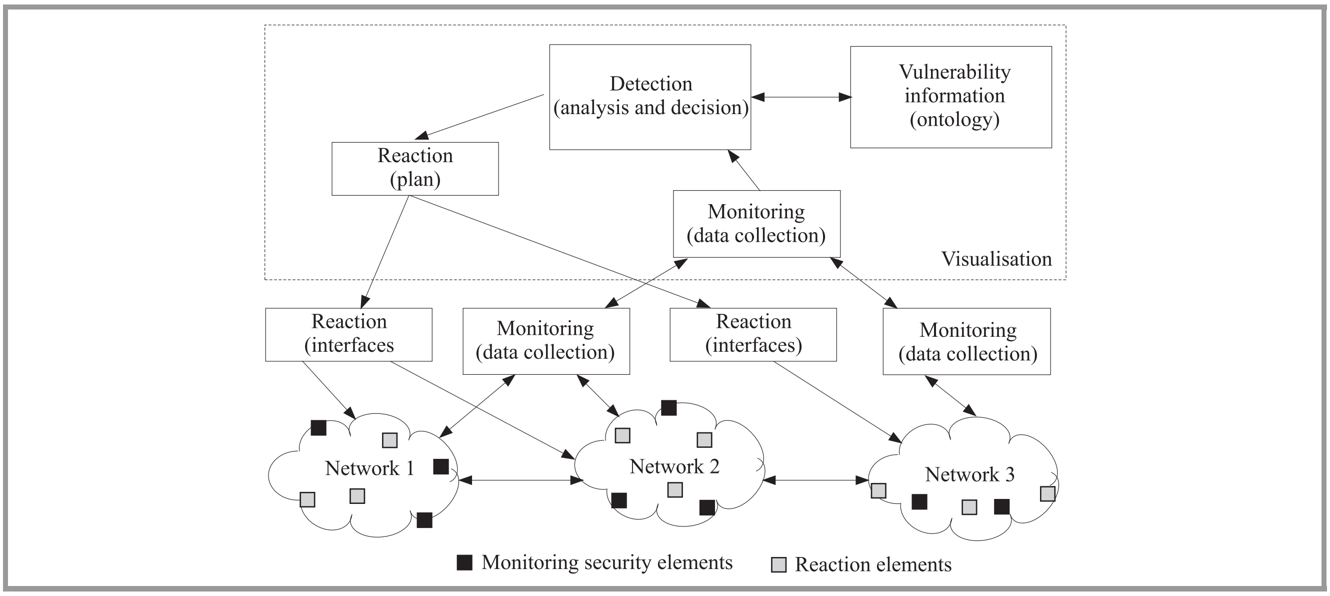


Fig. 6. INTERSECTION framework.

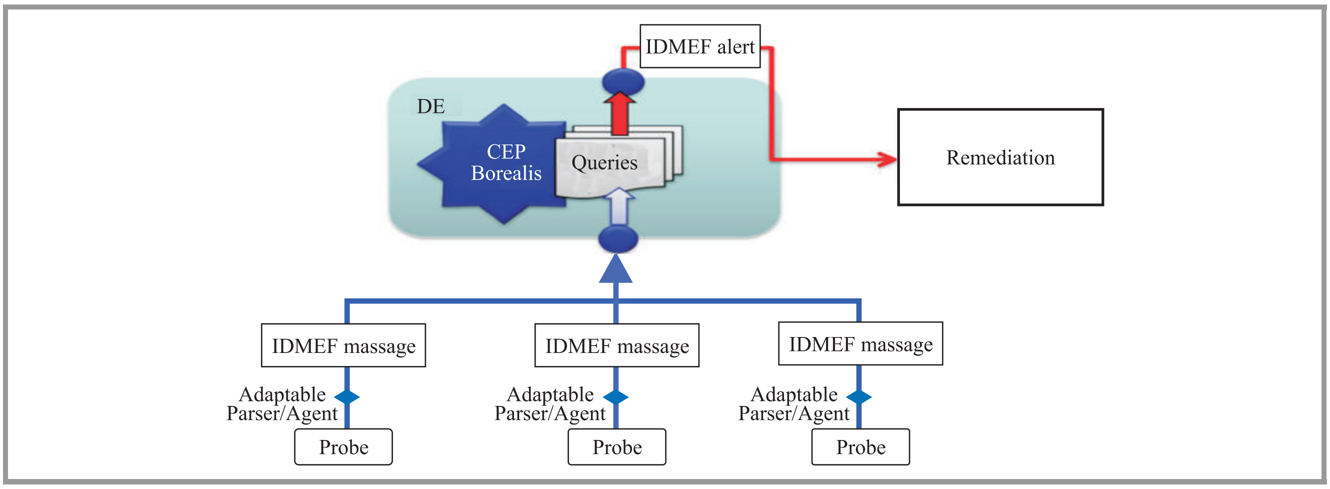


Fig. 7. Usage of IDMEF in INTERSECTION.

Figure 7 shows how IDMEF format is used in INTERSECTION. The decision engine (DE) aggregates events gathered from multiple probes (Host IDSes, Network IDSes, DB monitors, etc.). Correlation of these events is performed by complex event processor (CEP), namely Borealis correlation engine, developed jointly by Brandeis University, Brown University and MIT.

IDMEF messages are used to transmit information from the probes, and to send alerts to the remediation component.

5.2. IPFIX

IPFIX is an IETF working group standard [14]. It was created from the need for a common, universal standard for exporting the Internet protocol flows information from routers, probes, and other devices that are used by mediation systems and network management systems to facilitate services such as measurement, accounting and billing. Within INTERSECTION IPFIX was used for the measurement task – a task that can be initiated by one of three components: measurement controller, IDS or visualization. Probes in INTERSECTION are called OpenIMP probes. Figure 8 shows that the monitoring system uses multiple measurement units (probes), which are distributed within the network and passively monitor network traffic. In addition, the monitoring system includes a postprocessor, collector, management and control interface.

The following section describes the proposed test scenarios within INTERSECTION project.

6. Test Scenarios

The INTERSECTION defined the suite of test scenarios to evaluate performance and detection, remediation and visualization capability of the proposed framework. This

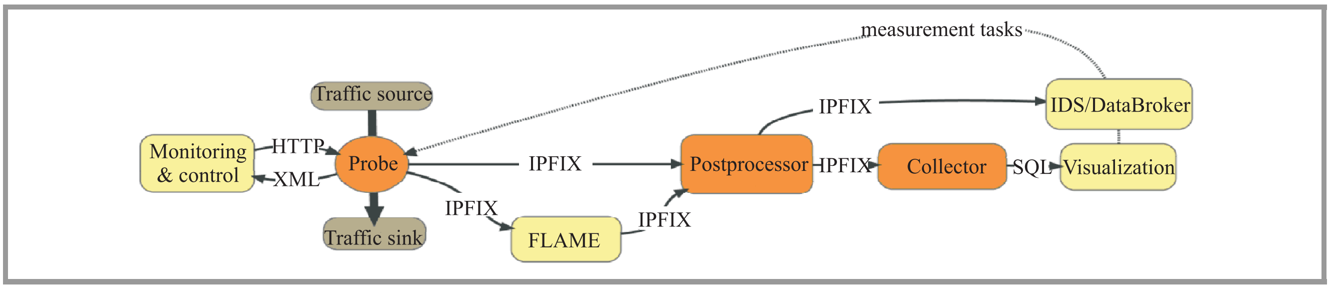


Fig. 8. Role of IPFIX in INTERSECTION.

section enumerates five different demo scenarios that have been designed to show how the INTERSECTION framework can effectively detect and resolve attacks by analyzing different pieces of information obtained from different networks. Furthermore, some scenarios show how the INTERSECTION framework is capable of detecting attacks by correlating data that, when analyzed separately would not provide enough information to detect the attack and to correct the system configuration settings. Each one of the five scenarios has been designed to be run over an interconnected infrastructure, the INTERSECTION demo network presented in Fig. 9, which is setup by the project partners. This interconnection of networks is necessary since an important premise, when designing the demo scenarios, was heterogeneity. In fact, the heterogeneity in the demo scenarios is addressed in the following ways:

- Each demo scenario involves at least two demo labs of different access technology interconnected, thus showing that the designed INTERSECTION framework can deal with access network heterogeneity. The interconnected infrastructure of the different demo labs, called INTERSECTION demo network, consists of five laboratories of different communication technologies (including satellite, wireless and wired networks) connected in a full-mesh network.

- Demo scenarios show how the INTERSECTION framework combines detection techniques from different access technologies with other detection techniques independent from the access technology, thus providing a richer framework for detection of attacks. Even if the attack exploits a vulnerability related to a specific access technology, information from other networks can contribute to the detection of the attack.

The demo scenarios are based on exploiting specific vulnerabilities that are currently present in networks. In summary, a demonstration case is a realistic story about how a vulnerability of a certain technology or equipment can be exploited, how the attack will be detected, how some mechanisms will be activated to solve the attack and how this process of detection and remediation can be shown to the network administrator through a visualization framework. Unlike a usual attack scenario, in these demo cases the attack is not detected by just analyzing the network where the attack is performed, but by correlating the information from different networks involved in the monitored environment. The demo scenarios investigated by the INTERSECTION project are:

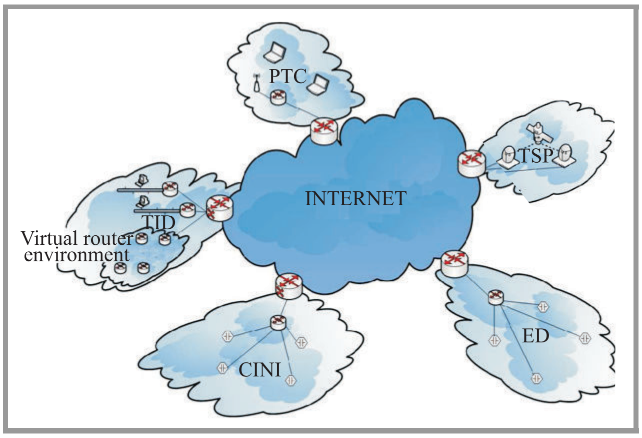


Fig. 9. he INTERSECTION demonstrator.

- loss of access to a content provider by hijacking prefixes of its corresponding autonomous system,
- satellite PEP spoofing,
- multistage attack on high profile roaming user and his network,
- injection of bogus packets in a wireless sensor network,
- distributed denial of service attack.

One of the INTERSECTION objectives is to show the applicability of the advanced decision support tools for mitigation, response and recovery in a heterogeneous environment. For this reason, a set of the aforementioned demo scenarios have been identified. Even though INTERSECTION is a research framework it is designed to be deployed by any operator already using commercial solutions that are incorporating IDMEF and IPFIX protocols.

7. Towards Converged Security

Numerous network security systems are currently available on the market. Authors envisage two categories of systems:

– systems that can manage one or more security areas, but not the end-to-end security environment of the organization,

– systems that have the capability of aggregating information from multiple sources and managing the whole environment.

First branch of systems include (but is not limited to) network access control, self-defending networks, and security gateways whereas the second is focused on so called management solutions. The latter include COTS products used for collecting, maintaining and reporting network traffic providing services such as centralized log system, user notification, activity monitoring. It can be seen that INTERSECTION does not address all security issues like privacy, communication flow security, non-repudiation and access control required for a complete security management system. However, a strong correlation of INTERSECTION to standard protocols for data exchange and proposed strategy for interconnecting networks of independent operators and their customers shows the way towards unified approach to network security in our highly interconnected world. One way to evaluate security solution is to map it against the security maturity model. Attributes such as company size, industry regulations, liability, technical complexity, culture, risk tolerance, and the level of dependence on physical and logical assets all create distinct requirements for risk management and security convergence. However, there are several attributes common to a mature, converged security organization. In [15] maturity attributes of a company are presented (Table 3).

Table 3
Converged security maturity attributes – excerpt from [15]

Maturity attribute	Defense-in-depth
Immature (ad hoc)	There is no formal security structure
Aware (repeatable but intuitive)	Security is focused on perimeter defense
Management and risk-based (defined)	Safeguards extended beyond the perimeter, but remain technically focused
Common (optimized)	There is true defense-in-depth encompassing people, policy, and processes with technology. Thrid-party and mobility issues are included

INTERSECTION framework, while not covering some of the aspects of the converged rank in maturity model in some way may stretch the model beyond current definition. INTERSECTION framework envisions participation in a solution that not only includes internal policies and processes of an organization but provides enhancements and introduces capability possible only with a management solution that is at a higher level than any of the connected systems alone.

8. Security as a Service

As the software paradigm shifts towards cloud computing the more important it appears not only to provide means for better security but also to incorporate security solutions that span across domains and gain from the knowledge/experience of "first" victims in order to protect others. In computer networks there is a problem of extremely high speed of data/message exchange between host/networks during attack. The so called zero-day exploits are the effects of malicious activity of attackers that may affect huge number of network users (from individual to corporational). Thus important dimensions for improved threat detection and prevention (also tolerance) are time and knowledge. Time factor covers the time period to detect malicious activity as well as time to find and apply countermeasures best matching to the context. On the other hand knowledge sharing is essential in keeping security best practices up to date each time security flaw is detected and providing framework for building and exchanging rules to apply (e.g., in the context of security threat) remediation policy of an organization. Some aspects limiting the proper take-off of the 3S paradigm are related to both business view (security maturity of an organization, information exchange strategy) and regulatory framework of a given country (obligation for anonymization of logs). Deployment of INTERSECTION enables implementation of the paradigm of security as a service external to an organization. One can imagine that monitoring and decision engines are located outside of a company network and managed by trusted third party.

9. Conclusions

The growth of Internet connectivity results in increased security requirements for enterprises to achieve services availability as described by SLA agreements (for end users and between operators). Inherent heterogeneity of the networks increases risk factor and new security threats emerge due to the variety of network types and their vulnerabilities. The solution proposed by INTERSECTION aims at providing security-level interoperability between many operators using different network technologies and different security solutions. The real benefit of exploiting INTERSECTION as an example of applied security framework paradigm can be capitalized if the key operational assumptions are fulfilled. The network owners and service providers should agree on the need to foresee security related data exchange as an important substrate of a successful security policy. The wide spread of malicious code that is remotely commanded to trigger distributed DoS attacks at any time decided by a hacker, calls for the real cooperation that is fostered by telecommunication regulatory institutions. Currently the Polish telecommunication law for instance states that cooperation between telecommunication operators is the obligation of the operator only at times of crisis situations [16]. So it is up to the operator to make its internal information accessible for other operators. The continual improvement (as a business process) of individual organizations security infrastructure is essential but there is an even more important aspect in holistic security supremacy that is only possible when security information exchange requirement is fulfilled. It should be at least as important to investigate technology and business governance as to formalize the need for decisions on security cooperation in the process of managing and planning security. From this perspective the meaning of the attribute of converged security maturity of an organization is stretched as INTERSECTION provides enhancements and introduces capability possible only with a management solution that is at a higher level than any of the connected systems alone.

Acknowledgements

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 216585 (INTERSECTION Project).

References

[1] J. Dawkings, "A framework for unified network security management: identifying tracking security threats on converged networks", *J. Netw. Sys. Manag.*, vol. 13, no. 3, 2005.

[2] C. Onwubiko, A. P. Lenaghan, L. Hebbes, "An integrated security framework for assisting in the defense of computer networks", in *Proc. Mobile Future 2006 and the Symposium on Trends in Communications SympoTIC'06*, 2006, pp. 52–55.

[3] B. Von Solms, "The ten deadly sins of information security management", *Comp. Secur.*, vol. 23, pp. 371–376, 2004.

[4] P. Hunter, "Lack on integration undermines IT security", *Netw. Secur.*, vol. 2003, no. 1, pp. 5–7, 2003.

[5] M. Sourour, B. Adel, and A. Tarek, "Ensuring security in depth based on heterogeneous network security technologies", *Int. J. Inf. Secur.*, vol. 8, no. 4, pp. 233–246, 2009.

[6] Technical Report ISO/IEC TR 15947 "Information technology-security techniques-IT intrusion detection framework. Part 1: Network security management", 2002.

[7] M. Withworth, "Outsourced security – the benefits and risks", *Netw. Secur.*, vol. 2005, no. 10, pp. 16–19, 2005.

[8] International Standard ISO/IEC 18028-1 "Information technology-security techniques-IT network security. Part 1: Network security management", 2006.

[9] ISO 7498-2:1989 – CCIT Rec. X.800 (1991).

[10] International Standard ISO/IEC 18028-2 "Information technology-security techniques-IT network security. Part 2: Network security architecture", 2006.

[11] International Standard ISO/IEC 18043 "Information technologu-security techniques-selection, deployment and operations of intrusion detection systems", 2006.

[12] D. Moore, V. Paxson, S. Savage, C. Shannon, S. Stainford, and N. Weaver, "Inside the Slammer worm", *IEEE Secur. Priv.*, vol. 1, no. 4, p. 33–39, 2003.

[13] Ł. Saganowski, M. Choraś, R. Renk, W. Hołubowicz, "A novel signal-based approach to anomaly detection in IDS systems", *Lecture Notes in Computer Science*, vol. 5495, pp. 527–536, 2009.

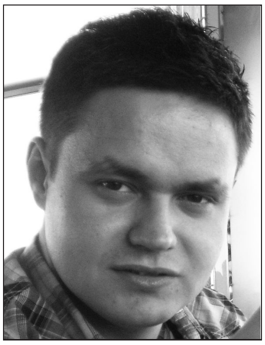
[14] Request for Comments RFC 5101, "Specification of the IP flow information export (IPFIX) protocol for the exchange of IP traffic flow information", 2008.

[15] K. Anderson, "Convergence: a holistic approach to risk management", *Netw. Secur.*, vol. 2007, no. 5, pp. 4–7, 2007.

[16] *Polish Telecommunication Law*, act from 16th July 2004.

Adam Flizikowski obtained his M.Sc. from University of Technology and Agriculture in Bydgoszcz, Poland, in 2000. Since 2003 he holds a research position of professor assistant at the University of Technology and Life Sciences. He also had managed projects dealing with telecommunication aspects like end to end QoS management, policy based networking, mobile service automated evaluation, methodology of evaluation of video servers and STB platforms, and so on. He has working experience with IT system design methodology and especially applied ontology based systems. Currently he works on dissertation on admission control in 4G WiMAX networks.

e-mail: adamfli@utp.edu.pl
Istitute of Telecomunications
University of Technology and Life Science
Prof. Kaliskiego st 7
85-796 Bydgoszcz, Poland



Mateusz Majewski obtained his M.Sc. from University of Technology and Agriculture in Bydgoszcz, Poland, in 2009. Since 2008 he has been working in the ITTI Ltd. in Poznań. In the recent years he participated in EU projects as a junior consultant. During the years he was involved in projects covering such topics as QoS aspects in wireless networks, simulation methodologies, security framework and applied security.

e-mail: mateusz.majewski@itti.com.pl
ITTI Ltd.
Rubież st 46
61-612 Poznań, Poland

Maria Hołubowicz obtained her M.Sc. from Poznań University of Technology, Poland, in 2009. She has been working in ITTI Ltd. in Poznań since 2007. She has contributed to the preparation of a number of proposals to FP7 ICT and Security calls. She has participated in EU projects as a junior consultant and in projects that involve GUI design, knowledge based systems, network security systems and ontology design for knowledge based systems for EDA projects.

e-mail: carolina@itti.com.pl
ITTI Ltd.
Rubież st 46
61-612 Poznań, Poland

Zbigniew Kowalczyk obtained M.Sc. and MBA from Kozmiński University in Warsaw. Since 2000 he has worked as an IT Architect, IT Consultant, and Project Manager for Compugen of Richmond Hill (Canada). He holds number of professional certifications, including PMP and CISSP. Since the beginning of 2009 he is involved in EU projects as part of the team of Polska Telefonia Cyfrowa (PTC). He contributes to Tuesday Technology Report, a bi-monthly electronic newspaper, writing on IT security trends and issues.

e-mail: zkowalczyk@era.pl
Polska Telefonia Cyfrowa
Jerozolimskie Av. 181, B2.13
02-222 Warsaw, Poland

Simon Pietro Romano received the degree in computer engineering from the University of Napoli Federico II, Italy, in 1998. He obtained a Ph.D. degree in Computer Networks in 2001. He is currently an Assistant Professor at the Computer Science Department of the University of Napoli. His research interests primarily fall in the field of networking, with special regard to QoS-enabled multimedia applications, network security and autonomic network management. He is currently involved in a number of research projects, whose main objective is the design and implementation of effective solutions for the provisioning of services with quality assurance over Premium IP networks. He is member of both the IEEE Computer Society and the ACM.

e-mail: spromano@unina.it
Computer Science Department
Universita' di Napoli Federico II
Via Claudio 21
80125 Napoli, Italy

Paper

Anomaly Detection Framework Based on Matching Pursuit for Network Security Enhancement

Rafał Renk and Witold Hołubowicz

ITTI Ltd., Poznań, Poland
Adam Mickiewicz University, Poznań, Poland

Abstract—In this paper, a framework for recognizing network traffic in order to detect anomalies is proposed. We propose to combine and correlate parameters from different layers in order to detect 0-day attacks and reduce false positives. Moreover, we propose to combine statistical and signal-based features. The major contribution of this paper are: novel framework for network security based on the correlation approach as well as new signal based algorithm for intrusion detection using matching pursuit.

Keywords—anomaly detection, intrusion detection, matching pursuit, network security, signal processing.

1. Introduction and Motivation

Intrusion detection systems (IDS) are based on mathematical models, algorithms and architectural solutions proposed for correctly detecting inappropriate, incorrect or anomalous activity within a networked systems. Intrusion detection systems can be classified as belonging to two main groups depending on the detection technique employed: anomaly detection and signature-based detection. Anomaly detection techniques, that we focus on in our work, rely on the existence of a reliable characterization of what is normal and what is not, in a particular networking scenario. More precisely, anomaly detection techniques base their evaluations on a model of what is normal, and classify as anomalous all the events that fall outside such a model. If an anomalous behavior is recognized, this does not necessarily imply that an attack activity has occurred: only few anomalies can be actually classified as attempts to compromise the security of the system.

Anomaly detection systems can be classified according to:

- the used algorithm,
- analyzed features of each packet singularly or of the whole connection,
- the kind of analyzed data - whether they focus on the packet headers or on the payload.

Most current IDS systems have problems in recognizing new attacks (0-day exploits) since they are based on the signature-based approach. In such mode, when system does not have an attack signature in database, such attack is not recognized. Another drawback of current IDS systems is that the used parameters and features do not contain all the necessary information about traffic and events in the network.

Therefore, in this paper we present the framework in which anomaly detection system based on correlation and diversity approaches are used, such as:

- Item diversity – different network layers parameters are monitored and used. In such approach we do not have information from transport layer only – such information is merged/correlated with application layer events.
- Correlation – correlation is used twofold (during decision):
  - item both anomaly and signature-based approaches are correlated,
  - parameters/features from various network layers are correlated,
  - statistical and signal-based features are used and correlated.

2. Technical Solution

In this paper, a new solution for aanomaly detection system (ADS) based on signal processing algorithm is presented. ADS analyzes traffic from Internet connection in certain point of a computer network. The proposed ADS system uses redundant signal decomposition method based on matching pursuit algorithm. ADS based on matching pursuit uses dictionary of base functions (BFD) to decompose input 1D traffic signal (1D signal may represent packets per second) into set of based functions called also atoms. The proposed BFD has a ability to approximate traffic signal. Number and parameters of base functions was limited in order to shorten atom search time process.

Since some attacks are visible only in specific layer (e.g., SQLIA), in our approach, we propose to use network parameters from different layers.

Transport layer, network layer and application layer parameters are used.

In the further step, we use the presented parameters to calculate characteristics (features) of the observed traffic. Some of the parameters are used for statistical features calculation and/or for signal-based feature calculation respectively. Feature extraction methods are presented in the following subsections.

2.1. Statistical Features

The chi-square multivariate test for anomaly detection systems can be represented by:

$$X^2 = \sum_{i=1}^{p} \frac{(X_i - \overline{X}_i)^2}{\overline{X}_i}, \tag{1}$$

where $X = (X_1, X_2, \ldots, X_p)$ denote an observation of p variables from a process at time t and $\overline{X} = (\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_p)$ is the sample mean vector.
Using only the mean vector in Eq. (1), cause that chi-square multivariate test detects only the mean shift on one or more of the variables.

2.2. Signal Processing Features

Signal processing techniques have found application in network intrusion detection systems because of their ability to detect novel intrusions and attacks, which cannot be achieved by signature-based approaches [1]. It has been shown that network traffic presents several relevant statistical properties when analyzed at different levels (e.g., self-similarity, long range dependence, entropy variations, etc.) [2].
Approaches based on signal processing and on statistical analysis can be powerful in decomposing the signals related to network traffic, giving the ability to distinguish between trends, noise, and actual anomalous events. Wavelet-based approaches, maximum entropy estimation, principal component analysis techniques, and spectral analysis, are examples in this regard which have been investigated in the recent years by the research community [3], [4], [5], [6], [7]. However, discrete wavelet transform provides a large amount of coefficients which not necessarily reflect required features of the network signals.
Therefore, in this paper we propose another signal processing and decomposition method for anomaly/intrusion detection in networked systems. We developed original anomaly detection type IDS algorithm based on matching pursuit.
In the rest of the paper, our original ADS method will be presented in details. Moreover, results of experimental setup will be given. We tested our method with standard traces in worm detection scenario as well as in anomaly detection scenario. Discussion on redundant dictionary parameters and final conclusions will be provided.

Matching pursuit

Matching pursuit signal decomposition was proposed by Mallat and Zhang [8].
Matching pursuit is a greedy algorithm that decomposes any signal into a linear expansion of waveforms which are taken from an over complete dictionary D. The dictionary D is an over complete set of base functions called also atoms.

$$D = \{\alpha_\gamma : \gamma \in \Gamma\}, \tag{2}$$

where every atom α_γ from dictionary has norm equal to 1, $\|\alpha_\gamma\| = 1$, Γ represents set of indexes for atom transformation parameters such as translation, rotation and scaling.

Signal s has various representations for dictionary D. Signal can be approximated by set of atoms α_k from dictionary and projection coefficients c_k

$$s = \sum_{n=0}^{|D|-1} c_k \alpha_k. \tag{3}$$

To achieve best sparse decomposition of signal s (min) we have to find vector c_k with minimal norm but sufficient for proper signal reconstruction. Matching pursuit is a greedy algorithm that iteratively approximates signal to achieve good sparse signal decomposition. Matching pursuit finds set of atoms α_{γ_k} such that projection of coefficients is maximal. At first step, residual R is equal to the entire signal $R_0 = s$.

$$R_0 = \langle \alpha_{\gamma_0}, R_0 \rangle \alpha_{\gamma_0} + R_1. \tag{4}$$

If we want to minimize energy of residual R_1 we have to maximize the projection. $|\langle \alpha_{\gamma_0}, R_0 \rangle|$. At next step we must apply the same procedure to R_1

$$R_1 = \langle \alpha_{\gamma_1}, R_1 \rangle \alpha_{\gamma_1} + R_2. \tag{5}$$

Residual of signal at step n can be written

$$R^n s = R^{n-1} s - \langle R^{n-1} s \big| \alpha_{\gamma_k} \rangle \alpha_{\gamma_k}. \tag{6}$$

Signal s is decomposed by set of atoms

$$s = \sum_{k=0}^{N-1} \langle \alpha_{\gamma_k} | R^n s \rangle \alpha_{\gamma_k} + R^n s. \tag{7}$$

Algorithm stops when residual $R^n s$ of signal is lower then acceptable limit.

Our approach to intrusion detection algorithm

In basic matching pursuit algorithm atoms are selected in every step from entire dictionary which has flat structure. In this case algorithm causes significant processor burden. In our coder dictionary with internal structure was used. Dictionary is built from:

- atoms,
- centered atoms.

Centered atoms groups such atoms from D that are as more correlated as possible to each other. To calculate measure of correlation between atoms function $o(a, b)$ can be used [9]

$$o(a,b) = \sqrt{1 - \left(\frac{|\langle a, b \rangle|}{\|a\|_2 \|b\|_2} \right)^2}. \tag{8}$$

The quality of centered atom can be estimated according to

$$O_{k,l} = \frac{1}{|LP_{k,l}|} \sum_{i \in LP_{k,l}} o\big(A_{c(i)}, W_{c(k,l)}\big), \tag{9}$$

where: $LP_{k,l}$ is a list of atoms grouped by centered atom, $O_{k,l}$ is mean of local distances from centered atom $W_{c(k,l)}$ to the atoms $A_{c(i)}$ which are strongly correlated with $A_{c(i)}$. Centroid $W_{c(k,l)}$ represents atoms $A_{c(i)}$ which belongs to the set $i \in LP_{k,l}$. List of atoms $L_{k,l}$ should be selected according to the equation:

$$\max_{i \in LP_{k,l}} o\big(A_{c(i)}, W_{c(k,l)}\big) \leq \min_{t \in D \backslash LP_{k,l}} o\big(A_{c(t)}, W_{c(k,l)}\big). \tag{10}$$

In the proposed IDS solution 1D real Gabor base function (equation was used to build dictionary) [9], [10], [11]

$$\alpha_{u,s,\xi,\phi}(t) = c_{u,s,\xi,\phi}\,\alpha\Big(\frac{t-u}{s}\Big)\cos\big(2\pi\xi(t-u)+\phi\big), \tag{11}$$

where:

$$\alpha(t) = \frac{1}{\sqrt{s}} e^{-\pi t^2}, \tag{12}$$

$c_{u,s,\xi,\phi}$ – is a normalizing constant used to achieve atom unit energy.

In order to create over complete set of 1D base functions dictionary D was built by varying subsequent atom parameters: frequency ξ and phase ϕ, position u, scale s.

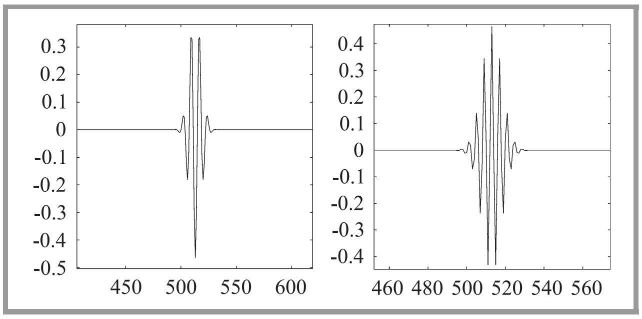


Fig. 1. Example atom from dictionary.

Base functions dictionary D was created with using 10 different scales (dyadic scales) and 50 different frequencies. In Fig. 1 example atoms from dictionary D are presented.

3. Experimental Results

Percentage of the recognized anomalies as a function of encoded atoms from dictionary of base functions is presented in Fig. 2. Five dictionaries with different parameters (different number of scales and frequencies) were used in our ADS system.

Percentage of the recognized anomalies for dictionary of base functions with approximately constant number of atoms is presented in Fig. 3. In this case we try to leave approximately constant number of atoms in dictionary but with different proportions of scales and frequencies.

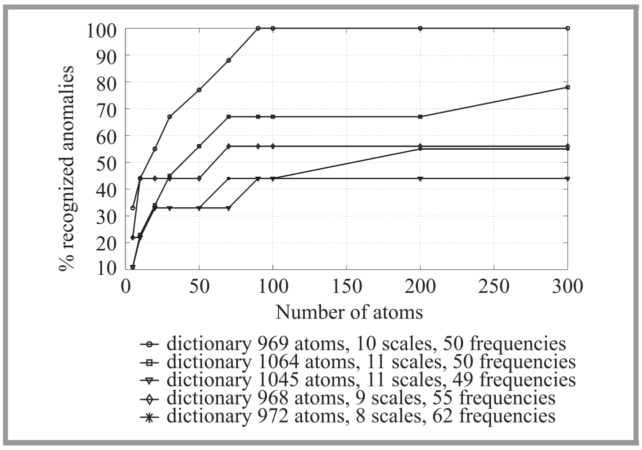


Fig. 2. Percentage of the recognized anomalies as a function of encoded atoms.

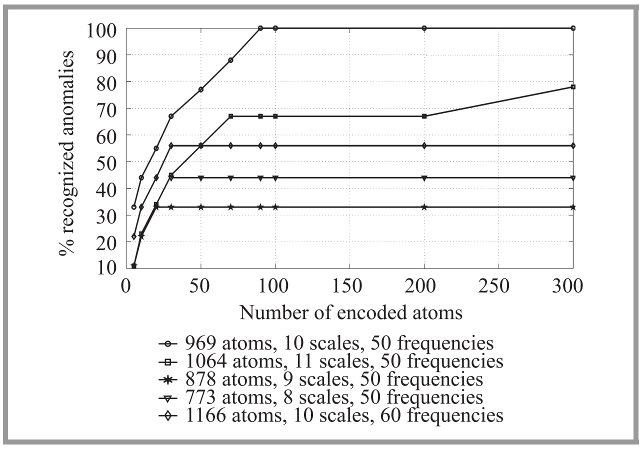


Fig. 3. Percentage of the recognized anomalies for Dictionary of Base Functions with approximately constant number of atoms.

In Tables 1, 2, 3, 4 there are example results taken from our ADS system. Traffic traces were analyzed by proposed ADS with the use of 20 minutes windows (most attacks (more than 80%) last no longer then 20 minutes). In every window we calculate matching pursuit mean projection parameter in order to recognize suspicious traffic behavior. Analyzed traces are infected by worms (Tables 1 and 2), DDos (Table 4) and DDoS SYNFlood (Table 3) attacks.

4. Conclusions

In this paper a framework for recognizing attacks and anomalies in the computer networks is presented. Our methodology is based on both statistical and signal based features. The major contribution and innovation is the application of matching pursuit algorithm to calculate network

Table 1
Matching pursuit mean projection for TCP trace [12]. Traces are analyzed with the use of 20 minutes windows

TCP trace	Window1 MP-MP	Window2 MP-MP	Window3 MP-MP	Mean MP-MP for trace	Mean MP-MP for normal trace
Mawi 2004.03.06 tcp	210,34	172,58	239,41	245,01	240,00
Mawi 2004.03.13 tcp	280,01	214,01	215,46	236,33	240,00
Mawi 20.03.2004 (attacked: worm Witty)	**322,56**	**365,24**	**351,66**	346,48	240,00
Mawi 25.03.2004 (attacked: worm Slammer)	**329,17**	**485,34**	**385,50**	400,00	240,00

Table 2
Matching pursuit mean projection for UDP trace [12]. Traces are analyzed with the use of 20 minutes windows

UDP trace	Window1 MP-MP	Window2 MP-MP	Window3 MP-MP	Mean MP-MP for trace	Mean MP-MP for normal trace
Mawi 2004.03.06 tcp	16,06	13,80	17,11	15,65	16,94
Mawi 2004.03.13 tcp	20,28	17,04	17,40	18,24	16,94
Mawi 20.03.2004 (attacked: worm Witty)	**38,12**	**75,43**	**61,78**	58,44	16,94
Mawi 25.03.2004 (attacked: worm Slammer)	**56,13**	**51,75**	**38,93**	48,93	16,94

Table 3
Matching pursuit mean projection for TCP trace [13] (traces consist of DDoS SynFlood attacks). Traces are analyzed with the use of 20 minutes windows

TCP trace	Window1 MP-MP	Window2 MP-MP	Window3 MP-MP	Mean MP-MP for trace	Mean MP-MP for normal trace
One hour trace from unina1	**1211**	**3271**	**3007**	**2496,333**	860,00
One hour trace from unina2	**1906**	**1804**	**1251**	**1653,667**	860,00

Table 4
Matching pursuit mean projection for TCP trace [14] (traces consist of DDoS attacks). Traces are analyzed with the use of 20 minutes windows

TCP trace	Window1 MP-MP	Window2 MP-MP	Window3 MP-MP	Mean MP-MP for trace	Mean MP-MP for normal trace
Backscatter 2008.11.15	147,64	**411,78**	**356,65**	305,35	153,66
Backscatter 2008.08.20	**208,40**	161,28	153,47	174,38	153,66

traffic features. The effectiveness of the proposed approach has been proved in attack and anomaly detection scenarios. Our framework can be applied to enhance military networks since it uses signal-based features. Such features can be calculated for encrypted traffic since flow characteristics are extracted without considering the payload. Future work focuses on algorithms optimization so that our framework can be applied to real-time network security enhancement and to protect federated network systems (e.g., in national project SOPAS).

Acknowledgment

Within the project supported by Polish Ministry of Science and Higher Education funds allocated for the years 2010–2012 (Research Project number OR00012511).

References

[1] M. Esposito, C. Mazzariello, F. Oliviero, S. Romano, and C. Sansone, “Real time detection of novel attacks by means of data mining techniques”, *Enterprise Information Systems*, VII 2006, Part 3, pp. 197–204.

[2] M. Esposito, C. Mazzariello, F. Oliviero, S. Romano, and C. Sansone, "Evaluating pattern recognition techniques in intrusion detection systems", in *Proc. 5th Int. Worksh. Pattern Recogn. Inf. Sys. PRIS 2005*, Miami, USA, 2005, pp. 144–153.

[3] C.-M. Cheng, H. T. Kung, , K.-S. Tan, "Use of spectral analysis in defense against DoS attacks", in *Proc. IEEE Glob. Telecommun. Conf. GLOBECOM'02*, Taipei, Taiwan, 2002, vol. 3, pp. 2143–2148.

[4] P. Barford, J. Kline, D. Plonka, and A. Ron, "A signal analysis of network track anomalies", in *Proc. Internet Measur. Worksh. ACM SIGCOMM 2002*, Pittsburg, USA, 2002.

[5] P. Huang, A. Feldmann, and W. Willinger, "A non-intrusive, wavelet-based approach to detecting network performance problems", in *Proc. Internet Measur. Worksh. ACM SIGCOMM 2001*, San Diego, USA, 2001.

[6] L. Li and G. Lee, "DDoS attack detection and wavelets" in *Proc. 12th Int. Conf. Comp. Commun. Netw. ICCCN'03*, Dallas, USA, 2003, pp. 421–427.

[7] A. Dainotti, A. Pescape, and G. Ventre, "NIS04-1: wavelet-based detection of DoS attacks", in *Proc. IEEE Glob. Telecommun. Conf. GLOBECOM'06*, San Francisco, USA, 2006, pp. 1–6.

[8] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries", *IEEE Trans. Sig. Process.*, vol. 41, no. 12, pp. 3397–3415, 1993.

[9] P. Jost, P. Vandergheynst, and P. Frossard, " Tree-based pursuit: algorithm and properties", *IEEE Trans. Sig. Process.*, vol. 54, no. 12, pp. 4685–4697, 2006.

[10] J. A. Tropp, "Greed is good: algorithmic results for sparse approximation", *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[11] R. Gribonval, "Fast matching pursuit with a multiscale dictionary of Gaussian chirps", *IEEE Trans. Sig. Process.*, vol. 49, no. 5, pp. 994–1001, 2001.

[12] "WIDE Project: MAWI Working Group Traffic Archive" [Online]. Available: http://tracer.csl.sony.co.jp/mawi/

[13] "Network Tools and Traffic Traces", Universita' degli Studi di Napoli "Federico II" [Online]. Available: http://www.grid.unina.it/Traffic/Traces/ttraces.php

[14] C. Shanon and D. Moore, "The CAIDA Dataset on the Witty Worm", March 19–24, 2004 [Online]. Available: http://www.caida.org/passive/witty

Rafał Renk – responsible for ITTI business development and management of key projects. Since 2004 professional associated with the Adam Mickiewicz University in Poznań as a researcher and lecturer in the field of software engineering and programming languages, applications in telecommunications. He also run courses covering the elements of a non-technical work in computer science. For over 10 years executes and manages number of projects addressing topic of telecommunication networks in technical and business aspect, issues of IT systems associated with the construction of new systems, systems of distance education, data warehousing, ERP, organization security and aspects of crisis management. He participated in several international projects, including these under the 5th, 6th and 7th EC Framework Program, the PASR, Phare, Leonardo da Vinci, Force Protection. He is certificated as Lead Auditor of lead information security management system according to BS 7799. He is the author or coauthor of numerous publications and presentations at national and international conferences.

e-mail: rafal.renk@itti.com.pl
ITTI Ltd.
Rubież st 46
61-612 Poznań, Poland

Division of Applied Informatics
Faculty of Physics
Adam Mickiewicz University
Umultowska st 85
61-614 Poznań, Poland

Witold Hołubowicz, Ph.D., is a graduate of the Poznań Technical University, Electrical Engineering Faculty. His professional career is split into two areas: the academic world and consulting. He was a professor in the area of telecommunications at several universities: Poznań University of Technology, Polytechnic University in New York, Franco-Polish School of New Information and Communication Technologies (EFP) in Poznań and most recently, since 2003 at the Adam Mickiewicz University in Poznań, where he is currently the Head of the Division of Applied Informatics. Throughout all of his professional career, he has been an active consultant. He participated in or coordinated more than sixty consulting, implementation and research projects on radio communications, teleinformatics and telecommunication systems, including numerous international projects. He is the president and cofounder of ITTI Ltd.. It is an independent company started in 1996, which carries out projects, both applied research and consulting, in the area of IT and telecom.

e-mail: witold.holubowicz@itti.com.pl
ITTI Ltd.
Rubież st 46
61-612 Poznań, Poland

Division of Applied Informatics
Faculty of Physics
Adam Mickiewicz University
Umultowska st 85
61-614 Poznań, Poland

Paper

Tunneling Activities Detection Using Machine Learning Techniques

Fabien Allard, Renaud Dubois, Paul Gompel, and Mathieu Morel

Thales Communications, Colombes Cedex, France

Abstract—Tunnel establishment, like HTTPS tunnel or related ones, between a computer protected by a security gateway and a remote server located outside the protected network is the most effective way to bypass the network security policy. Indeed, a permitted protocol can be used to embed a forbidden one until the remote server. Therefore, if the resulting information flow is ciphered, security standard tools such as application level gateways (ALG), firewalls, intrusion detection system (IDS), do not detect this violation. In this paper, we describe a statistical analysis of ciphered flows that allows detection of the carried inner protocol. Regarding the deployed security policy, this technology could be added in security tools to detect forbidden protocols usages. In the defence domain, this technology could help preventing information leaks through side channels. At the end of this article, we present a tunnel detection tool architecture and the results obtained with our approach on a public database containing real data flows.

Keywords—cyberdefense, network security, decision trees, hidden Markov models, HTTPS tunnel, RandomForest.

1. Introduction

Controlling flows going through network boundaries is a key point of information systems security. The filtering of these flows and the verification of their conformance to the network security policy is done in security gateways by application level gateways (ALG) and firewalls. In particular, these tools enforce the restrictions on forbidden protocols over the network. This task is achieved by packets filtering techniques and deep inspection of carried payloads.

Nonetheless, firewalls and ALG may become completely ineffective in two cases: if a permitted protocol is used to embed a forbidden one or if the flow is ciphered. This enables a legitimate or malicious user to infringe the security policy of an information network, using covert application-layer tunnels to bypass security gateways (Fig. 1).

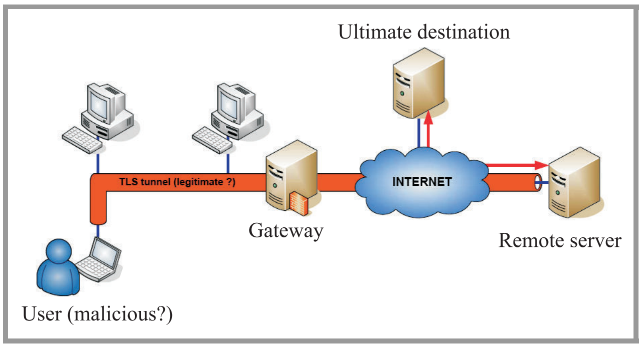


Fig. 1. High level scheme of a TLS tunnel.

Tunneling tools such as `HTTPHost` [1] or `STunnel` [2] are easily available on the Internet, and may be used by a legitimate user to establish a forbidden connection with an external Internet server. These connections consist in a protocol usually filtered by the gateway (e.g., ICQ, FTP, SSH, Skype, Gnutella, BitTorrent, etc.) embedded in a hypertext transfer protocol (HTTP) or hypertext transfer protocol secure (HTTPS) connection. The resulting data exchange is not controlled by the security gateway and may lead to critical information leaks or malware intrusions. For example, an invited participant to a meeting on a military vessel may use a hidden tunnel to leak out classified information via a VoIP protocol. Moreover, similar hidden tunnels are used by attackers on the Internet to communicate with local hosts that have previously been infected by a backdoor.

In this paper, we propose a solution to this problem based on machine learning techniques. Our system relies on a statistical analysis of ciphered flows enabling identification of the carried inner protocol, and therefore, detection of tunneling activities. This solution consists in computing features for each flow and comparing these parameters to a statistical model previously built. The parameters used are derived from the size and the inter-arrival delays of the packets in the flow.

2. Related Work

Many flow level classifiers have been presented in former works and applied to protocol identification [3], [4], [5], [6]. These studies use different parameters and machine learning techniques (Bayesian methods, support vector machine, etc.) to classify the flows into several categories (SSH, HTTP, P2P, GAMES, etc.), with promising results. However, none of these studies specifically address the security issues. Therefore, they use parameters easily tampered with by an attacker, such as port numbers or transmission control protocol (TCP) flags.

To our knowledge, the methods presented in [7] and [8] are the only ones that share our goal to classify encrypted or encapsulated traffic. Nonetheless, both of these works use only the first packets of a connection to classify the entire flow. Thus, by simulating a legitimate flow using only the first packets, an attacker can easily bypass these systems.

Considering the security approach specifically, i.e., tunnels detection, we describe a classification method based on a decision trees forest. This method leads to better results than other machine learning algorithms. A study dealing with the impact of transport layer security (TLS) encapsulation on flows features used for classification is also presented. Then, we present a tunnel detection tool architecture and the classification results obtained with our approach. Finally, we propose a means to decrease the false positive rate.

3. Machine Learning Techniques Applied to Tunnel Detection

Many different machine learning tools have been applied to the flow classification problem. A machine learning algorithm is used to classify a vector among several predetermined classes. It consists in two phases:

- A learning phase, taking as input a set of vectors for each class and returning a classifying model. During this phase, the class of each vector is known.
- A challenge phase taking as input a set of vectors, each belonging to a hidden class, the model and returning the class of each vector.

In our case, the classes are the protocols (HTTP, etc.), and the vectors are the flows (TCP, etc.) over the gateway.

However, related studies were conducted on different databases, with different parameters, and results cannot be compared from one paper to another. An interesting qualitative survey of several methods is presented in [6], but no quantitative comparison is carried out.

In order to determine the most effective algorithm and the best parameters to use for classification, we conducted several experiments on a public database described in [9] and [10]. This database is composed of more than 20,000 flows captured on a real network. The distribution of the database flows by traffic classes are presented in Table 1.

Table 1
Distribution of the database flows by traffic classes

HTTP	Mail (POP, SMTP, ...) IMAP	FTP	Attack	Peer-to-peer	Multimedia (WM player, real player, ...)	Services (X11, DNS NTP, ...)	Interactive (SSH, Telnet)
5707	3519	3107	1822	5717	649	2150	283

First, we selected the parameters that will be used to build statistical models. In order to classify the ciphered or encapsulated flows, these parameters must not be related to the packets payload. We thus kept only the parameters calculated from the sizes of exchanged packets and the inter-packets delays. In order to select the most discriminating ones, a correlation based feature selection with BestFirst search was applied, as described in [11]. A subset of 10 parameters was determined by this means:

- the number of transmitted packets, client to server direction,
- the number of transmitted bytes, client to server direction,
- the IP packets mean size, client to server direction,
- the IP packets maximum size, client to server direction,
- the minimum inter-arrival delay between two IP packets, client to server direction,
- the maximum inter-arrival delay between two IP packets, client to server direction,
- the number of transmitted bytes, server to client direction,
- the maximum IP packets size, server to client direction,
- the variance of the IP packets size, server to client direction,
- the number of uploaded bytes/total number of exchanged bytes' ratio.

Afterwards, we applied six different machine learning algorithms to the database, using a cross-correlation method to classify the entire database. These methods are: support vector machine (SVM), Gaussian mixture model (GMM), K-Means, naïve Bayes method, C4.5 decision tree and RandomForest (a forest of random decision trees). For each algorithm, several criterions were evaluated, such as correct classification rate, false positive rate, computation time, etc. Figure 2 shows the correct classification rates obtained for each algorithm.

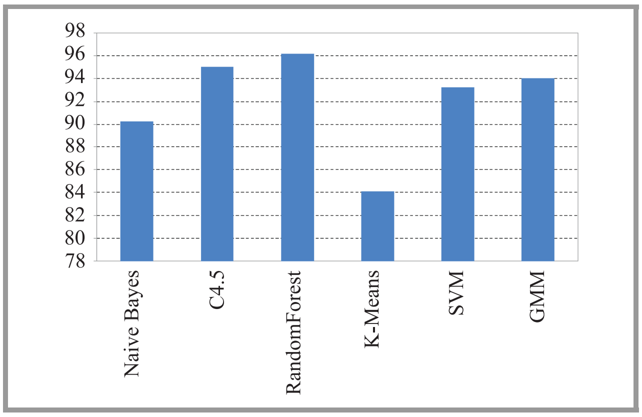


Fig. 2. Correct classification rates for tested machine learning algorithms.

It appears that RandomForest, a machine learning tool never applied before to flow classification, leads to the best performances in terms of correct classification rate and computation time.

4. Impact of TLS Encryption on Classification Parameters

Previous experiments were carried out on a database made of clear flows. Unfortunately, there is no publicly available payload trace set composed of ciphered flows. Our work aims at demonstrating the feasibility of tunnel detection for ciphered flows, and thus it is necessary to prove that results similar to those mentioned above would be obtained on ciphered flows. We conducted a complementary study to evaluate the impact of encapsulation on classification parameters. In particular, we studied the effect of TLS encryption on the set of 10 parameters we use to classify a flow (note that TLS encryption is used to establish an HTTPs tunnel) following these steps:

- pairs of clear/ciphered flows and extracted are generated for different protocols (HTTP, SCP, SSH, etc.),
- the classification features are extracted for each flow,
- an affine transformation function from clear to ciphered was estimated for each parameter,
- the accuracy of these transformation functions was estimated by calculating the residual quadratic error of approximation.

The results obtained showed that the transformation induced by TLS encryption on classification parameters can be correctly approximated by affine functions for 8 features out of 10. On the opposite, two of them (minimum inter-arrival delays between packets from client to server and variance of the size of packets from server to client) were transformed in a more complex way.
We can reasonably conclude from this results that TLS encryption will not lead to a significant loss of performance for the classification algorithm mentioned above.

5. A Tunnel Detection Tool Architecture

The biggest drawback of statistical methods is their high rate of false positive (i.e., legitimate flows classified as malicious). We propose a specific tunnel detection tool architecture designed to lower the false positive rate. Figure 3 describes this architecture.
The system consists in a network capture tool (such as `TCPDump` [12]) combined with a flow demultiplexer. Classification features are then extracted from each flow, and a RandomForest model is used to determine the class of each connection. In order to minimize false positive cases due to errors of classification, a set of heuristic rules is applied to generate an analysis report composed of a list of alerts. These rules take into account past results of classification, and a level of confidence for each classification. No alert is raised if the confidence level is too low, if the IP address of the local or remote host is on a white list, etc.

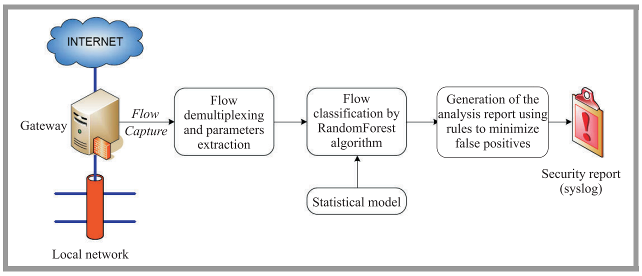

INTERNET
Gateway
Flow
Capture
Flow demultiplexing and parameters extraction
Flow classification by RandomForest algorithm
Generation of the analysis report using rules to minimize false positives
Security report (syslog)
Statistical model
Local network

Fig. 3. High level tunnel detection tool architecture.

The analysis report generated by the application of this set of rules could have the syslog format, for future integration in a complex intrusion detection system.
The proposed architecture was implemented on an experimental platform and give very encouraging qualitative results. These results are presented in the next section.

6. Qualitative Results of the Proposed Solution

6.1. Network Simulation

At first, we implemented our detection tool on a network simulator. The simulator consisted in 3 machines, simulating respectively the local network, the gateway and the Internet. This simulator has been used to measure the TLS impact (Section 4) and the efficiency of the detection tool. The resulting detection rates for the protocols shown in Table 2 are close to 100%. However, this did not provide a convincing proof because the diversity of the flows is reduced compared to a real network:

- the topology of the network is too simple,
- the behavior of the user is unique,
- the material is also unique (one OS, one hardware, etc.).

Table 2
Distribution of the database flows according to the protocols

HTTP	HTTPs	SSH	SMTP	DNS (over TCP)	FTP	Active directory	POP3s	NetSteward
2500	2500	2500	2500	2500	2500	1069	1503	1611

The results obtained for the TLS impact remain valid, but in order to evaluate the accuracy of the tool, a more complex set of flows had to be tested.

6.2. A Flows Database in Order to Evaluate Our Detection Tool

The public database containing real data flows used for our experimentations is provided by the MAWI working

Table 3
Confusion matrix obtained using the RandomForest method to classify the database

HTTP	HTTPs	SSH	SMTP	DNS	FTP	Active directory	POP3s	NetSteward	Protocols
93.08	4.36	0.0	1.08	0.04	0.24	0.08	0.36	0.76	HTTP
2.36	**91.56**	0.08	3.2	0.0	0.48	0.48	2.36	0.28	HTTPs
0.0	0.12	**99.44**	0.08	0.0	0.08	0.0	0.28	0.0	SSH
0.96	2.28	0.0	**91.12**	0.2	3.48	1.12	0.72	0.12	SMTP
0.0	0.0	0.0	0.32	**99.64**	0.0	0.04	0.0	0.0	DNS
0.08	0.6	0.0	3.0	0.0	**95.88**	0.2	0.24	0.0	FTP
0.19	0.09	0.0	0.47	0.0	0.0	**99.16**	0.0	0.09	Active directory
0.13	1.2	0.27	0.73	0.0	0.0	0.0	**97.67**	0.0	POP3s
1.37	0.06	0.0	0.0	0.0	0.0	0.0	0.0	**98.57**	NetSteward

group [13]. The database is a recording of the whole set of flows carried by a transpacific 150 Mbit/s network line between Japan and USA, during 96 hours. The payloads have been removed and the headers from layers 1 to 4 from the OSI model have been anonymised.

In order to illustrate the performance of our solution, we classified nine kind of network flows. For each protocol, the number of flows contained in the database and used for the experimentation is shown in the Table 2.

Note that the flows used for the experimentation are mostly clear flows, i.e., unciphered flows. Indeed, there is unfortunately no public database of ciphered flows precising for each flow which protocol is ciphered. Nevertheless, our analysis with this database is interesting and can be extended to ciphered flows for the following reasons:

- the flow classification features can be calculated with ciphered flows exactly as for the clear flows,
- the impact of ciphering on the parameters is limited.

Parameters like the delay induced by the user behavior (as the password capture for a secure shell (SSH) session or the frequency of HTTP request while surfing) are not affected by the encryption.

6.3. Classification Results

Table 3 shows the corresponding confusion matrix obtained with this algorithm. The procedure used to get the confusion matrix is:

1. For each flow, compute the features regarding the full connection.
2. Train the classifying model (i.e., RandomForest) on a subset (the learning set) of flows.
3. Challenge the model on the remaining vectors (the challenge set.
4. Report the results.

For example in this table:

- the number 93.08 in the first row indicates that 93.08% of HTTP flows have been correctly classified as HTTP,
- the number 4.36 in the first row indicates that 4.36% of HTTP flows have been erroneously classified as HTTPs.

Therefore, the correct classification rates are on the table's diagonal. The average rate of correct classification is 95.81%.

In a standard configuration, the only allowed protocol might be HTTP and HTTPs. Any flow classified in an other class (e.g., SSH, POP3, ...) would then be considered as malicious. Hence, if we set this configuration, the tool detects 98.68% of illegitimate flows (corresponding to 1.32% of false negatives) with 4.72% of false positives (i.e., false alarms). This last rate is too high for an actual use, since most of flows are legitimate. In Subsection 6.7 we propose a way to decrease the number of false alarms sent by the tunnel detection tool.

6.4. Classification Computation Time

As shown in Table 4, the classification computation time is quite short. The implementation has been realized on a 3.06 GHz PC platform running under a Debian distribution. The langage used is Java, therefore this computation time could be reduced using a faster langage such as C if needed.

Table 4
Computation time with a 2500 flows database

Phase	Time
Learning phase	1143 ms
Challenge phase	223 μs

6.5. Impact of the Flows Length

The procedure described in Subsection 6.3 works with a full connection. Thus, it does not allow the gateway to take a real time decision such as ending a session as soon as an illegitimate flow is detected (the decision is a posteriori). In order to take a proactive decision, a small number of packets can be used rather than the full connection. As a consequence, it increases dramatically the computation power required by the security gateway. Our study showed that the decision can be taken with only very few packets (about 3 packets). This could be explained by the fact that the considered protocols have different behaviors from the beginning of the connection, which helps to distinguish them with a small number of packets.

6.6. Impact of the Database Size

Another issue is the size of the learning database. Depending on the context, it may be hard to generate a large database for each flow. For example, the database built with our simulator had to be manually filled. Figure 4 illustrates the impact of the database size on the detection accuracy.

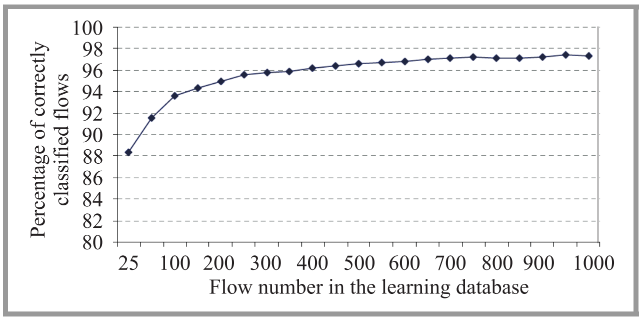


Fig. 4. Impact of the database size on the detection accuracy.

6.7. A Simple Method to Lower the False Positive Rate

We saw in Subsection 6.7 that the false positives rate (i.e., legitimate flows classified as malicious) is too high for an actual use while the illegitimate flows rate is, on the opposite, very good. Depending on the use case, it could be better to limit the number of false positives, because it could disturb most of the network users.

For this reason, we propose to set a confidence indicator. Therefore, a flow with a confidence indicator below a specific threshold will be automatically considered as legitimate. This rule can be added in the heuristic part of the tunnel detection tool architecture (Fig. 3).

Figure 5 shows the rates of false positives and false negatives obtained by applying this simple heuristic, based on the confidence indicator set. We can see that such a rule can reduce the false positives rate. However, this method seems too 'naive', because the increase of false negatives rate (i.e., illegitimate flows allowed by the security gateway) is significantly faster than the decrease of false positives rate.

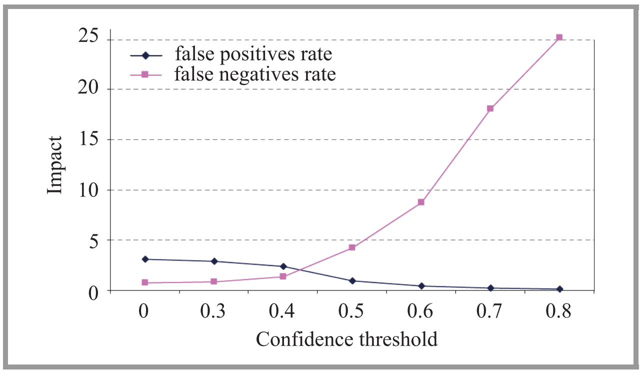


Fig. 5. Impact of a rule based on a confidence indicator on the rates of false positives and false negatives.

7. Conclusion

In this paper, we presented a solution to the key problem of encapsulated illegitimate flows detection across network boundaries. In a first part, we compared the performances of different machine learning algorithms and identified the best one in our specific case. In a second part, we conducted a complementary study showing that the effect of TLS encryption on classification features should not significantly affect classification performances. Finally, in a last part, we described a high-level tunnel detection tool architecture. We pointed out qualitative results using this tool with a public database and the impact of variation around the protocol on its accuracy. Finally we proposed, regarding the results obtained, a simple method to lower the false positive rate.

The construction of our solution is generic and can be tuned to be used for automatic classification, pro-active reaction or small learning database. In a global cyberdefense system, the proposed architecture could be efficiently used with a classical security tool, such as an IDS, in order to improve the security level.

References

[1] HTTPHost [Online]. Available: http://www.htthost.com

[2] STunnel [Online]. Available: http://www.stunnel.org

[3] J. Erman, A. Mahanti, and M. Arlitt, "Internet traffic identification using machine learning", in *Proc. IEEE GLOBECOM'06*, San Francisco, USA, 2006.

[4] T. Karagiannis, K. Papagiannaki, and M. Faloutsos, "BLINC: Multilevel traffic classification in the dark", in *Proc. ACM SIGCOMM'05*, Philadelphia, USA, 2005.

[5] A. W. Moore and D. Zuev, "Internet traffic classification using bayesian analysis techniques", in *Proc. ACM SIGMETRICS'05*, Bauff, Canada, 2005.

[6] T. T. Nguyen and G. Armitage, "A survey of techniques for Internet traffic classification using machine learning", *IEEE Communications Surveys and Tutorials*, vol. 10, no. 4, pp. 56–76, 2008 [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=04738466

[7] L. Bernaille and R. Teixeira, "Early recognition of encrypted applications", in *Proc. PAM 2007*, Louvain-la-neuve, Belgium, 2007.

[8] M. Dusi, M. Crotti, F. Gringoli, and L. Salgarelli, "Detection of encrypted tunnels across networks boundaries", in *Proc. IEEE ICC'08*, Beijing, China, 2008.

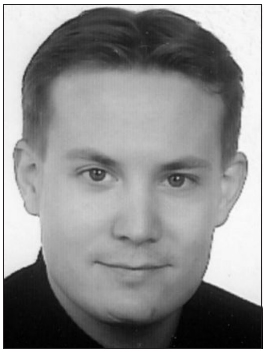
[9] A. W. Moore, D. Zuev, and M. L. Crogan, "Discriminators for use in flow-based classification", Techn. Rep., 2008.

[10] W. Li, M. Canini, A. W. Moore, and R. Bolla, "Efficient application identification and the temporal and spatial stability of classification schema", *Comp. Netwo.*, vol. 53, no. 6, 2009.

[11] N. Williams, S. Zander, and G. Armitage, "A preliminary performance comparison of five machine learning algorithms for practical IP traffic flow classification", in *Proc. ACM SIGICOMM'06*, Pisa, Italy, 2006.

[12] Tcpdump/Libpcap [Online]. Available: http://www.tcpdump.org

[13] MAWI Working group traffic archive [Online]. Available: http://mawi.wide.ad.jp/mawi/



Fabien Allard graduated in 2004 with an M.Sc. in computer sciences at University of Rennes (France), and in 2005 with a specialized M.Sc. in telecommunications at Telecom Bretagne. In October 2005, he joined Orange Labs in order to do a Ph.D. His research focuses on the context transfer mechanism and his application for security protocols. Then, in 2009 he joined Thales Communications in an IT security dedicated team and currently works as technical expert and architect for national defence projects.
e-mail: fabien.allard@fr.thalesgroup.com
Thales Communications
160 Boulevard de Valmy – BP 82
92704 Colombes Cedex, France

Renaud Dubois graduated in 2003 with an M.Sc. in applied mathematics at University Pierre et Marie Curie (Paris VI), and in 2004 with a specialized M.Sc. in cryptography at University Bordeaux I. He joined the cryptographic lab of Thales Communications in 2005 after an internship. He works as a cryptographic expert for national defence projects and lead cryptographic R&D studies.
e-mail: renaud.dubois@fr.thalesgroup.com
Thales Communications
160 Boulevard de Valmy – BP 82
92704 Colombes Cedex, France

Paul Gompel got his M.Sc. in mathematics and computer science at University Paris IX – Dauphine and M.Sc. in telecommunication at Telecom Bretagne in 2005. Then he joined Thales Communications in an IT security dedicated team for Defense activities. He acted until December 2009 as architect, expert and technical leader in security activities, for both national and NATO major projects. He also led different technical studies and workgroups, including R&D in covert channels detection.
e-mail: pgompel@gmail.com
Thales Communications
160 Boulevard de Valmy – BP 82
92704 Colombes Cedex, France

Mathieu Morel graduated in 2010 with a multidisciplinary M.Sc. from the Ecole Polytechnique and a additionnal M.Sc. in computer sciences at Telecom-ParisTech. He concluded his degree with a seven months internship at Thales Communications, focusing on covert channels detection using statistical techniques. Since September 2011, he works for the French Home Office, leading a team of computing specialized engineers with a series of projects.
e-mail: mathieu.c.morel@gmail.com
Thales Communications
160 Boulevard de Valmy – BP 82
92704 Colombes Cedex, France

Paper

Success Factors for SOA Implementation in Network Centric Environment

Joanna Śliwa[a] and Marek Amanowicz[a,b]

[a] Military Communication Institute, Zegrze, Poland
[b] Military University of Technology, Warsaw, Poland

Abstract—This paper discusses challenges and success factors for service oriented architecture (SOA) implementation in network centric environment. The authors identify 9 fundamental challenges for the SOA approach in order to make the biggest benefit for the NATO NEC (NNEC) and increase the mission effectiveness to the highest extent. They cover the areas of applicability to existing military communications and the ability to reflect military processes. Their range is quite broad, pointing out technological as well as SOA governmental problems. The authors emphasize that any COTS solution available on the market today is able to overcome all of them at once. However, they propose solutions to some of the problems and present quick wins that can speed up the process of achieving capabilities in a heterogeneous multinational NEC environment.

Keywords—NEC, SOA challanges, SOA success factors, tactical networks.

1. Introduction

Modern coalition operations are conducted in a dynamic environment, usually with unanticipated partners and irregular adversaries. This new situation has forced the NATO Alliance to pursue the achievement of the so-called “NATO network enabled capabilities” (NNEC) concept, which is the “ability to collect, fuse and analyze relevant information in near real time so as to allow rapid decision making and the rapid delivery of the most desired effect”[1]. The main tenet of the net-centricity is to achieve information superiority by sharing reliable information collected from various sources, creating situational awareness and distributing it among mission participants, across domain, context and organizational boundaries on the basis of extended collaboration.

The NNEC concept, followed in the “NNEC Data Strategy” [2] emphasizes two primary objectives in this process, i.e., the necessity to increase the data that is available to communities in the network-enabled environment; and to ensure that the data is usable by authorized anticipated and unanticipated users and applications. In order to accomplish this goal, the change of focus must take place: from the idea of standalone, stovepiped systems (i.e., platform-oriented) to the idea of shareable, universal information. This would allow unanticipated (but authorized) users to discover information, as opposed to being pushed to them via a pre-defined mechanism.

The improvement of collaboration and information sharing in a highly dynamic, unpredictable NEC environment is a great challenge. It assumes transfer of information with a required quality of service and security, independently of the underlying infrastructure as well as a common access to relevant information by the authorized users. These requirements are to be satisfied by the use of service oriented architecture (SOA), that succeeded in commercial world lately and is recommended by NATO as the crucial NEC enabler [3]–[5]. SOA can make military information resources available in the form of services that can be discovered and used by all mission participants that do not need to be aware of these services in advance.

2. Service Oriented Architecture in the Context of NEC

Service orientation is a conceptual architecture which asymmetrically provides services to arbitrary service consumers, facilitating the information sharing in a heterogeneous environment, and thus supports to some degree the open-ended aspects of net-centricity.

By OASIS definition [6], service oriented architecture (SOA) is “as an architectural paradigm for organizing and utilizing distributed capabilities that may be under the control of different ownership domains”. It provides a “uniform means to offer, discover, interact with and use capabilities (as well the ability to compose new capabilities from existing ones) all in an environment that transcends domains of ownership”.

The use of service oriented architectures has emerged as a major trend within the commercial sector and among nations developing NNEC type capabilities, because of the flexibility they provide in sharing information and information processing capabilities. SOAs provide mechanisms for using existing information services as well as providing the basis for developing new, more advanced information services. Such mechanisms will allow many command and control processes to be satisfied by linking together existing information sources in a modular, flexible fashion that can be readily adapted to changing operational context. The flexibility provided through the use of SOAs is particularly

[1] Citation from [1].

well suited to support the needs of coalition based network-centric operations using systems of various nations, on different levels of transformation, without the need to replace, but only to integrate them into the SOA environment. The concept is that clients see other servers and applications as "services", accessible using a known and common technology, independently of the underlying implementation of the service (platform – independent approach). In fact, this interaction does not have to be between client applications on user terminals and the servers, but can be between peer applications that relate to each other as a client and a service. This approach allows both user-to-system – and direct system-to-system – interactions that have not previously been generally feasible except by system specific and proprietary implementations.

It is also worth noting that, in a true service oriented environment, developers of services do not know who will access their services at run time. There is a great flexibility in how, when and by whom the services will be used. This idea of "unanticipated users" is a key element of SOA, particularly advantageous in the dynamic NATO environment, where the components of missions change over time and the consumers of services today may not be the consumers of the services tomorrow.

The technological background of NNEC implementation, i.e., networking and information infrastructure (NII) strategy assumes that the NII will be implemented as a federation of systems[2] (FoS), involving the use of SOAs [1], [3], [4], [7]. The NNEC feasibility study [4] has made it clear that the Information and integration services (IIS) layer of NII will be formed by a federation of services, within which any NATO or national information system will be autonomous and will provide specific services by means of implementing a standardized service interface [4].

The most mature implementation of SOAs, recommended by NATO and widely applied in the commercial sector, are web services (WS) and other extensible markup language (XML) technologies. Current trends show great levels of maturity and adoption of these technologies – within the IT industry, within the NATO nations, and within various multinational programs – leading to the belief that this is a direction already being followed and that there is already much force behind it. WSs are described by a wide range of standards that deal with different aspects of WS realization, transport, orchestration, semantics, etc. They provide the means to build a very flexible environment that is able to dynamically link different system components to each other. These XML-based standards have been designed to operate in high bandwidth links. XML gained wide acceptance and became very popular for the reason that it solves many interoperability problems, is human- and machine-friendly and facilitates the development of frameworks for a software integration, independent of the programming language. Nevertheless, it undoubtedly adds significant overhead, both in terms of computation and network resources while being transported.

The value of SOA is though, that it provides a simple scalable paradigm for organizing large systems that require the interoperability to realize the value inherent in the individual components. Moreover, apart from its inherent ability to scale and evolve, the SOA – based infrastructure is also more agile and responsive than the one built on an exponential number of pair-wise interfaces. Therefore, SOA can also provide a solid foundation for developing operational context, based on business agility and adaptability.

The remainder of this paper is organized as described below. Section 2 describes a set of challenges that SOA application in military NEC-centric environment must overcome. They derive both from the architecture itself and from the characteristics of the environment, that must integrate systems owned and governed by different nations/organizations, built by use of different (modern and legacy) technologies, lacking standardization in many areas (e.g., management, cross-domain security, QoS, etc.), and facing disadvantaged communications links. Section 3 presents the SOA success factors that should be taken in order to support dynamic, flexible and scalable SOA – based environment to conduct a net-centric multinational operation. We conclude this paper in Section 4.

3. SOA Challenges

Service oriented architecture, as presented above has a great potential deriving from the paradigm and framework it relates to. It has undoubtedly many benefits, however its

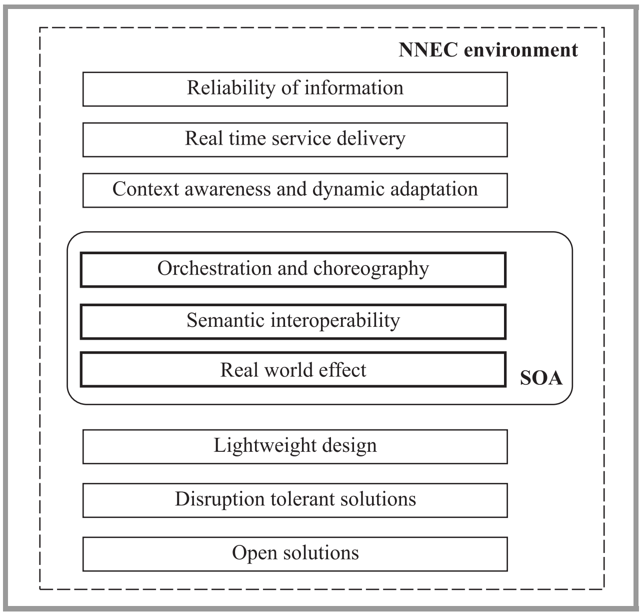

NNEC environment
Reliability of information
Real time service delivery
Context awareness and dynamic adaptation
Orchestration and choreography
Semantic interoperability
Real world effect
SOA
Lightweight design
Disruption tolerant solutions
Open solutions

Fig. 1. SOA challenges deriving from the SOA concept and NNEC environment.

[2] Federation of systems – complex environment built of heterogeneous autonomous systems governed independently, taking advantage of cooperation. "*(...) formed by the synergistic amalgamation of a dynamic set of globally interconnected, multi-national, autonomous systems, each comprised of networking and information infrastructure components, providing information capabilities, associated processes, and personnel for collecting, processing, storing, disseminating and managing information to authorized users on demand, on an end-to-end basis*" [4].

application in a highly dynamic, heterogeneous and disadvantaged NEC environment is a challenge.

This paper identifies 9 fundamental challenges for the SOA approach in order to make the biggest benefit for the NNEC and increase the mission effectiveness to the highest extent (see Fig. 1). They cover the areas of applicability to existing military communications and the ability to reflect military processes. Their range is though, quite broad, pointing out technological as well as SOA government problems. No COTS solution available on the market today is able to overcome all of them at once.

3.1. Real World Effect

A service is a mechanism that enables access to one or more capabilities, based on the use of a predefined interface and exploited consistently with constraints and policies as specified by the service description.

Following OASIS SOA Reference model [6], the main tenets of SOA services are:

- visibility (provided by the service description),
- interaction (described in the service contract and realized between SOA components),
- real world effect.

The first two of them are inherent to the interaction patterns that will exist in the system. However, the last one is strongly related to the business model that exists over the technical model. The design of the SOA services layer must start with identifying business processes that are the heart of the "military enterprise". Only after identifying actors that take part in the net-centric operation and defining the information needs and flows between them, it is possible to create services of a true value to the enterprise.

Business processes are considered to be the basis for developing service oriented architecture [3], [4]. At the national level, it is necessary to initiate activities that are to develop overarching architecture of the C4I on the basis of the service oriented architecture in order to have a SOA – based C4I system that supports creation of the national NNEC. It should integrate new and legacy systems and reflect real user requirements. It is impossible to carry out this process without having previously defined and properly described the operational processes conducted in every moment of a net-centric operation. These actions would lead to development of the service map and service layer based on real information needs.

3.2. Semantic Interoperability

Taking advantage of the wide spectrum of services that will arise in the federation of systems (FoS) built based on SOA, requires that all of the domains connected speak with "the same language" and "understand" each other. Achieving the interoperability in terms of network protocols (e.g., based on TACOMS STANAGS) or data (e.g., based on the NC3A Data Strategy), does not cover all its aspects. It must be ensured that the meaning of terms used by applications in domains owned by different countries, different kinds of military services (land, air, navy) is the same. It is necessary to establish a high degree of correlation and similarity between the details of their respective descriptions and definitions, on which they intend to reach an agreement. This implies that a shared understanding requires shared definitions. It is, though highly recommended to provide ontological models for different domains of the NEC environment within NATO that could be used by every country joining a multinational operation.

The role of ontologies in transformation to NNEC has been emphasized by the NC3A in the NC3A Technical Architecture [7] supplement. It points out that in a dynamic multidomain environment, the goal is to implement semantic web solutions that would enable the users to locate, select, employ, compose and monitor the web services automatically. Ontologies are needed especially in the area of services description, enabling semantic services selection, as well as providing the possibility to employ orchestration and choreography. To make use of a NNEC service, software agents need a computer – interpretable description of that service, and the means by which it is accessed. An important goal for semantic web languages is to establish a framework within which these descriptions are made and shared.

It should be emphasized that for automated tasks of the systems, it is also necessary to provide semantic metadata descriptions needed, e.g., for provision of the quality of service, service level agreement (SLA) support, security, management, etc. What is more, in order to achieve the understanding of data distributed among systems, common data ontologies for domains (e.g., particular communities of interest) are needed.

The application of ontologies and providing foundation for creating the, so called, semantic web, does not restrict itself to the formal semantic description of service resources for machine-to-machine exchange and automated integration and processing. One important feature of formally describing resources is to allow computers to reason by inference. Once the resources are described using facts, associations, and relationships, inference engines, also called reasoners, can derive new knowledge and draw logical conclusions from the existing information.

Sharing information among different systems creates the problem of understanding what the data means. Ontologies and inference services, by expressing the meaning in ontology for the specific data sources and defining the relationships among the different concepts or terms appearing in those apparently different ontologies, are able to perform the data and information integration. Moreover, the application of ontologies provides the possibility of an automated reasoning, semantic search for information, enabling the decision support and advanced searches for information in the SOA environment. It is though, very important to start work on developing domain ontologies enabling to provide such information integration. This may be done on the basis of the JC3IEDM model, developed by the MIP community and applied by many NATO countries to provide automated data exchange and replication. It must be emphasized, however, that JC3 is very broad, and not divided into smaller ontologies, would be inefficient to be processed by devices with limited computational resources.

3.3. Orchestration and Choreography

One of the most important advantages of the service orientation is that, it supports, by the use of service reusability, the automation of processes. The orchestration and choreography makes it possible to create new services and arrange them into process flows on demand. Therefore, in order to take full advantage of what SOA brings, it is necessary to provide a high level management of the business processes mapped into services, enabling a dynamic reaction on new information needs, change into business process to create dynamic service environment that flexibly changes in order to meet the information needs. These functionalities that may be built on top of SOA services are getting more and more popular as the amount of work on building ontologies also increases.

In order to support the realization of operational processes' goals, services form a set that can be orchestrated. These orchestrations may be reconfigured differently, when needed, to regain support of an operational process after its ad hoc arrangement has changed to suit new or changed operational needs.

A service orchestration, in general, refers to an executable business process that may interact with both internal and external services, capable of satisfying certain operational objectives that cannot be achieved by any of the services alone. It requires the various composing systems to collaborate in a controlled (orchestrated) manner. Depending on the purpose, it may not be enough to only determine which services are used. It may also be necessary to resolve timing issues, semantic misunderstandings, and the quality of service discrepancies, which may appear when services interact.

Orchestration leads to the emergence of higher level services, where the combined use of services is to deliver a higher level functionality or effect. In case of web services, this creates a composite web service.

On the other hand, choreography is more collaborative in nature. Each party involved in the process describes the part they play in the interaction. Choreography describes a process flow between services and processes themselves and tracks the sequence of messages that may involve multiple parties and multiple sources.

Orchestration and choreography become more and more popular among solutions for semantic search for services, trying to correlate offered inputs and required outputs. More and more often, the search for services bases also on proprietary quality of service descriptions that makes it possible to provide services meeting user preferences, adapted to the possibilities of the network. The utilization of orchestration and choreography together with ontologies, enables to automate many processes realized during the course of the operation and to provide the possibility to gather all the necessary information for the operators involved.

Furthermore, it is necessary to emphasize that the semantic description used on the daily basis in the software applications need to be tailored to the computational resources of that devices. Web ontology language (OWL), most commonly used to express ontologies, is a very expressive language, and enables to perform reasoning over ontologies and infer knowledge that is not explicitly stated. However, the reasoning engines for OWL typically require a lot of resources, and are, therefore, not well suited for resource-limited handheld devices. It has been stated in [3] that in an ad hoc wireless environment utilization of ontologies is possible only when ontologies are small, so that the handheld devices can process them and reason over. That is why creating ontologies need to be carefully carried out, preferably based on one of the commonly applied methodologies (e.g., methontology), acquiring them to the environment where they will be used.

3.4. Open Solutions

Realizing the benefits of the SOA approach will require agreeing on a standardized set of protocols, data formats and foundational (core) services (e.g., covering such areas as service discovery security, metadata management, identity management, service management and mediation), that provide means to establish the interoperability in the technical field. NII is to be formed based on the Internet model, major advantage of which is the use of common set of protocols enabling to create dispersed, dynamic environment for sharing information without central governing authority [3], [4]. In order to take advantage from the success of the Internet, the utilization of open standards that will help to ensure interoperability is necessary.

Obviously, a military environment differs from the Internet in many aspects (e.g., security constraints, policy), so that it will be impossible to adhere only to commercial standards. However, in order to provide coherent NATO – supported coalition network environment enabling to act efficiently in multinational missions, it is recommended to revise the consequences of using open standards and assess the risk that is related to the increased interoperability problems when using some proprietary solutions. There should be also taken initiatives to standardize the mechanisms and protocols that are of great importance to provide a secure and dynamic service-oriented net-centric environment (e.g., cross-domain security solutions, quality of service principles, common ontologies). Such initiatives are taken, e.g., in NATO (by the working groups and other research initiatives provided by NATO Research and Technology Organization[3]) and by Network Centric Operations

[3] More information on RTO web site, http://www.rta.nato.int

Industry Consortium (NCOIC[4]), that plays an important role in establishing a common view of net-centric capabilities and provides technical solutions to enhance the interoperability in a multinational environment.

It is also very important to emphasize the need to create a common technical framework for developing SOA-based solutions that is based on open or agreed standards. This process is carried out, e.g., by the standardization bodies (e.g., OASIS, W3C), but the military environment often requires additional functionalities, not supported by organizations working mainly for the commercial sector. It has been shown in many multinational exercises (e.g., Combined Endeavour CE, Common Warrior Interoperability Demonstration CWID – currently CWIX, MultiNational Experimentation MNE, etc.), that the interoperability is crucial for the mission success in a NATO community. Systems created for the sole purpose of the country (e.g., crisis management systems) must also interoperate with many national systems. The interoperability is though, necessary on many levels. That is why adherence to the open, agreed standards makes it easier to solve the problem on the technical grounding.

3.5. Disruption Tolerant Solutions

The utilization of SOA-based systems in a NEC environment has been shown in many international experiments [5]. They prove that SOA technologies improve the collaboration, interoperation and information sharing in complex environment of heterogeneous systems. However, the NATO concept of FoS relying on the exchange of information implies that communications are of critical importance to the entire (C2) system. In order to achieve an efficient information exchange between the users, the SOA solutions need to work with different types of information and communication systems. The challenge is though, to use this – simple in concept and providing a big flexibility – means of communications on every echelon of command – from the strategic and operational to the tactical and individual soldier's level.

Going down in the command structure, the network environment is getting more and more degraded providing worse and worse communications, e.g. low bandwidth, high level of unpredictability of quality factors, lack of connectivity guarantee, changing topology, radio silence and high error rates. The fragile nature of tactical radio networks requires robust communication mechanisms which current SOA solutions (e.g., mainly enterprise SOA implementations, such as Enterprise Service Buses (ESBs)) do not provide. This includes methods to deal with large delays, communication failures, network splits and merges. It is necessary to provide SOA solutions that enable to integrate network elements on all command levels and a smart communication infrastructure which would deal with peculiarities of the wireless medium. This will allow to create the situational awareness also on the lowest command levels and supply users of the tactical systems with necessary information and decision support.

3.6. Lightweight Design

In order to adhere to the open standards tenet and take advantage of the most common WS-based realization of SOA, it must be ensured that the solutions implemented, originally derived from the commercial world, do not overload the network nor overuse the existing resources (network and terminal ones). SOA solutions, very often based on XML, known from significant overhead it adds, will have to be effective in the whole NEC environment. They should have though, a lightweight design that minimizes the demand on the network and implements a minimum range of functionality.

The requirement for minimum functionality is the most important in the tactical domain, for nodes with a limited memory, storage capacity, processing power and battery life. Such nodes are usually man-portable nodes or sensor nodes powered by battery, where the minimum functionality can also serve the important goal of energy conservation. However, the use of complicated, "heavy" software components on low command levels should be avoided.

3.7. Real Time Service Delivery

The most obvious tenet of net-centricity is the information exchange (see Fig. 2). This capability should, however, be seen not only in terms of what derives from the usual relation between the commander and the subordinate. Looking down to the battalion, platoon and lower levels, apart from regular exchange of formalized messages, operators more and more frequently exchange information horizontally, sharing information within small formations (platoon or squad), e.g., positions, alarms, video streams, pictures and other important elements of building situational awareness at the tactical level. Individual soldiers are often equipped with high-quality sensors and hardware that make them complex technical systems on their own. Each actor engaged in an operation can thus be considered both a consumer and a provider of data. Down the echelons of command, actions are getting more and more dynamic, decision-making time is shorter, so that actors need real-time information. Moreover, the information granularity, low in the military hierarchy, is greater, which means that military staff needs detailed information about their area of operation.

SOA solutions must be applicable to the tactical domain providing the possibility to create common operational picture (COP) on various levels of command. COP created at high command levels, tailored to the needs of the operators and giving them the overview of coalition, neutral and enemy forces, enabling to plan and conduct operation in real time, should be also achieved at the lowest levels including the individual soldier. This would support creating shared situational awareness and enable military staff to make reliable decisions in a short timeframe, work to-

[4] More information on NCOIC web site, https://www.ncoic.org/home/

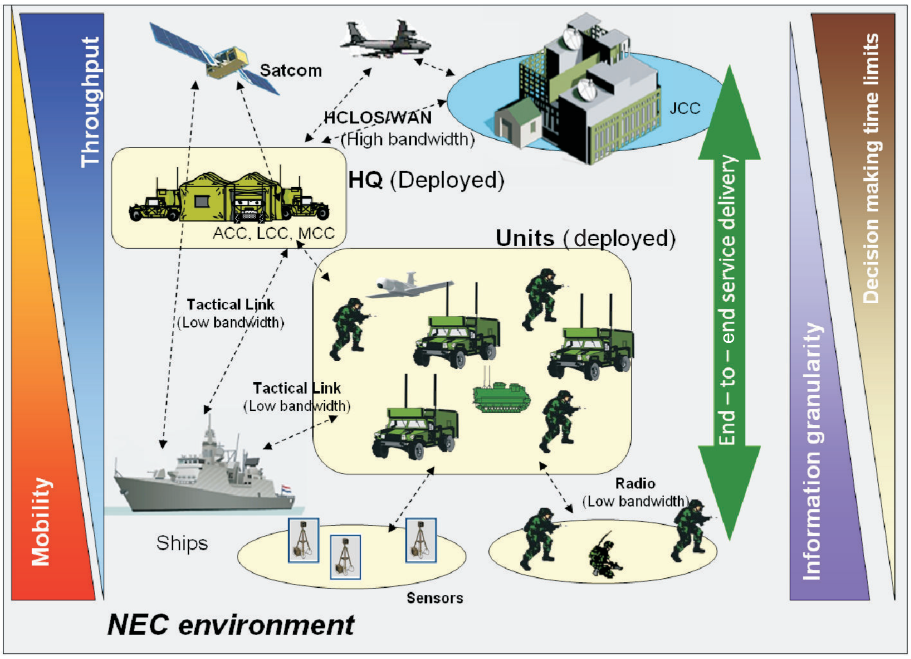


Fig. 2. SOA application in NEC environment on different command levels.

gether in new, more effective ways and thus, to improve the speed of command, leading to a dramatic increase of mission effectiveness.

It is though necessary to provide technical means to realize services in real time. This includes messaging mechanisms that do not necessarily base on web services often considered not well applicable for real time requirements but provide good efficiency in disadvantaged networks and enable the realization of time-constrained services.

3.8. Context Awareness and Dynamic Adaptation

SOA, as an architecture covering the whole FoS will be used (obviously depending on the level of transformation particular systems acquire) to integrate all connected systems. In order to adhere to different types of communication networks and meet the requirements of users on every command level, SOA solutions must be “aware”of the context of the service call. Middleware layer of the SOA should though, maintain a shared perception of the network state. In order to do so, it should serve as a mediator for collecting, organizing, and disseminating relevant context information to the upper layers (application) and lower layers (transport mechanisms). The context may be also used for selecting the best service realization mode and can include device, network characteristics, service and user activities and requirements. It can characterize static elements that can be defined before a service call is executed, as well as dynamic elements that depend on the type of request, type of response and temporary network QoS parameters.

The solution to this problem, related to providing interoperation with other layers of the architecture, should enable realizing SOA services in a constrained environment and facing another challenge, which is to provide the possibility to adapt to environment limitations.

There are no standard solutions that provide the possibility for SOA to adapt to limitations of the underlying network. However, this problem is crucial in terms of SOA success in a military environment. In order for the services' layer to be responsive and provide dynamic access to information for the users that change their location, privileges and activity during the course of the operation, SOA solutions must provide end-to end service delivery in the horizontal and vertical dimension of the NEC environment. This derives the necessity of dynamic service discovery based also on the user/service requirements, QoS needs and geo-location, as well as the provision of technical solutions that enable SOA to adapt to current network situation.

SOA middleware should have the ability to recognize changes to its execution context and to adapt its behavior appropriately. An example would be:

- an adjustment of services provided to an application (e.g., reduction in frame rates for a real-time video streaming application), based on middleware awareness of reduced network throughput,

- providing the possibility to realize the service in a different pattern (e.g., peer-to-per instead of client-server in case of central server unavailability), and deliver messages when network is degraded, or

- searching for services in the surrounding of the user in case of service registry unavailability.

3.9. Reliability of Information

Mission – critical systems that support NEC operations these days are to deliver appropriate, fused and processed information in the right time. They assume that the infrastructure enables to appropriately deliver binary data. The problem SOA brings in terms of QoS is related not only to the quality of service that covers the mechanisms on the physical, network and data/object levels, but to the reliability of information provision that relates to the reliability of information sources, trustworthiness of data from that sources, support for using the information and supporting the realization of procedures, processes, strategy and doctrine, quality of information and quality of operations as well.

In terms of security, it is necessary to provide appropriate identity management together with the common understanding of privileges, access rights, as well as an agreed security policy. The solutions must enable to identify the users from different domains, appropriately handle their access rights and authorize to appropriate network/information resources according to the valid national security policy. The problem tackles the need for cross-domain handling of the public key infrastructures (PKIs), certificate chains, digital signatures supporting the process of authentication, authorization, messages integrity, non-repudiation, privacy, etc. This is particularly important in terms of the dynamic nature of NNEC environment and the main SOA tenet of common access to relevant information by the unknown but authorized users.

In terms of web services, the most commonly used OASIS WS security (WSS) [8] standard addresses message security and focuses on credential exchange, message integrity, and message confidentiality. It is possible to integrate it with XML digital signature (DSIG) [9] (e.g., using X.509 certificates), and support a cross-domain authentication using existing security standards, such as security assertion markup language (SAML) [10]. This is a good foundation for SOA security in NEC environment, however the interoperability tests prove a gap existing in requirements in terms of obligatory elements included in the WS-security part of the SOAP message or SAML assertion.

4. Potential Solutions

The challenges presented in Section 2 can be met by different solutions. There is no complete product that can be used for that purpose, but there exist some smaller good practices and solutions that should be perceived as quick wins and make a good technological step forward.

NATO started its way towards the transformation by defining the NATO NEC (NNEC) feasibility study together with the roadmap, and foundation of the project. Some countries followed this idea carrying out their own studies, reflecting the challenges in terms of both technology and personnel. This has been done, e.g. by the Swedish Armed Forces [11] that defined the overarching architecture [12] and solid foundation for SOA implementation in their systems, as well as for integration of the legacy systems and their transformation towards full SOA [13]. Coherent introduction of service oriented architecture into currently employed and new systems can only be based on previously defined reference architecture of the C4I system and real information needs resulting from the planned operational scenarios.

In order to maximize the gain from linking the needs with capabilities, the role of business processes as a basis for developing service oriented architecture has been pointed out in [3] and [4]. The architecture engineering methodology (AEM) – developed by the NC3A algorithm for creating architectures, points out the necessity of creating a model of dependencies between components and applying them to a real operational context, to be sure that the proposed architecture will provide the required operational effect. For the graphical notation of operational process diagrams, the business process modelling notation (BPMN) has been recommended by the NC3A [3].

In the area of the semantic interoperability, there is a strong need to provide coherent and agreed standards enabling to share ontologies, as well as semantic descriptions of metadata and data itself. Achieving the interoperability on the semantic level requires the involvement of multinational activities that enable to agree on ontologies, proposed to be used in interdomain relations. Any proprietary solutions in this area will be unsuccessful unless other nations and systems developers agree on using them. The first step to provide semantic interoperability in a heterogeneous environment, is to develop a common semantic description of services for dynamic semantic search and adaptation purposes. This may be done using, e.g., more and more popular OWL-S. Semantic search engines by means of explicitly defining the semantics of the sources and by providing a relationship among terms (like a taxonomy) – can provide, e.g., a concept-based search. For information discovery, depending on the user preferences or the type of user, the system would be able to group the possible results that matched the query or refine the result list by filtering those that are of interest to the target audience. On the basis of semantic services descriptions (e.g. using OWL-S), semantic service discovery is able to provide results tailored to the user preferences, QoS, etc. as well as support orchestration and choreography.

The effort for providing ontologies for the NATO has been taken by the TIDE[5] (technology for information, decision and execution superiority) community, that aims to rapidly improve the operational capabilities through iterative processes based on horizontal and vertical integration of existing and emerging products. Within its framework, it stimulates coexistence of NATO and national programs and services providing proposition of standards embracing service oriented architectures and discussing them during regular meetings (so called TIDE sprints). It has proposed the service and information dis-

[5] More information can be found by the authorized users on the TIDE web site http://tide.act.nato.int/tidepedia/index.php?title=TIDE

covery protocols (for request-response and publish subscribe modes) used, e.g., within the NATO MSA community in BRITE and other national solutions (e.g., finish MEVAT system). So far, it has also delivered 33 ontologies used by different TIDE focus groups (e.g., MSA, SUCBAS, NIRIS etc.), based on RDF/RDFS and OWL. There are, e.g., subscribe-publish ontology, location ontology, symbology ontology (for APP6A and MS2525B visualization symbols), JC3IEDM ontology and many others.

The conceptual framework developed within the TIDE initiative describes how the network enabled capabilities will transform raw data into intended effects, and how they support achieving NATO's transformation goals and objectives. The effort TIDE members put into technology development has been used also in real life scenarios (e.g., in the BRITE system for creating NATO Maritime Situational Awareness – MSA). This should be continued and followed by other existing and possibly new teams that would support the interoperability on the data and metadata level, also using ontologies.

Another big step forward in SOA implementation is to put it into operation at the lowest command levels. According to the NEC principles and modern command processes, military operations are conducted in a dynamically, very often based on the "mission command" pattern. Low level commanders need to be aware of the possible consequences of their decisions. This makes it necessary to provide users down to the individual soldier level with the possibility to use the information systems and feed them with information crucial for the mission effectiveness.

The most common realization of the SOA environment is based on web services. In order for the operators to use the information from sources located on high command levels, web services realization must be made reliable and it must be adapted to the characteristics of the network.

It must be noted that there are no commercial works on applying SOA and web services in disadvantaged grids. There have been undertaken a few initiatives that focused on the information distribution over disadvantaged grids. In very low bandwidth environments the use of asynchronous replication based middleware, provides static information distribution between partners that have agreed to use a specific database format [14]. This solution is, however, not very convenient for a highly dynamic operational scenario. For this reason, there have been carried out researches on WS-based SOA solutions that provide the flexibility and interoperability, and are well suited to work in federation of systems. The NATO C3 Agency (NC3A) report on using WS in tactical domain [15] concludes with a statement that web services remain promising even in low bandwidth links, as long as very fast response times are not required. Other interesting works on this subject have been carried out in Norwegian Defense Research Establishment (FFI) in Norway [5], [16]–[19]. These projects focused on applying mechanisms that diminish XML-based disadvantages of WS and experiment on new transport protocols instead of SOAP/ HTTP (hypertext transfer protocol), e.g., data distribution service (DDS) or message handling system (MHS).

Data-rate constraints in tactical networks impose great challenges that have to be faced in order to fully deploy SOA supporting NEC. There are several solutions that can be applied to adapt SOA web services mechanisms to the capacity of the systems at various C2 levels. The ones which bring the biggest advantage are compression (e.g., very popular and available in the application servers GZip algorithm), filtering, caching and non-SOAP transport mechanisms.

Two main disadvantages of using XML are as follows: big overhead related to human-readable format of the documents and significant parsing and processing times of the XML-based messages. In [15] there has been shown that XML is very compression-friendly and in many cases, depending on the data set and the size of the message, more than 95% gain can be achieved. This very simple method should be set as a requirement for the use of WS in disadvantaged networks. It must be noted however, that in order not to complicate the interoperability, compression algorithms should be standardized and known by the collaborating parties.

Another very efficient method of limiting XML disadvantages is using its binary form. Generally, it reduces the verbosity of XML documents and the cost of parsing. It can lead, though, to faster document processing and lower memory and central processing unit (CPU) requirements, which is especially appealing on mobile devices.

The filtering enables to limit the amount of information sent to the end user in order to relieve the tactical network from sending heavy traffic. This method is often used for the access control based on XML guards (XML filters), or for providing subscriptions based on the messages content. However, the utilization of this method for WS adaptation to disadvantaged grids is a very complicated matter since it is very difficult to propose very general filter rules that apply to most Web Services, and that can provide the end user with the right set of information that he really needs at that moment.

Some of the filtering functionalities are present it HTTP - like, e.g., requiring only a part of the HTTP page from the server. In force tracking systems [20] filtering may mean sending information about objects that are away from the end user (service requester) by x km creating some kinds of circles of the service accuracy. For provision of still images, it can mean sending the image with the resolution adapted to the end user terminal. For video it may mean decreasing the frequency of frames per second.

HTTP is one of the most popular binding protocols primarily designed to transport SOAP messages. HTTP holds its connection after the SOAP request is sent until the SOAP response is returned in the HTTP acknowledgment. If the connection times out (e.g., because of delays), the SOAP response will not be delivered to the service consumer. Therefore, using HTTP over disadvantaged grids or a combination of heterogeneous networks may not work very well [17].

According to [15], sending new service requests is much more "expensive" (in a performance sense) than extending the existing ones. For instance, during the HTTP "handshaking" that takes place before each request, a lot of extra traffic is generated each time. Therefore, two 1000-byte requests will take far longer than one 2000-byte request. For the same reason, the performance of HTTP communications, which are made up of sets of data packets, are much more adversely affected by including more packets than more data. In other words, a 20-packet "conversation" equal 1000 bytes will be slower than a 15-packet conversation equal 2000 bytes.

It is a good practice to use a proxy element with a store-and-forward functionality that could cope with the unpredictability of QoS parameters of tactical communications networks, especially in terms of frequent network disconnections. Such an element can provide different transport mechanisms, like data distribution service [21] and military message handling system [18] that have been designed to work in disadvantaged environments. Caching is one of the functions of the proxy element that enables, e.g., storing information that crossed the proxy to be available for subsequent requests. If a user requests the same object, proxy can send it without the need to ask the server again. However, proxy must be able to handle stored information appropriately. This will not work well with short-lived information, that change frequently and information, the expiration of which cannot be assessed. Caching functionality should be placed at the edge of the low-bandwidth network – quite near the end user. It can decrease the network load on the server-proxy path and shorten the time to send the client response. Therefore, the good practice is to keep the frequently used (or infrequently changing) data at, or near the client, so that its retrieval does not impact the network considerably.

Web services compression and filtering can be used in the ultimate sender and receiver of information, however NNEC FS and other NATO documents [22] propose the utilization of edge proxies that provide the possibility to adapt web service realization to the possibilities of the network.

Norwegian Defense Establishment proposes, for that purpose, so called Delay and disruption tolerant SOAP Proxy (DS Proxy) [23] which is able to store-and-forward SOAP messages, compress them and provide prioritization of the traffic. It is placed between a web service consumer and a web service. When the web service is temporarily unavailable (also due to network disruptions), the DSProxy component will cache the Web service request, and retry the invocation at intervals, returning the Web service response when finally successful. The DSProxies can be working in a group. They are self organizing into an overlay network consisting of any number of DSProxy components, based on a mechanism which relies on UDP multicast. This provides the possibility to traverse multiple and heterogeneous networks.

Initial tests have proven that the DSProxy solution is able to bridge heterogeneous networks and offer store-and-forward capabilities. Using TCP and UDP transport protocols, web services were invoked in links with bandwidth equal 400 bit/s and 300 bit/s which is a very promising result.

Another approach presented in [24] assumes a situation when the client needs information, but the access network is degraded and he cannot receive it in a timely manner. In case when the QoS guarantees cannot be met, it is proposed that the client interacts with the proxy mediation service, which is able, on the basis of the current network parameters, to adapt his traffic to the possibilities of the network, and to enable sending it in some other way.

The proxy service aims at delivering web services to the users located in disadvantaged networks and minimizing negative effects of the wireless environment, including:

- Connection failures – due to store & forward capabilities and publish/subscribe approach;
- High delays and low throughputs – due to reduction of the packet size – e.g., compression, binary coding with compression (e.g., using Efficient XML);
- High error rates – due to reliability mechanisms and network monitoring.

When the client is temporarily disconnected from the network, the mediation service can store the data and forward them when it will be available again. The proxy actively interoperates with the service client and service producer, dynamically adapting contents of the XML messages, selecting the best communication means from within the available ones for the specific service call. Characteristics of the service call are described in the context of a call that provides the possibility to make appropriate and the most accurate decision.

The context consists of the user profile (static and dynamic), terminal profile, network profile (dynamic) and service description. All of them are described in OWL files. In order to take appropriate actions, there is also an adaptation ontology that defines all the actions that need to be taken in order to fulfil the user request.

It must be emphasized that for some types of services web service technology in the tactical domain can be replaced with, e.g., DDS middleware (data distribution service), that is a real time publish/subscribe data-centric platform for dynamic distribution of information in real time. DDS has strong and extensive supports for QoS and is used successfully in real life in many European armies.

5. Summary and Conclusions

The challenges and solutions presented in this paper have been gathered based on the available literature on this subject as well as the experience of the authors in SOA application in NEC environment. They show the basic steps in order to support dynamic, flexible and scalable SOA – based environment to conduct net-centric multinational operation. Briefly, the SOA success factors can be summarized to the list of Quick Win solutions (see Table 1). This set provides a guide of best practices that enable the undoubtedly successful architecture paradigm to be used in highly dynamic and heterogeneous NEC environment that should technically support multinational missions.

Table 1
SOA success factors

1	Develop service layer based on real information needs
2	Provide semantic description of services for dynamic semantic search and adaptation purposes
3	Provide high level management of the business processes mapped into services, enabling dynamic reaction on new information needs, change into business process to create dynamic service environment that flexibly changes in order to meet the information needs
4	Provide low transmission overhead
5	Optimize service realization (minimize the number of interactions between architecture components)
6	Minimize the utilization of end-terminal resources, especially in the tactical domain
7	Provide interoperation of the middleware layer with the transport and application layers
8	Provide the possibility to negotiate a service contract and adapt service realization to available resources
9	Address the security issues by providing cross-domain authentication given local authorization and security policy

The SOA's greatest advantage is that it provides seamless information exchange based on different policies and loose coupling of its components. The use of SOAs facilitates the application and data sharing and provide a flexible mechanism for reusing existing services to enable the development of new, value-added information services [3], [4].

It appears that the service-orientation and SOAs can facilitate the implementation of net-centric capabilities, but by themselves do not guarantee net-centricity. It must be emphasized that even we have seen many of so called – net-centric solutions, like often shown in demonstrations – common operational pictures (COPs), the truth is that current service oriented technology standards and products only support achieving mission effectiveness. It is important that net-centricity is as much a business model or organizational relationship issue as it is a technology one. Many organizations, including R&D institutions, NATO and industry are beginning to address the issue of service orientation across enterprise boundaries, in FoS environment, on all the echelons of command, but still much work remains to be done.

The article points out SOA technical factors that can approach to the success of its application in NEC environment. It is very important to note that only tailoring the SOA design to military environment limitations, management and security constraints as well as operational needs can provide a true benefit of the SOA paradigm.

Acknowledgment

This work was carried out within the project PBZ – MNiSW – DBO-02/I/2007 supported by the R&D funds allocated for the years 2007–2010.

References

[1] P. Bartolomasi, T. Buckman, A. Campbell, J. Grainger, J. Mahaffey, R. Marchand, O. Kruidhof, C. Shawcross, and K. Veum, *NATO Network-Centric Operational Needs and Implications for the Development of Net-Centric Solutions*, vol. 1 of *NATO Network Enabled Capability Feasibility Study*, version 2.0, NATO C3 Agency, 2004/2005.

[2] "NNEC Data Strategy", Headquarters, Supreme Allied Commander Transformation, IS-NNEC IPT, September 2004.

[3] G. Babakhani, J. Busch, C. Dumas, R. Fiske, B. Holden, H. Lægreid, R. Malewicz, D. Marco-Mompel, and V. Rodriguez-Herola, "Web trends and technologies and NNEC core enterprise services", version 2.0, Technical Note 1143, NATO C3 Agency, The Hague, Dec. 2006.

[4] M. Booth, T. Buckman, J. Busch, B. Caplan, B. Christiansen, R. van Engelshoven, K. Eckstein, G. Hallingstad, T. Halmai, P. Howland, V. Rodriguez-Herola, D. Kallgren, S. Onganer, R. Porta, C. Shawcross, P. Szczucki, and K. Veum, *Detailed Report Covering a Strategy and Roadmap for Realizing an NNEC Networking and Information Infrastructure*, vol. II of *NATO Network Enabled Feasibility Study*, version 2.0, NATO C3 Agency, June 2005.

[5] R. Haakseth, T. Gagnes, D. Hadzic, T. Hafsøe, F. T. Johnsen, K. Lund, and B. K. Reitan, "SOA – Cross Domain and Disadvantaged Grids", NATO CWID 2007, FFI-rapport 2007/02301, Norwegian Defence Research Establishment.

[6] C. M. MacKenzie, K. Laskey, F. McCabe, P. F. Brown, R. Metz, Booz, and A. Hamilton, "Reference Model for Service Oriented Architecture 1.0", *OASIS Standard*, 2006 [Online]. Available: http://docs.oasis-open.org/soa-rm/v1.0/soa-rm.html

[7] "NC3A Technical Architecture" [Online]. Available: http://nc3ta.nc3a.nato.int/website/home.asp?msg=logreq

[8] A. Nadalin, C. Kaler, R. Monzillo, and P. Hallam-Baker, "Web Services Security: 4 SOAP Message Security 1.1 (WS-Security 2004)", *OASIS Standard Specification*, 2006 [Online]. Available: http://www.oasis-open.org/committees/download.php/16790/wss-v1.1-spec-os-SOAPMessageSecurity.pdf

[9] M. Bartel, J. Boyer, B. Fox, B. LaMacchia, and E. Simon, "XML Signature Syntax and Processing (Second Edition)", *W3C Recommendation*, 2008 [Online]. Available: http://www.w3.org/TR/xmldsig-core/

[10] E. Maler *et al.*, "Assertions and Protocols for the OASIS Security Assertion Markup Language (SAML) V1.1", *OASIS Standard*, 2003 [Online]. Available: http://www.oasis-open.org/committees/download.php/3406/oasis-sstc-saml-core-1.1.pdf

[11] O. Winberg, "SOA for NBD: principles and considerations", Försvarets Materialverk, 2006.

[12] O. Winberg, "Overarching architecture for FMLS 2010 technical system", Försvarets Materialverk, 2006.

[13] O. Winberg, "Deign Rule: legacy integration", Försvarets Materialverk, 2006.

[14] "Information Management over Disadvantaged Grids", Techn. Rep. RTO-TR-IST-030, AC/323( )

[15] J. Busch, "An investigation into deploying Web services", TN–1229, NATO C3 Agency, Dec. 2007.

[16] R. Haakseth, D. Hadzic, K. Lund, A. Eggen, and R. E. Rasmussen, "Experiences from implementing dynamic and secure web services", in *Proc. 11th ICCRTS*, Cambridge, UK, 2006

[17] T. Hafsøe, F. T. Johnsen, K. Lund, and A. Eggen, "Adapting web service publish/subscribe technologies for use in NEC C2 Systems", in *Proc. 12th ICCRTS Conf.*, Newport, USA, 2007 [Online]. Available: http://www.dodccrp.org/events/12th_ICCRTS/CD/Launch_CD.html

[18] F. T. Johnsen, A. Eggen, T. Hafsøe, and K. Lund, "Utilizing military message handling systems as a transport mechanism for SOA in military tactical networks", in *Proc. IST 083 Symp.*, Prague, Czech Republic, 2008.

[19] K. Lund, A. Eggen, D. Hadzic, T. Hafsøe, and F. T. Johnsen, "Using web services to realize service oriented architecture in military communication networks", *IEEE Commun. Mag.*, Oct., pp. 47–53, 2007.

[20] "Interim NFFI Standard for Interoperability of FTS", AC322(SC5)N(2006)0025, NC3B Information Systems SC, 16 Dec. 2006.

[21] "Data Distribution Service for Real-time Systems Version 1.2", OMG Available Specification, Jan. 2007.

[22] R. Faucher, R. Ladysz, D. Miller, S. Musman, S. Raparla, and D. Smith, "Guidance on proxy servers fpr the tactical edge", DoD C3I FFRDC, The Mitre Corporation, Sept. 2006.

[23] E. Skjervold, T. Hafsøe, F. T. Johnsen, and K. Lund, "Delay and disruption tolerant web services for heterogeneous networks", in *Proc. MILCOM*, Boston, USA, 2009.

[24] J. Śliwa and D. Duda, "Adaptive web services supported by Qos IP network", in *Proc. Military Commun. Inf. Syst. Conf.*, Prague, Czech Republic, 2009, pp. 448–456 (MK-291).

Joanna Śliwa was born in 1979 in Warsaw. She graduated from the Faculty of Electronics and Information Technology of Warsaw University of Technology (2003). At the moment she is a researcher in Military Communication Institute in Zegrze, Poland. She is working on her Ph.D. in the area of efficient realization of web services in disadvantaged networks. Her main areas of interests are: new telecommunication technologies, Service Oriented Architecture, Network Enabled Capabilities and QoS provisioning.

e-mail: j.sliwa@wil.waw.pl
Military Communication Institute
Warszawska st 22A
05-130 Zegrze, Poland

Marek Amanowicz was born in Poland in 1946. He received M.Sc., Ph.D. and D.Sc. degrees from the Military University of Technology, Warsaw, Poland in 1970, 1978 and 1990, respectively, all in telecommunication engineering. In 2001 he was promoted to the professor's title. He was engaged in many research projects, especially in the fields of communications and information systems engineering, mobile communications, satellite communications, antennas and propagation, communications and information systems modeling and simulation, communications and information systems interoperability, network management and electronics warfare. He is an author or co-author of over 200 scientific papers and research reports.

e-mail: marek.amanowicz@wat.edu.pl
Military University of Technology
Faculty of Electronics
Gen. Sylwestra Kaliskiego st 2
00-908 Warsaw, Poland

e-mail: m.amanowicz@wil.waw.pl
Military Communication Institute
Warszawska st 22A
05-130 Zegrze, Poland

Paper

Authentication and Authorization of Users and Services in Dynamic Military SOA Environments

Bartosz Jasiul, Joanna Śliwa, Rafał Piotrowski, and Robert Goniacz

Military Communication Institute, Zegrze, Poland

Abstract—The problem of user authentication and authorization is usually being solved in a single system. Federated environment assumes heterogeneity of systems, which brings the problem of mutual users and services authentication and authorization. In this article the authors presented security requirements for cross domain information exchange in federated environments and a method of secure access to information resources on the basis of web services. Special attention was paid to authentication and authorization of users and services. As opportunities, there were presented solutions verified in multinational experimentations and exercises.

Keywords—authentication, authorization, military networks, SOA, web services.

1. Introduction

The main tenet of net-centricity is to achieve information superiority by sharing reliable information collected from various sources, creating situational awareness and distributing it among mission participants, across domains, context and organizational boundaries. However, the improvement of collaboration and information sharing in highly dynamic, unpredictable network enabled capability (NEC) environment is a great challenge. It assumes transfer of information between users of so called federation of systems with required quality of service and security independently of the underlying infrastructure as well as common access to relevant information by the authorized users.

Federation of systems (FoS) capability derives from the strategy for developing the networking and information sharing aspects of NATO NEC (NNEC) and focuses on joining together the networking and core information systems from NATO and NATO nations. The FoS concept refers to a set of different systems, which are not centrally managed, but are so connected or related as to produce results beyond those achievable by the individual systems alone [1]. This implies that networking and information infrastructures (NII) consists of national NIIs segments and a NATO networking and information infrastructure (NNII), which together will provide capabilities that no system can provide by itself. This concept is similar to the one known from Internet, where there is no central control, and the synergy for the federation is achieved through collaboration and cooperation. Operational capabilities needed to conduct modern military operation, from the technical point of view, impose the use of flexible, adaptable architecture enabling seamless information exchange in dynamically changing, unpredictable federation of systems. In order to satisfy these requirements, NATO recommends the use of service oriented architectures (SOAs), that succeeded in commercial world lately and are seen as the crucial NEC enabler [1], [2], [3].

Service orientation is a conceptual architecture which asymmetrically provides services to arbitrary service consumers, facilitating the information sharing in heterogeneous environment, and thus supports, to some degree, aspects of net-centricity. SOA can make military information resources available in the form of services that can be discovered and used by all mission participants that do not need to be aware of these services in advance. That is why the NII strategy assumes that the system infrastructure will be implemented as a FoSs, involving the use of SOA to expose software functions as consumable services that can be discovered and invoked across the network.

SOA's greatest advantage is the ability of seamless information exchange based on different policies and loose coupling of its components. However, this can be realized by the widespread use of open standards. One of the most mature realizations of SOA assumes application of web services (WS) – the most successful implementation of this paradigm.

WSs, based on extensible markup language (XML), SOAP, web services description language (WSDL) and related open standards, implemented in national systems, allow data and applications to interact without human intervention through dynamic, ad hoc coalition connections. WSs are in fact described by a wide range of standards that deal with different aspects of WS realization, transport, orchestration, semantics etc. They provide means to build a very flexible environment that is able to dynamically link different system components to each other. The most important and obvious advantage of this solution is its natural applicability to FoS, where it can be implemented in a wide variety of communications systems, can coexist with other technologies and software design approaches [4], and be adopted in evolutionary way without the need to modify the legacy systems.

2. Security Challenges

Sharing information among mission participants in FoS imposes many technological interoperability on data, application and communications levels. What is more, the unpredictability of this environment, mobility and dynamic nature of NEC operations impose several threats that do not occur in a system managed by single administration.

The security challenges inherent to the web services approach are alarming and unavoidable [2]. Many of the features that make web services attractive, including greater accessibility of data, dynamic application-to-application connections, and relative autonomy (lack of human intervention) are being seen by the security officers as threatening or even forbidden with traditional security models and controls. Even the idea of ubiquitous information sharing with different security classification between domains building FoS is currently seen as risky.

Security objectives [5] for federated SOA-based systems are the same as for all IT communication means and cover confidentiality, integrity, access control, non-repudiation, accountability and availability. Such attitude guarantees a controlled access to network elements, services and applications for identified users.

Both identification of users and services across domains and providing appropriate information for making authorization decision rise interoperability problems. It is even more visible in SOA- based systems, where services can be invoked by users not known in design time, so that SOA-based systems must face up additional requirements [6]:

- balancing information sharing with security,
- trust propagation between federated systems,
- minimizing vulnerabilities (e.g., in terms of software development),
- providing access to system resources for unanticipated users.

First of all, in order for the users to share the information and use the received ones, they must be sure that the source and sink are reliable. Information can be, though shared only among users/devices that have been identified and are approved for this kind of data. Only cross-domain authentication and authorization, based on trust relation between security providers and appropriate identity management are able to fulfill initial security requirements for FoS.

As shown in Fig. 1, in national C4I system the access to information is controlled and granted only for authenticated and entitled users from this domain. Access control is an internal issue of the system and can be realized in different ways. The authorization decision is made locally and usually bases not only on user rights attributes but also on invoked web service attributes and valid security policy. The set of user attributes required for making decision may differ depending on the system implementation. The mechanism usually secures all information flows inside the system.

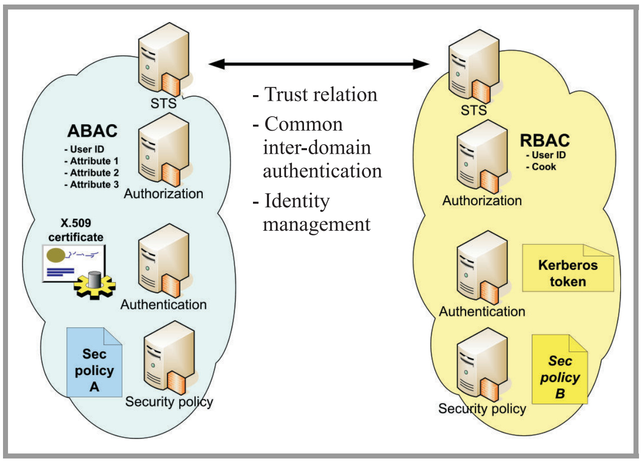


Fig. 1. Cross-domain authentication and authorization of users and services. Intradomain and interdomain security mechanisms.

The authentication mechanism is to confirm the identity of a user or service (see Fig. 2). To perform access control, web services need to identify and authenticate the requesters. In FoS it can be performed by different means. Each domain can have their own authentication services that base on login/password, X.509 certificates, biometric data, etc. Authentication services in different domains must interoperate with each other and accept the identity confirmation issued in other domain (some kind of a token). It must be emphasized that the user must be appropriately identified across domain boundaries.

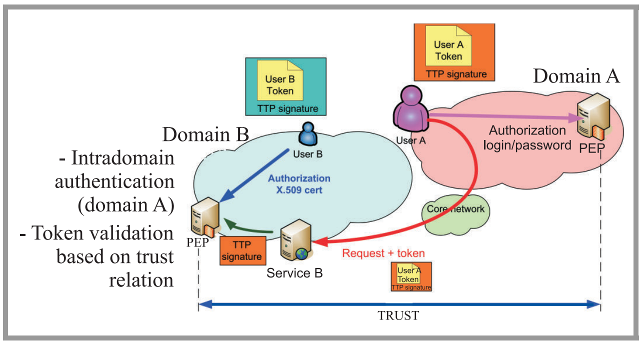


Fig. 2. Authentication, authorization and trust relation in cross-domain information sharing.

The access to services can be granted or denied based on the authorization service decision which can use different access control (AC) models (e.g., role based – RBAC, attribute based – ABAC, policy based – PBAC, risk adaptive – RAAC). In real life scenarios domains can use different combinations of the above, applying rules that are defined for this particular system. These rules can be generally called policy of information sharing, which do not imply that this is PBAC.

In disrupted networks, where connections are not stable, to achieve the reliability of the system, a distributed policy repository implementation is recommended. This solution allows all policy decision points (PDPs) present in the system to perform authorization decision invoking local copy of the policy. It should be noted that the policy rules may change dynamically during the system and mission lifetime (e.g., domains or users may become untrusted). The rules may change as a result of the administrator activities as well as the actual risk assessment. These changes must be immediately distributed to all policy repositories employed in the system. Effective policy database replication mechanism is essential to enable all PDPs to realize access control decisions according to current and valid rules.

Some users may frequently request particular set of services provided by own or federated domains. Getting authorization each time the user queries any resource may overload the security services and can make the process of retrieving results more cumbersome. Thus, an important issue is the provision of a mechanism allowing users to authenticate with one system to a single point (the user's identity provider) and, on the basis of that authentication statement (usually being some kind of a token), use other services and applications within SOA. This mechanism is called single sign-on (SSO) (see Fig. 3). SSO allows users to have more flexible access to system resources which can limit the time needed to get necessary data and number of authentication requests.

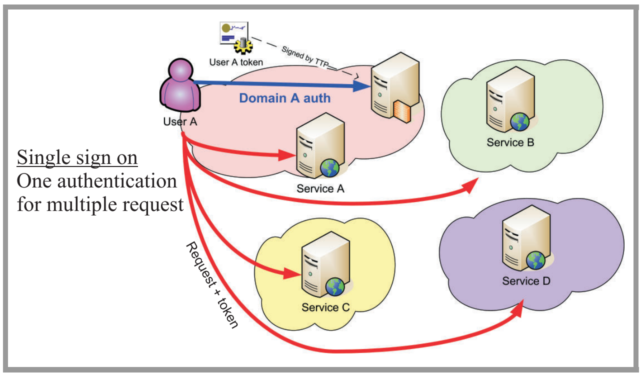


Fig. 3. Single sign on in FoS.

As it was already stated, one of the core requirements for sharing information on FoS is trust relation among cooperating nations. All entities involved in a transaction or process must trust one another. A level of trust must be in place for the parties willing to cooperate in common operations and to exchange sensitive information in a timely manner. The willingness to share and accept information depends heavily on trusting the receiver and the provider of information.

There are several trust relationship models that can be used in WS security. The most useful is federation of trust, which bases on the federated identity management. A federated trust model allows users and web services from various domains to interact with some level of security. It is based on both the brokered and bilateral trust models. Particular domains use so called trusted third party (TTP) that is to certify that a service or a requester can be trusted within the domain. Each domain has its own TTP and TTPs of these domains have mutual relationships. Establishing trust is very important during the client-provider transactions as well as in service-to-service interactions.

Provision of security in terms of authentication and authorization in federated environment goes beyond the challenges presented above. Cross-domain solutions that provide the identification of users, are able to present and then to analyze their credentials for the purpose of authorization. They usually base on assertions or tokens that prove local authentication and enable granting access to the resource. The problem gets more complicated when we imagine a service composition (the chain of services), when one application requests data from another one. In this case, the entity that initiated the process should be granted the access to particular resources. The problem can be solved by identity delegation that assumes passing identity of requesting user through the service chain, however its technical realization can be cumbersome in the range of one system, not mentioning a federated environment.

3. Opportunities

In order to show the opportunities of authentication and authorization in a federated environment, we present the results of experiments that were carried out during preparation for the multinational experimentation MNE 6 and demonstrated on the Combined Warrior Interoperability Demonstration (CWID) in 2009. The solution covers the basic requirements for authentication and authorization of users and services and was used to share data between maritime systems supporting creating multinational interagency situational awareness on the sea. This section describes the cross-domain authentication and authorization solution that was implemented in two different domains independently, given existing local interdomain security mechanisms.

3.1. Description of the Trial

The authentication and authorization mechanisms in presented system are developed in service oriented architecture as a set of loosely coupled security web services. To ensure a trust relation between heterogeneous domains forming federation of systems there was assumed mutual acknowledgement of public key infrastructures (PKIs) approved in each domain.

When the user wants to get access to a resource located outside his domain, he needs to be authenticated in his own domain. In order to prove his authentication in the other domain he gets security assertion (some kind of a passport) consisting of user identity, public key and attributes. Local authentication can be made with different security mechanisms, e.g., id and password, Kerberos ticket, X.509 certificate (see Fig. 4). This solution guarantees the independence of policy rules in each autonomous system. However, the assertion must be understandable to the cooperating system. The trust to the users is internal case of the system and may depend on the authentication method, because there is a huge difference of efficiency between X.509 and, e.g., id and password mechanism.

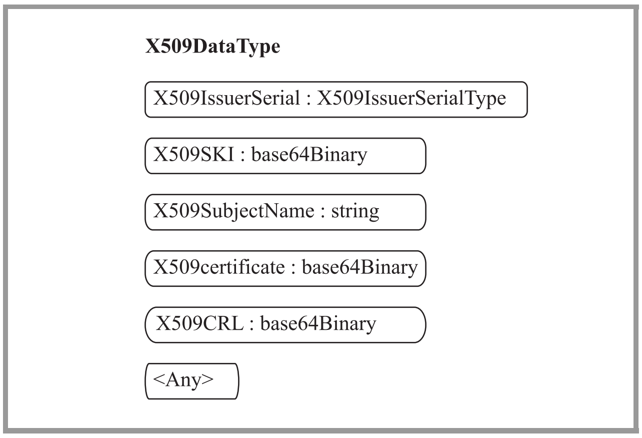


Fig. 4. X.509 certificate schema.

In the presented solution X.509 certificates [7] were implemented for the intradomain user authentication. Certificates based on open industry standards are supported on many platforms. Additionally, X.509 certificates can be used to provide confidentiality and data origin authentication at the message and transport layer. The identity of particular participant in a message exchange is unique and can be confirmed by verification of signature made using X.509 certificates.

For cross-domain authentication security assertion markup language (SAML) [8] tokens were chosen. They carry X.509 credentials and additional values, e.g., signatures and user attributes. SAML is an XML-based standard introduced by Security Services Technical Committee of the OASIS for exchanging authentication and authorization data between security domains. In the presented model, SAML assertions are transferred from identity provider (that can be, e.g., secure token service) to service provider. It must be noted that signatures included in SAML tokens guarantee the message integrity and non-repudiation (signatures unable to fake identity of user and his attributes).

The advantage of utilizing SAML is its flexibility and adaptability to carrying variety of properties which can be used for securing communication. In fact, only a few elements are mandatory in SAML assertion and the rest is optional. Assertion contains statements that the service provider uses to make access control decisions.

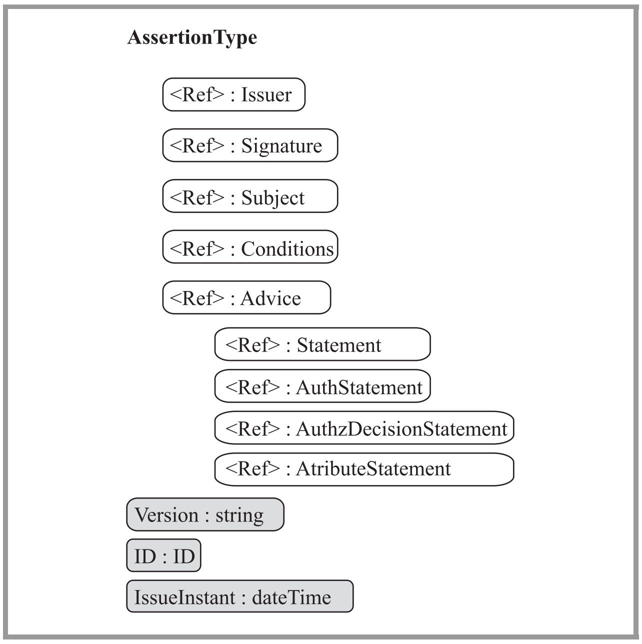


Fig. 5. SAML assertion schema (grey tags are for obligatory elements, white ones – for optional).

SAML provides three statements (see Fig. 5.):

- Authentication – asserts the service provider that the principal (e.g., STS, PEP) authenticated the requester at a particular time using a given method of authentication (e.g., user/password, biometric data, X.509 certificates, Kerberos, etc);
- Attribute – provides attributes of requester that could be used to make access control decision;
- Authorization decision statements – provide permissions for particular actions.

SAML assertion is composed of obligatory and optional elements. This allowed us to incorporate information about local authentication of the user based on X509 certificates and send user attributes for the purpose of authorization. SAML assertion used for the trial consisted of the following elements:

- Issuer – the unique identifier of the requesting service provider/the unique identifier of STS, PEP;
- Subject – SOAP message source unique identifier (requester or a service provider).
  Both the issuer and the subject data are extracted from X.509 certificates and consist of common name, organization unit, organization, country (CN, OU, O, C);
- Signature – a value obtained from signing the whole request/response message; it provides the message integrity and guarantees non-repudiation;
- Conditions – the validity period; it consists of values: NotBefore and NotOnOrAfter;

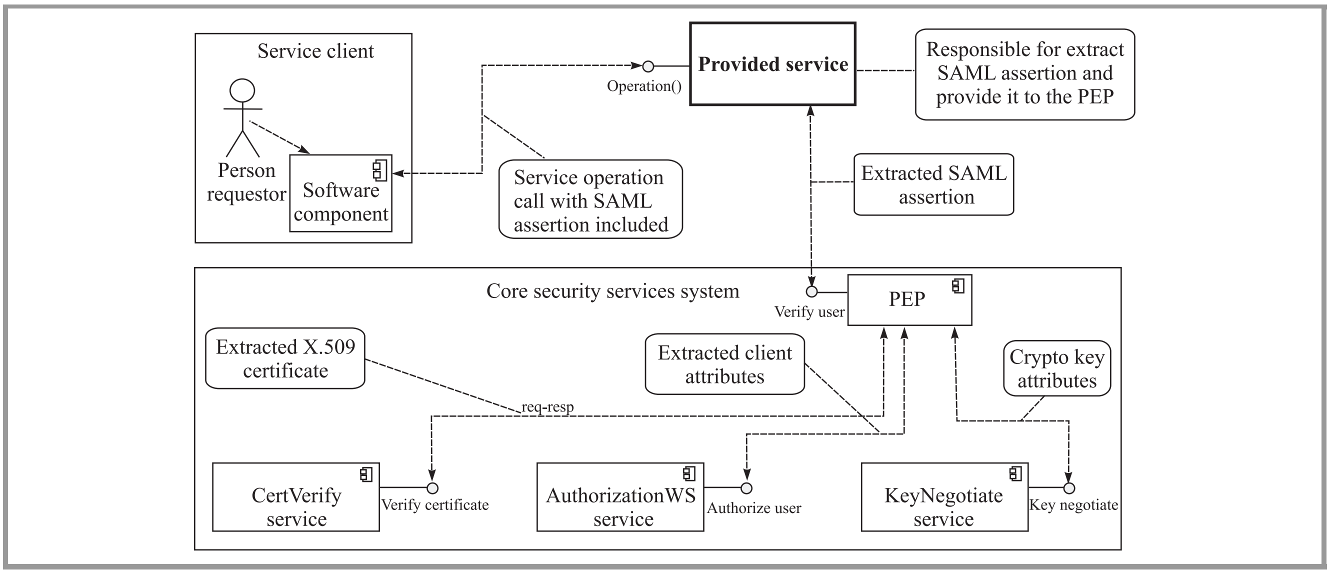


Fig. 6. Core security services system workflow diagram.

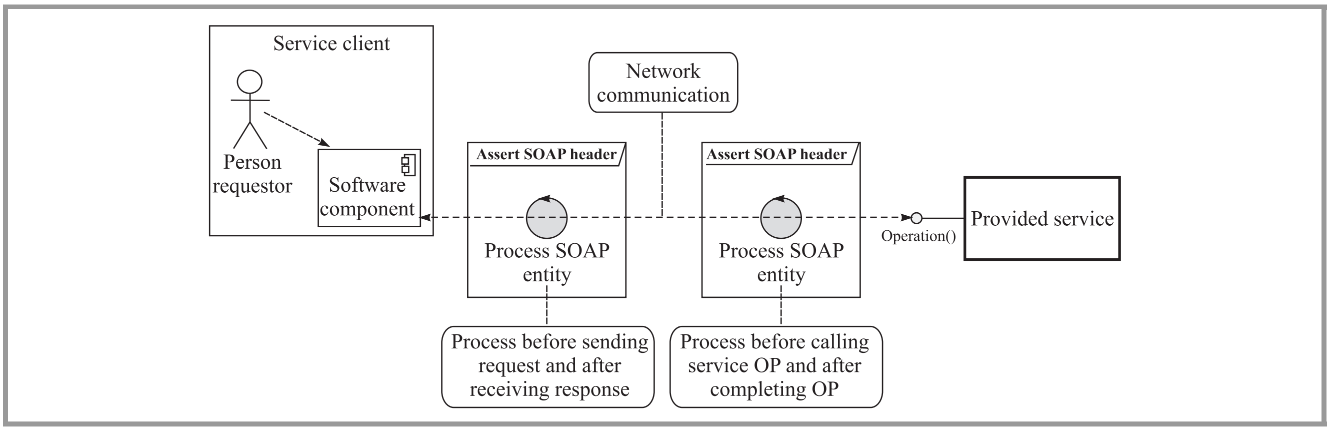


Fig. 7. Workflow of processing SOAP request and response messages.

- Authentication statement – a method used for the authentication of the user;
- Attribute statement – attributes of the user used for authorization, e.g. military rank, function in organization, secrecy permissions.

3.2. Implementation of the Solution

Each request to a service must be augmented with appropriate SAML assertion, with a token proving the user authentication and its credentials necessary for local authorization. Such an assertion the user usually gets from so called security token service, that is able to verify its identity (based on the local authentication mechanisms) and to prepare an appropriate SAML assertion, understandable to the requested site. Appropriately prepared SOAP message is sent to the service.

For the purpose of the trial, a set of security services was implemented (see Fig. 6). They are to provide cross-domain authorization and authentication based on the trust relation. Policy enforcement point is treated as main SAML assertion processing module (engine). It analyzes part of the SOAP message header, i.e., SAML assertion and delegates tasks to particular auxiliary security modules: CertVerify service, AuthorizationWS service and KeyNegotiate service. Together with PEP they create the core security services system.

PEP and auxiliary modules are implemented as web services communicating with each other with SOAP messages. As mentioned before, PEP with auxiliary services, is responsible for validating SAML assertions. It:

- validates X.509 certificate and digital signature of SOAP message embedded in assertion,
- extracts and validates user authorization attributes,
- obtains symmetric crypto key from KeyNegotiate service.

The certificate included in the SOAP message belongs to the requester and it is validated against public key of its

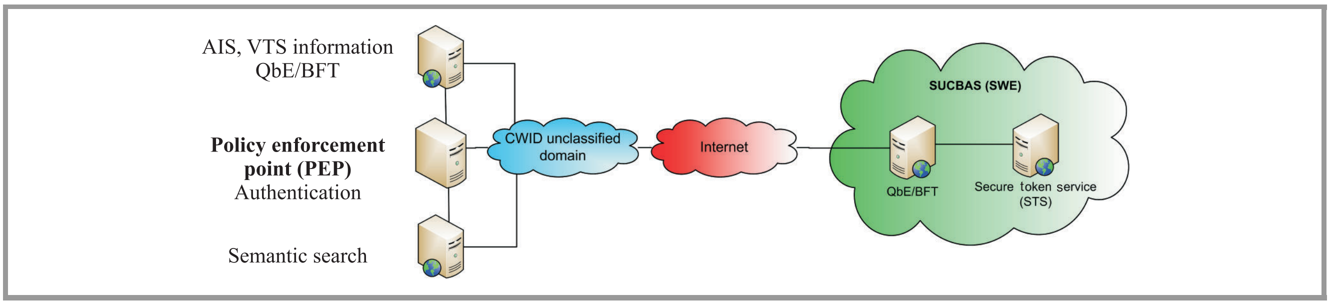


Fig. 8. CWID test infrastructure.

reverse processing workflow takes place while sending each of the SOAP response messages, e.g., to authenticate the service to the requester. After the client has had an access granted, the requested service communicates with authority which was bilaterally exchanged at the memorandum of cooperation (at the level of establishing trust). User attributes are credentials describing the user place in the organization hierarchy. Depending on the level of trust between domains, this credentials may be used to take appropriate authorization decision.

KeyNegotiate service is to exchange seeds that allows to calculate symmetric crypto key. It could be derived from shared secret values or from asymmetric computations. This crypto key is used to encrypt communication between the user and the requested service, which fulfils the confidentiality requirements.

If all of these processing stages return valid and correct values, PEP module, in turn, allows web service to proceed further proprietary operations. The only stated requirement from the service provider is to call PEP operations before processing any proper web service task, just after receiving SOAP message request. The implementation was based on open source web services frameworks and APIs, e.g., JAX-WS Java API for web services, Bouncy Castle Java Cryptography API and EJBCA – Open Source PKI library.

PEP in order to have its response message appropriately prepared. Just before sending web service response, PEP is involved in preparation of the SAML assertion. It gains user attributes and crypto key and prepares whole SOAP message including SOAP header. This is implemented by using output handler (JAX-WS handler), which delegates the workflow to PEP. General logical workflow of processing request and response SOAP messages is presented in Fig. 7.

3.3. Verification of the Solution

General overview of tests carried out during workshops and CWID 2009 exercises is depicted in Fig. 8. Tests and implementations were prepared and developed by Polish and Swedish teams engaged in realization of objective 4.2 titled Multinational Interagency Situational Awareness – Extended Maritime (MISA-EM), which is part of sixth edition of MultiNational Experimentations (MNE-6).

It must be emphasized that the Swedish and Polish implementation were prepared separately, without exchanging any line of code. General concept of the cross-domain security solution for both realizations based on the same assumptions and was agreed before experiments. However, the software products and solutions of security web services are different. The Swedish implementation does not include PEP, however its functionalities were adopted to security proxy service adding/removing security assertions and secure token service (STS).

The tests covered a few scenarios. In the first scenario (see Fig. 9) the Swedish side invokes Polish blue force tracking (BFT) service. At the client side, STS provided exclusively by the Swedish side, was responsible for preparing appropriate SAML assertions embedded in the SOAP messages. After receiving message by BFT service provider, SAML assertion was analyzed by PEP, provided exclusively by the Polish side. In case of any errors (i.e., wrong or not valid X.509 certificates, not appropriate private keys, modifications injected into SOAP messages), PEP informs the BFT provider about such a situation and the processing chain can be stopped (access denied). In case of successful SAML assertion validation, BFT access was granted and the service was called. Before sending BFT response to the Swedish client, PEP was responsible for preparing appropriate SAML assertion embedded in the SOAP response message. After receiving BFT response, STS analyzed SAML assertion prepared at the Polish side. Similar checking workflow was executed by STS module and in case of any incompatibilities, BFT access was not granted and the client was informed about that.

The second scenario was very similar to the mentioned above. Security services were used in this case to provide cross-domain access control to the query by example services (QbE) implemented at both sides. This scenario tested bidirectional cooperation (SWE service – POL client, POL service – SWE client).

In third scenario, QbE service was called indirect through semantic search service. Details regarding the implementation and working of semantic search are out of scope of this article but positive results of the tests prove proper work of the security solutions.

It should be emphasized that all the tests were successful. During the preparation phase for these experiments, the

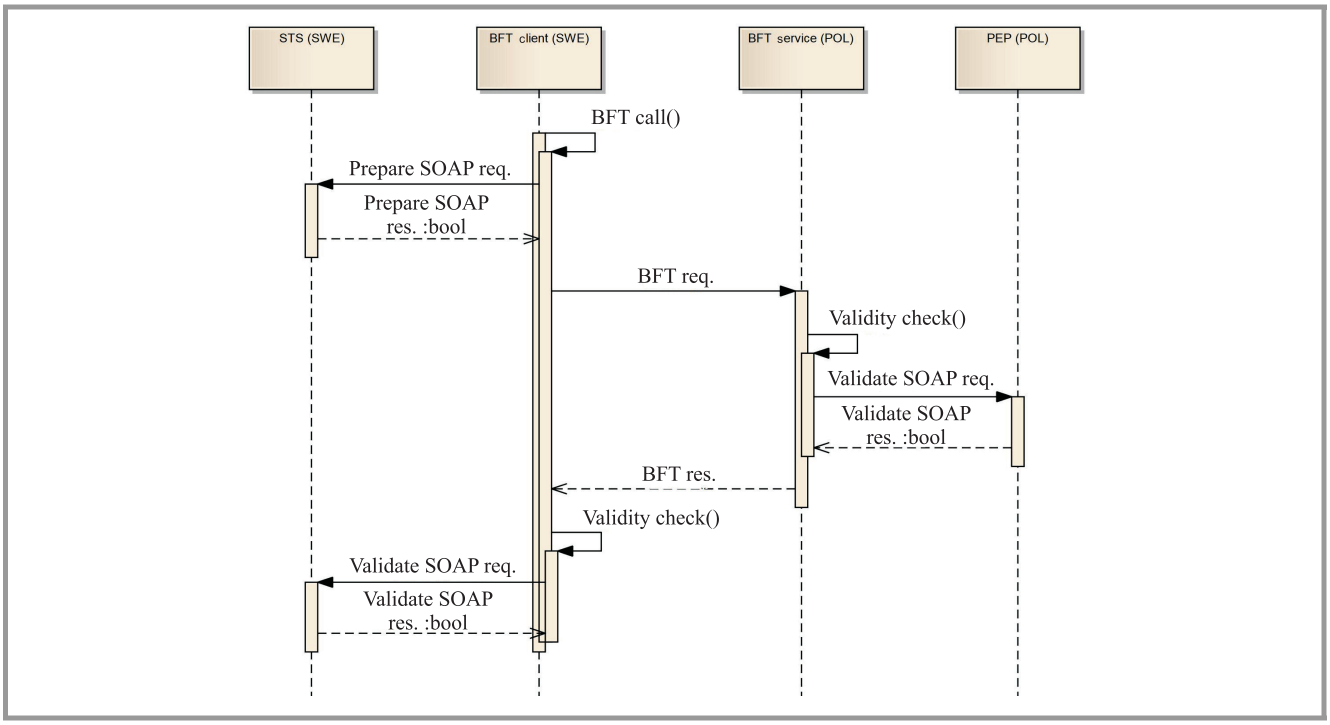


Fig. 9. Basic test scenario sequence diagram.

teams agreed on the solution and set the SAML assertion specification for the purpose of the trial. The implementations of PEP and STS security modules were independent, without the necessity of sharing any piece of code.

4. Way Forward, Further Works

The solution presented in the article provides basic cross-domain authentication allowing to make the authorization decisions locally, in the domain of the service provider. It relies on the trust between cooperating domains and supports the following security goals:

– authentication in the domain of the user which can be verified in the domain of the service;

– local authorization in the domain of the service;

– integrity of the SOAP messages;

– non-repudiation of sending the message.

The work under cross-domain security is a broad subject that should be deeply analyzed in multinational communities and experimented frequently in order to find the most suitable solution, acceptable for every nation, easy for implementation and interoperable in heterogeneous environment. The work under these issues in MCI is being continued to provide coherent solution for cross-domain secure information exchange in SOA-based dynamic environment. Currently, we are at the stage of making arrangements for tests of security solutions with the NC3A Core Services Testbed, which will be carried out in 2010.

The first step forward is to test the cross-domain authentication of users which utilize different intradomain authentication methods, e.g., login/password, Kerberos and biometric data. These authentication methods should be appropriately identified by the authorization service and reflected in the local security policy.

When dealing with the web services security, it needs to be emphasised that the publish-subscribe mode of operation should also be reflected. Since in this case the roles of the service and the client are reversed (the client has the service interface listening for notifications, the service has the client side sending notifications) validity of the assertion and the user certificate should be also checked during the subscription period.

Another issue worth considering, and very much visible in the web service environment is the concept and implementation of service chaining and single-sign-on mechanisms. Service chaining is being used more and more often in service orchestration and is an intrinsic feature of the SOA-based world. In order to guarantee appropriate authentication, authorization, integrity and non-repudiation of sending the message, there needs to be a mechanism providing the identity delegation in the chain of services allowing to recognize the initiator of the process and grant the access to the service based on his credentials.

In the area of reliability of the information exchange the need for combining the mechanism providing security and quality of service must be emphasized. In this area the work is conducted in scope of:

– joint security and QoS policy for application and network layer resources access control,

– dynamic policy modification according to changing conditions, e.g., risk evaluation, detected threats and changing situation,

– access control rules (policy) negotiation between autonomous systems using XACML before they start information exchange.

Acknowledgment

This work was partially supported by Research Project PBZ – MNiSW – DBO-02/I/2007.

References

[1] NATO Network Enabled Feasibility Study Volume II: Detailed Report Covering a Strategy and Roadmap for Realizing an NNEC Networking and Information Infrastructure (NII), version 2.0.

[2] *Guide to Secure Web Services*. NIST Special Publication 800-95, August 2007.

[3] NATO Network Enabled Feasibility Study Volume I: Overview of the NATO Network-Centric Operational Needs and Implications for the Development of Net-Centric Solutions, version 2.0.

[4] "NEC Security Research Strategy", RTO-TR-IST-045 Tech. Rep.

[5] ITU-T Recommendation X.805 – Security architecture for systems providing end-to-end communications.

[6] "J. J. Brennan Information Assurance for SOA", The Mitre Corporation, 2009.

[7] RFC 5280: Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile, IETF, 2008.

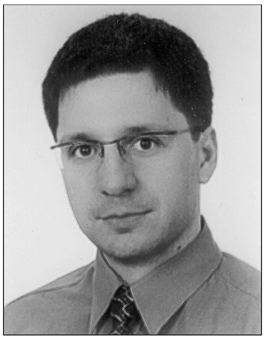
[8] "Assertions and Protocols for the OASIS, Security Assertion Markup Language, (SAML) V2.0, OASIS Standard", 15 March 2005 [Online]. Available: http://docs.oasis-open.org/security/saml/v2.0/saml-core-2.0-os.pdf

Bartosz Jasiul was born in 1975 in Zgorzelec, Poland. He graduated from the Faculty of Cybernetics of Military University of Technology (1998). At the moment he is a scientist in Military Communication Institute in Zgorzelec, Poland. His main areas of interest are: security, public key infrastructure, network enabled capability security, MANET networks, service oriented architecture, web services.

e-mail: b.jasiul@wil.waw.pl
Military Communication Institute
Warszawska st 22A
05-130 Zegrze, Poland

Rafał Piotrowski was born on July 1972 in Radom, Poland. He received the M.Sc. degree from Military University of Technology, Warsaw, in 1997 and the Ph.D. degree from Wrocław University of Technology, in 2001. He served in Polish Army as an Engineer Officer from 1997 to 1999. From 1999 he works for Military Communication Institute, Zegrze, Poland. His research interests include various aspects of communication: digital signal processing, compatibility, network security.

e-mail: r.piotrowski@wil.waw.pl
Military Communication Institute
Warszawska st 22A
05-130 Zegrze, Poland



Robert Goniacz was born on March 1972 in Chełmno, Poland. He graduated from the Faculty of Telecommunication of Military University of Technology (1997). From 1997 he has been working as a scientist in Military Communication Institute, Zegrze, Poland. His main research interests are: IT network management and security, object-oriented programming techniques, GIS techniques and platforms, service oriented architecture, web services.

e-mail: r.goniacz@wil.waw.pl
Military Communication Institute
Warszawska st 22A
05-130 Zegrze, Poland

Joanna Śliwa – for biography, see this issue, p. 53.

Paper

Web Services Efficiency in Disadvantaged Environment

Joanna Śliwa[a], Tomasz Podlasek[a], and Marek Amanowicz[a,b]

[a] *Military Communication Institute, Zegrze, Poland*
[b]*Military University of Technology, Warsaw, Poland*

Abstract—The article presents results of web services (WSs) efficiency tests carried out in the testbed emulating disadvantaged network environment. The authors discuss the advantage of different WS adaptation techniques that allow to minimize the XML message size (i.e. compression, filtering and binary coding) and the size of JPEG image attachment (i.e., resolution reduction, decreasing colour depth, JPEG compression). The presented results show the efficiency of selected methods that adapt the web services realization to the possibilities of the network. The article is summarized by conclusions and recommendations in terms of sending XML SOAP messages in disadvantaged networks.

Keywords—compression, filtering, image resize, SOA in tactical networks.

1. Introduction

Service oriented architecture (SOA) is one of the crucial enablers to achieve network enabled capability (NEC) and mission effectiveness [1]. It provides means for information sharing among various elements of federation of systems (FoS). It also offers services to arbitrary service consumers, meets their information needs, takes part in a business process carried out within military operation and thus, supports achieving the net-centricity.

Many of the countries have chosen web services (WS), as the most interoperable and easily extendable SOA platform, widely used in commercial applications and supported by the biggest and most advanced software developer companies in the world. WSs are described by a wide range of standards that deal with different aspects of services realization, transport, orchestration, semantics, etc. They provide means to build a very flexible environment that is able to dynamically link different system components to each other. These standards are based on the extensible markup language (XML) that was developed for the enterprise systems and operates in high bandwidth links. XML became very popular because it solves many interoperability problems, even though it adds a significant overhead.

The utilization of web services in SOA-based systems operating in NEC environment has been shown in many international experiments (e.g., coalition warrior interoperability demonstration (CWID) [2], currently CWIX). They prove that the WS technology improves collaboration, interoperation and information sharing in federation of systems (FoS). In order to achieve efficient information exchange between the users and to give the biggest advantage to the operators, the SOA solutions need to work with different types of information and communication systems. The challenge is though, to use this simple in concept and providing big flexibility means of communications on all command levels, including low bandwidth tactical communications systems.

Tactical network environment used by the lowest echelons of command has many limitations that make it difficult to provide reliable communications. It usually copes with high error rates, intermittent connectivity problems, radio silence and frequent disruptions. It also often changes its topology. Data rates of the low tens of kilobits per second and below are common.

Given their maturity level and the software tool support, web services would be the best choice for all command levels. However, the above-mentioned factors characterizing tactical disadvantaged environment provide limitation for the XML-based SOA information distribution mechanisms. The application of web services in tactical systems has been though, subject to experiments. According to the NATO C3 Agency (NC3A) report on using WS in tactical domain [3], web services continue to function adequately even at the lowest levels of network capacity, although their performance is diminished. They remain promising even in low bandwidth links, as long as very fast response times are not required.

In order to fully take advantage of SOA and web services methods for adapting WS realization in tactical networks must be provided. It is, though, very important to discover the level of WS applicability in a disadvantaged environment and to investigate the efficiency of different optimization mechanisms that enable the performance of simple XML SOAP communications to improve.

The remainder of this paper is organized as follows: Section 2 presents the test scenario and the results of tests that were carried out by our team, and discusses the possibilities of WS realization optimization methods; Section 3 presents the summary and recommendations for SOA solutions in low bandwidth tactical environment.

2. SOA Web Services Efficiency Tests' Results

The objective of the tests was to discover the edge network parameters below which web services cannot be realized. As the web services provider, the authors used own implementation of the blue force tracking (BFT) service that sends unit tracking information consistent with the STANAG 5527 NFFI (NATO Friendly Forces Information) standard [4]. This service was augmented with the possibility to send image sensor information (JPEG files). The authors focused on verifying the efficiency of:

- pure SOAP message exchange,
- SOAP message filtered out (allowing to send only obligatory NFFI information about the unit),
- compressed SOAP message exchange,
- exchange of JPEG images of different sizes.

2.1. SOAP Messages Filtering

The first experiments were focused on showing how the filtering (extracting only obligatory NFFI XML schema data) and the Gzip compression of NFFI messages can decrease the time needed to receive response from the service. The authors assumed that a response SOAP message consists of 30 objects.

BFT service based on NFFI can send tracks of objects. The track following NFFI 1.1 XML schema contains 3 types of information about the object:

- positional data – obligatory: track source, data/time, coordinates; optional: bearing, speed, reliability, inclination;
- identification data (optional) – unit symbol, unit short name;
- operational status data (optional) – footprint, strength, status code, alert, remarks.

Filtering is a process of cutting out all optional information elements of the NFFI message from the original message (so called "long track"), leaving only the obligatory ones (so called "short track"). This process results in limiting the SOAP message size (see Fig. 1).

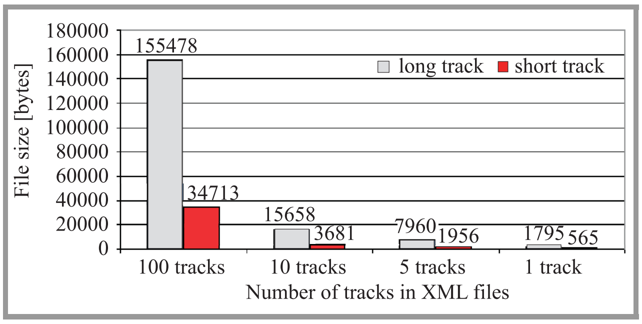


Fig. 1. Filtering gain.

The results of the experiment show that the use of filtering with Gzip compression remarkably reduces the size of SOAP messages and makes it possible to exchange NFFI messages properly through low bandwidth networks. By using compression we are able to provide information exchange in circumstances where the available bandwidth is about 2 kbit/s. Simultaneously, the time when the application receives a response was decreased below 7 s (see Fig. 2).

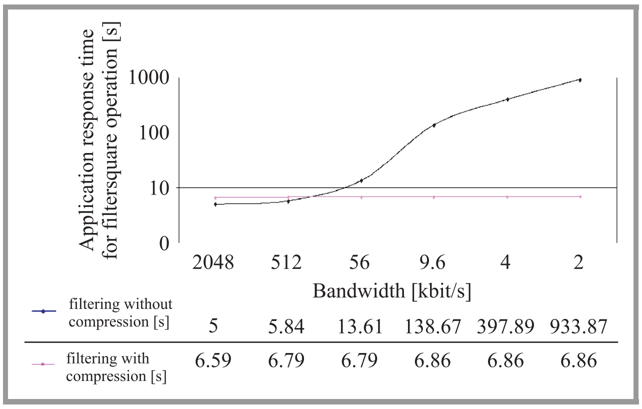


Fig. 2. Application response time for mediation service operation (in seconds) of available throughput.

As depicted in Fig. 2, the same information sent without using compression through links with low throughputs takes much more time (application response time is about 934 s). It is unacceptable in case of exchanging short-life information (e.g., quickly changing positions of forces, sensor information). The results show that it is recommended to use the compression when the traffic flows trough tactical, low bandwidth networks, particularly where the available throughput is below 56 kbit/s.

2.2. Gzip Compression and Efficient XML

Standard web services traffic bases on the exchange of SOAP messages usually sent in the request-response mode. The most popular and widely supported by SDKs SOAP binding is HTTP over TCP/IP. This protocol stack performs well in a stationary high bandwidth networks and over the Internet. However, its performance in a tactical environment is usually diminished [3]. In order to evaluate the level of web services performance in a disadvantaged environment, the authors carried out tests of "pure" SOAP communications and compression/binary encoding techniques that significantly reduce the message size. The web service producer was the Blue Force Tracking service.

In general, we distinguish "lossy" and lossless compression techniques. In case of a data transmission system, designers are interested in using only lossless ones. These compression algorithms usually exploit statistical redundancy in such way, as to represent the sender's data more concisely without error. The lossless compression uses the fact that most of the real-world data has statistical redundancy. However, lossless data compression algorithms will always fail to compress data that has high entropy (high disorder). This applies to already compressed data, random data or encrypted data. In such cases the attempts to compress data will usually result in an expansion.

One of the most popular and efficient lossless compression algorithms is Gzip[1], which was created as a fusion of two algorithms: Lempe-Ziv (LZ77) and Huffman Coding. It is based on the DEFLATE algorithm, which was designed to replace LZW and other patented data compression. It finds duplicated strings in the data to be compressed. The second occurrence of a particular string is replaced by a pointer to the previous one, in form of a pair represented by the distance and length. Distances are limited to 32 kB and lengths are limited to 258 B. When a string does not occur anywhere in the previous 32 kB, it is represented as a sequence of accurate bytes.

This characteristic of Gzip indicates that it can be satisfactory only with data that has some order. This makes it inappropriate, e.g., for already compressed files and the encrypted ones. In the tests, the authors were using this algorithm to compress SOAP messages – a text document in XML format. For this data type, Gzip compression achieves a high gain. However, for other data formats, e.g., already compressed JPEG images, which can be also sent as attachments to SOAP, it does not yield satisfactory results.

In Fig. 3, there has been depicted the message size comparison for different number of NFFI tracks sent in original XML SOAP messages and in SOAP messages compressed with Gzip.

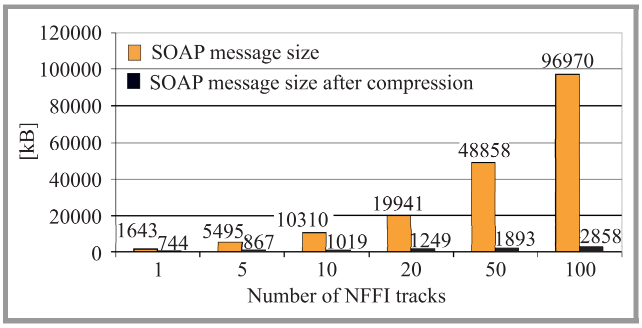


Fig. 3. SOAP message size comparison – original SOAP with compressed SOAP message.

It is clear that the bigger the message is – the best performance we get. This performance can be described as compression gain (CG) which the authors define as follows:

$$CG = 100 - \frac{\mathit{size\_of\_compressed\_message}}{\mathit{size\_of\_original\_message}}\,100 \qquad (1)$$

Compression gain can be, though, even 97% for 100 kB XML SOAP message, but reaches only 54% when the original message size is 1.6 kB (see Fig. 4).

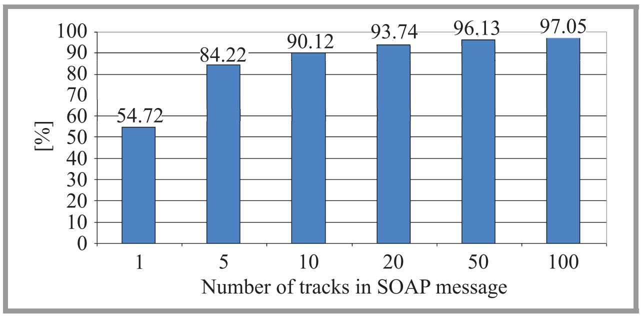


Fig. 4. Compression gain.

It should be noted that the process of compressing the message is resource consuming and takes considerable amount of time. On the server with processor with 4 cores (2.8 GHz) it took even 30 ms for the biggest 100 – track message (see Fig. 5).

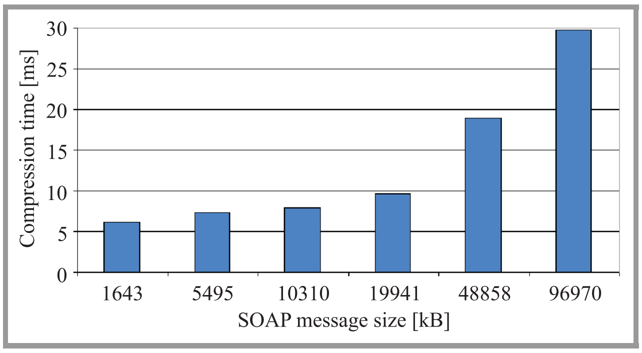


Fig. 5. Compression time for different sizes of SOAP message.

As compression is a resource consuming process, it has been proven that in an Ethernet with 100 Mbit/s throughput, it does not give considerable effects. The compression gain resulting from minimizing the message size is limited by the fact that the machine needs time for compressing and decompressing the message (see Fig. 6). The application response times for a compressed and a not compressed SOAP message transmission are though very similar when the network links are quite fast and reliable.

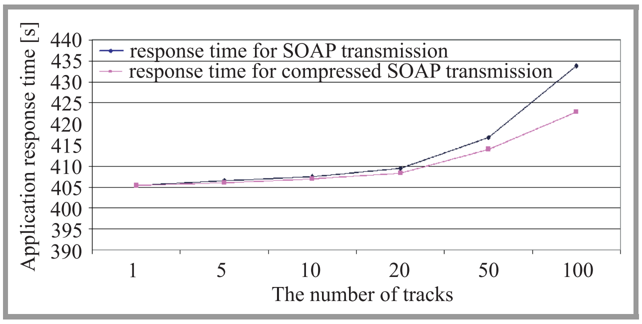


Fig. 6. Application response time for original SOAP transmission and with Gzip compression.

The true advantage of compression can be seen when the network is degraded with limited throughput, high error

[1] More information can be found Gzip home page, http://www.gzip.org/

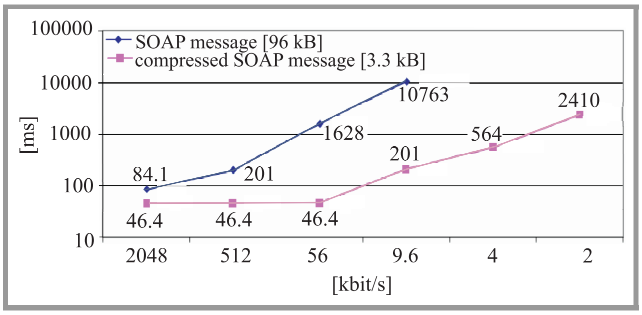


Fig. 7. Application response time in disadvantaged network for compressed and original 100 – track SOAP message. Network parameters: changing throughput (2048 kbit/s, 512 kbit/s, 56 kbit/s, 9.6 kbit/s, 4 kbit/s, 2 kbit/s), packet error rate = 0, delay = 0.

rates and high delays. The authors carried out tests of Gzip compression in an emulated disadvantaged environment, where links were configured by variable values of the throughput, delay and packet error rate (PER) (see Fig. 7, Table 1).

Table 1
Minimum throughput requirements at different values of per and delay

	PER [%]	Delay [ms]	Minimum throughput [kbit/s]
SOAP without compression	≤ 10	≤ 1000	9.6
	≥ 25	≥ 100	512
SOAP with compression	≤ 10	≤ 100	2
	≤ 10	> 100	4
	≥ 25	> 1000	56
	≥ 50	0	2048

In a network with very good quality parameters (PER = 0, delay = 0), an original 100 – track SOAP message could not be sent in a network with throughput equal 4 and 2 kbit/s. For 9.6 kbit/s, the application response time amounted to about 11 seconds, however, when the network parameters turn to degrade, this time is being extended (e.g., 134 s for delay = 100 ms, PER = 0). In comparison, a compressed SOAP message is transmitted in 2.56s for 9.6 kbit/s link, 5.62 s for 4 kbit/s link and in 28.8 s for 2kbps link, through links with the same QoS.

The results (see Table 1) showed that SOAP without compression can be sent through links not slower than 9.6 kbit/s, with PER $\leq$ 10% and delay $\leq$ 1000 ms. When PER $>$ 10%, the throughput must not lower than 512 kbit/s with a delay not lower than 100 ms. Compressed SOAP can be sent by 2 kbit/s links with PER $\leq$ 10% and delay $\leq$ 100 ms. If the delay is higher (over 100 ms), the throughput should amount to at least 4 kbit/s. When PER $>$ 10%, the lowest throughput is 56 kbit/s (delay $\leq$ 1000 ms). Sending compressed SOAP through a very bad link, where PER $\geq$ 50% can be done with the lowest throughput of 2048 kbit/s.

Apart from using the Gzip compression algorithm, it is possible to reduce the size of SOAP message with other mechanisms, including an additional efficient coding, like Efficient XML (EXI) [5], [6] or Fast Infoset.

Fast Infoset specifies a standardized binary encoding for the XML information set [7]. It uses the existing and proven ASN.1 standards. The specification is standardized as an ITU-T Recommendation within ITU-T Rec. X.891 [8] and ISO/IEC 24824-1 [9].

Efficient XML is a specification of binary coding of the XML data. EXI is a very compact representation of the XML information set that is intended to simultaneously optimize the performance and the utilization of computational resources. The EXI format uses a hybrid approach drawn from the information and formal language theories, plus practical techniques verified by measurements for entropy encoding of the XML information. To efficiently encode XML event streams, the EXI format uses a relatively simple algorithm, which is acquired for a fast and compact implementation, and a small set of data type representations. The EXI specification consists of the grammar production system and the format definition. EXI is compatible with XML at the XML information set level, rather than the XML syntax level. It permits to encapsulate an efficient alternative syntax and grammar for XML, while facilitating at least the potential for minimizing the impact on XML application interoperability. EXI is schema-"informed", which allows utilizing the available schema information to improve the compactness and performance. It also uses a grammar-driven approach that achieves very efficient encodings.

Table 2
Compression gain and data processing time using Gzip, FI, EXi and EXI with compression for different message sizes

Compression mechanism	SOAP message size					
	0.6 kB		12.3 kB		406 kB	
	CG [%]	RTD [ms]	CG [%]	RTD [ms]	CG [%]	RTD [ms]
XML	–	2.2	–	2.8	–	9
Gzip	49.3	1.7	95.3	2.5	98.2	19.5
FastInfoset	30	1.9	53.5	3	55.2	11.9
EfficientXML	49.7	0.23	93.2	0.9	94.4	3
EfficientXML + compression	71.8	0.44	97.4	1.7	99	6.8

The subject of the tests was the efficiency of these compression techniques. The MCI team also investigated this problem in terms of using: an open source Gzip compression algorithm, a binary form of XML Fast Infoset and EXI (the EXIficient implementation[2]). The authors tested three different sizes of SOAP messages (0.6 kB, 12.3 kB

[2] More information can be found on EXIficient web site, http://exificient.sourceforge.net/

and 406 kB) in a simple web service, and measured the compression gain (CG) and the round-trip delay (RTD) time (see Table 2). The verification proved the high gain in compressing XML messages more efficient with large files. The most efficient compression mechanism is binary form of efficient XML, especially with compression, which, for big amounts of data, achieved compression gain equal 99%.

2.3. Image Resolution Change

Apart from XML, SOAP messages can easily carry image attachments. In the emulated environment, the authors also tested the efficiency of reducing JPEG images resolution. WANem application was used to emulate high error rates, delay and PER values. Three different JPEG images were sent (JPEG image with original size 486 kB and its two smaller equivalents –50% and 10% of it original size) (see Table 3).

Table 3
Minimum throughput requirements for images and their transmission time at different values of PER

JPEG (2706 × 3657)	PER [%]	Minimum throughput [kbit/s]	Transmission time [s]
JPEG 486 kB (100%)	0	56	80–100
JPEG 240 kB (50%)	≤ 10	9.6	250–299
JPEG 58 kB (10%)	≤ 10	9.6	62–87

The results show that it is possible to send still images in low throughput channels, however with significant value of the transmission time. Reducing their resolution can decrease this time (see Table 3, Fig. 8).

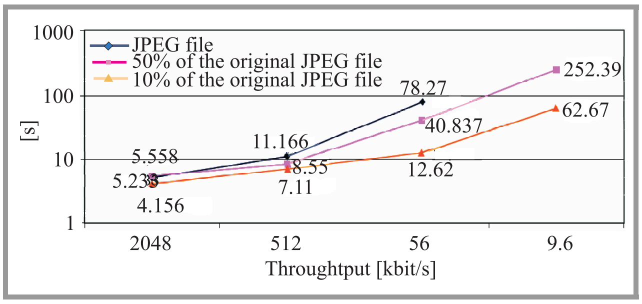


Fig. 8. Application response time for different image sizes. Network parameters: changing throughput (2048 kbit/s, 512 kbit/s, 56 kbit/s, 9.6 kbit/s), packet error rate = 0, delay = 0.

In the tests, reducing the image size was decreasing their resolution (see Fig. 9). However, image modifications can also include decreasing the colour depth and decreasing the image quality (inherent in the JPEG coding). Yet, changing the colour scale to greyscale in fact does not give much gain (673 kB of original size to 565 kB in the greyscale – see Fig. 10).

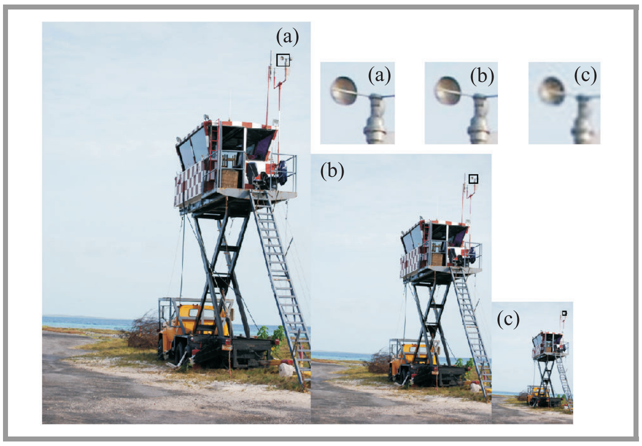


Fig. 9. Effect of reducing the image resolution for three dimensions 2706 × 3657 (a); 1913 × 2585 (b); 855 × 1156 (c).

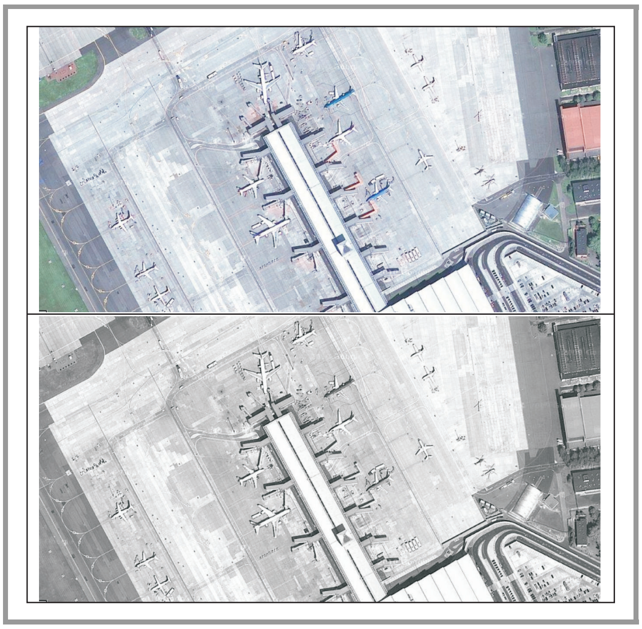


Fig. 10. Changing the original JPEG colour depth to greyscale.

In contrast, the reduction of the JPEG image quality can give surprisingly good results. The most common and widely supported one is a lossy JPEG compression inherent to the nature of this image coding method. The degree

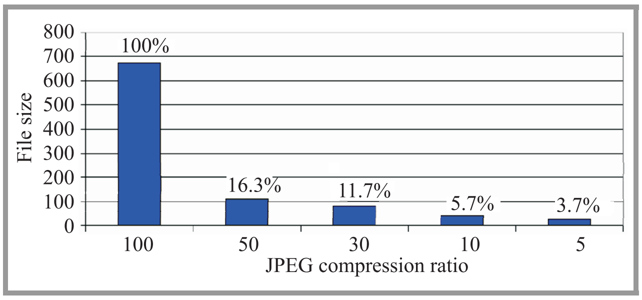


Fig. 11. JPEG compression ration versus image file size.

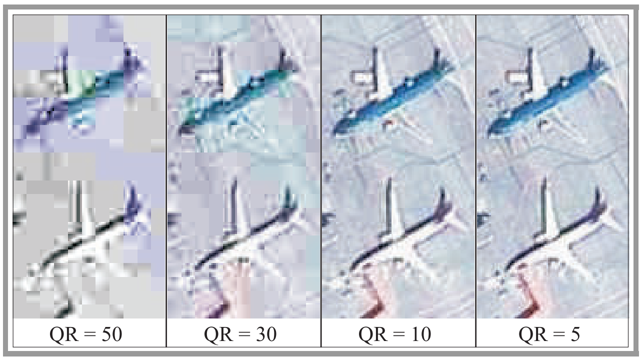


Fig. 12. Image quality reduction for a segment of an image and different values of JPEG quality ratio (QR).

of the compression can be adjusted, allowing a selectable tradeoff between the storage size and the image quality. In fact, decreasing the image quality, significantly reduces its size. Compression ratio equal 50 can reduce the size at over 83% of the original size (see Fig. 11), not degrading significantly the subjective image quality that can be assessed by the viewer. This is most visible when the user wants to zoom in a particular segment of the image where the details are most important (see Fig. 12).

3. Summary and Recommendations

The tests carried out by the authors proved that adaptation of the SOAP message by compression and modification of the image attachment (if present) can reduce the size of this message and improve its transmissions performance in a disadvantaged environment. The message size reduction can be achieved by the use of Gzip compression run in the application server at the HTTP level. This is a method very simple to use and widely supported by the application servers. As shown, compression is a time consuming process but its application during communication over links that have 56 kbit/s and below is highly recommended.

There should be also considered the application of binary coding techniques, especially the EXI, that provide good efficiency both in terms of compression gain and processing time. The measured web service response time for EXI was shorter than for Gzip, but the authors expected a greater difference for EXI solution. It may be caused by some deficiency in the used open source EXI implementation [9]. The measured compression ratio proved that EXI can significantly limit the size of SOAP messages (95% for medium messages), however, Gzip has also very good results (94% for medium messages). FI binary encoded messages gave worse results (53% for medium messages). The results proved that the Efficient XML standard is very promising, however, its efficiency strongly depends on the implementation.

Apart from tracking information, units located at the lowest command levels need recognition of the surrounding terrain. The sensor image services, though, need to be provided also through the disadvantaged tactical networks.

The results of the tests provided in this article prove that sending high resolution big images through disadvantaged networks is related to a long transmission time (79 s for 56 kbit/s, impossible transmissions in 9.6 kbit/s network, almost 1 minute for the smallest file experimented). It is necessary to establish other methods for sending image sensor information through disadvantaged low bandwidth links. These can be fragmentation into smaller parts, or, e.g., decreasing the quality of the image. This latter one is very efficient, however only for a user that does not need to zoom in the image to see it in great details.

The presented results show the efficiency of selected methods that adapt the web services realization to the possibilities of the network. The original service producers can be equipped with such a functionality or the WS traffic can be redirected through a content proxy that would help to improve the communication at the lowest command levels. What remains to be done is the mechanism that would limit the impact of connection losses and other connectivity problems.

Acknowledgment

This work was carried out within the project PBZ – MNiSW – DBO-02/I/2007 supported by the R&D funds allocated for the years 2007-2010.

References

[1] M. Booth, T. Buckman, J. Busch, B. Caplan, B. Christiansen, R. van Engelshoven, K. Eckstein, G. Hallingstad, T. Halmai, P. Howland, V. Rodriguez-Herola, D. Kallgren, S. Onganer, R. Porta, C. Shawcross, P. Szczucki, and K. Veum, *Detailed Report Covering a Strategy and Roadmap for Realizing an NNEC Networking and Information Infrastructure*, vol. II of *NATO Network Enabled Feasibility Study*, version 2.0, NATO C3 Agency, June 2005.

[2] R. Haakseth, T. Gagnes, D. Hadzic, T. Hafsøe, F. T. Johnsen, K. Lund, and B. K. Reitan, "SOA – Cross Domain and Disadvantaged Grids", NATO CWID 2007, FFI-rapport 2007/02301, Norwegian Defence Research Establishment.

[3] G. Babakhani, J. Busch, C. Dumas, R. Fiske, B. Holden, H. Lægreid, R. Malewicz, D. Marco-Mompel, V. Rodriguez-Herola, "Web trends and technologies and NNEC core enterprise services, version 2.0", Technical Note 1143, NATO C3 Agency, The Hague, Dec. 2006.

[4] "Interim NFFI Standard For Interoperability of FTS", AC322(SC5)N(2006)0025, NC3B Information Systems SC, 16 Dec. 2006.

[5] M. Cokus, T. Kamiya, "Efficient XML Interchange Working Group Public Page" [Online]. Available: http://www.w3.org/XML/EXI/

[6] J. Schneider, T. Kamiya, "Efficient XML Interchange (EXI) Format 1.0", W3C Proposed Recommendation, 20 Jan. 2011 [Online]. Available: http://www.w3.org/TR/2011/PR-exi-20110120/

[7] J. Cowan, R. Tobin, "XML Information Set (Second edition)", W3C Recommendation, 4 Febr. 2004 [Online]. Available: http://www.w3.org/TR/xml-infoset/

[8] "Information technology – Generic applications of ASN.1: Fast Infoset", ITU-T Rec. X.891, May 2005.

[9] "Information technology – Generic applications of ASN.1: Fast Infoset", ISO/IEC 24824-1, May 2007.

[10] K. Lund *et al.*, "Using Web Services to Realize Service Oriented Architecture in Military Communication Networks", Norwegian Defence Research Establishment, *IEEE Commun. Mag.*, pp. 47–53, 2007.

Tomasz Podlasek was born in 1984 in Grójec, Poland. He graduated from the Faculty of Electronics and Telecommunications Technology of Military University of Technology (2009). At the moment he is a researcher in Military Communication Institute in Zegrze, Poland. He is working in the area of efficient realization of web services in disadvantaged networks. His main areas of interests are: Java programming, advanced programming technologies, data structures, semantic technologies and web services.
e-mail: t.podlasek@wil.waw.pl
Military Communication Institute
Warszawska st 22A
05-130 Zegrze, Poland

Joanna Śliwa, Marek Amanowicz – for biographies, see this issue, p. 53.

Paper

Adaptation of the Kademila Routing for Tactical Networks

Tobias Ginzler[a] and Marek Amanowicz[b,c]

[a] Fraunhofer FKIE, Wachtberg, Germany
[b] Military Communication Institute, Zegrze, Poland
[c] Military University of Technology, Warsaw, Poland

Abstract—In this paper a modification of the widely used Kademlia peer-to-peer system to tactical networks is proposed. We first take a look at the available systems today to cover the range of possibilities peer-to-peer systems offer. We identify candidates for use in military networks. Then we compare two candidate systems in an environment with highly dynamic participants. The considered environment is focused on the special conditions in tactical networks. Then we give rationale for choosing Kademlia as a suitable system for tactical environments. Since Kademlia is not adapted to military networks, a modification to this system is proposed to adapt it to the special conditions encountered in this environment. We show that optimizations in the routing may lead to faster lookups by measuring the modified algorithm in a simulation of the target environment. We show also that the proposed modification can be used to extend the battery lifetime of mobile peer-to-peer nodes. Our results show that peer-to-peer systems can be used in military networks to increase their robustness. The modifications proposed to Kademlia adapt the system to the special challenges of military tactical networks.

Keywords—Kademlia, network enabled capabilities, peer-to-peer, wireless tactical military networks.

1. Introduction

Today peer-to-peer applications and protocols have gone far beyond the notorious file sharing. Applications like remote assistance search, distributed data storage or VoIP systems like Skype make use of peer-to-peer (P2P) systems. Starting with only a handful of protocols, an overwhelming variety of systems for quite every imaginable purpose has been developed. Peer-to-peer networks span the globe and consist of hundreds of thousands concurrent participants. Despite the success in civilian applications, no broad use in military applications is known yet. Especially the resilience of peer-to-peer networks is able to increase the availability of military communication infrastructure. Centralized networks have a single point of failure and facilitate effective adversary actions against the network. Peer-to-peer systems offer a distributed approach contrary to traditional server-centric architectures. We show that peer-to-peer systems exist which are able to work under difficult network conditions encountered in military network environments. Until now peer-to-peer networks have focused on wired infrastructure. In military environments, not only wired networks but also a large variety of wireless networks with mobile devices is used. An adaption of the broadly used Kademlia peer-to-peer system is proposed to adapt it to the military environment. The communication devices in the considered environment have to cope with limited CPU power, small bandwidths, high delay and many connection disruptions due to the nature of the wireless medium and their mobility. According to the network enabled capabilities (NEC) principle it is necessary to interconnect the closed legacy networks of today. An adapted peer-to-peer network available today may significantly improve the availability of the right information at the right place at the right time.

2. P2P Systems and Solutions

An overview of the state of the art of peer-to-peer systems is given in the following section. The scope of the overview of existing peer-to-peer systems is limited to the usage purpose of the system. It should offer a search functionality to find information elements which were previously stored in the network and a method to retrieve them. The network should be scalable so that thousands or even millions of participants can take part without a degradation of service. This respects the fact that in the NEC concept, sensors and systems may also be equipped with information technology resulting in a potentially huge number of network participants. The peer-to-peer system – more specific – the overlay network infrastructure employed by the system, should be resilient against network failures and the unexpected failure of participating nodes. We identify structured overlay systems to be the most promising as they have advantages regarding resilience against attackers and networks failures.

2.1. Early P2P Systems

Peer-to-peer systems which rely on a dedicated server [1] for search or login purposes have disadvantages. Such systems are a single point of failure. Load issues also render this approach unsuitable in a mobile environment. The so called unstructured overlay networks overcome the dependency on servers but searching for content is more difficult in such networks. In server based architectures searching and indexing is trivial and may be enriched with range queries or semantic search. A simple approach to find content in an overlay network without a server is to flood it with a search query [2]. This imposes heavy load on the network. If the search is limited to a fixed number of hops to increase scalability, the search may fail even if the content exists. The search is then considered *incomplete*. Super-peer networks are an alternative to flooding networks. All following approaches are considered to be *complete*, meaning the search succeeds if the content is available in the network or fails if it has not been stored.

2.2. Super-Peer Networks

Super-peer networks use a two-tier architecture. Long-lived or high-bandwidth nodes are declared as supernodes, while other nodes are declared as ordinary nodes. Super-peer networks form clusters of peers around supernodes (Fig. 1). The supernode answers search and storage requests on behalf of its connected peers. Every ordinary peer has to be connected to a supernode. The supernodes communicate by a dedicated protocol.

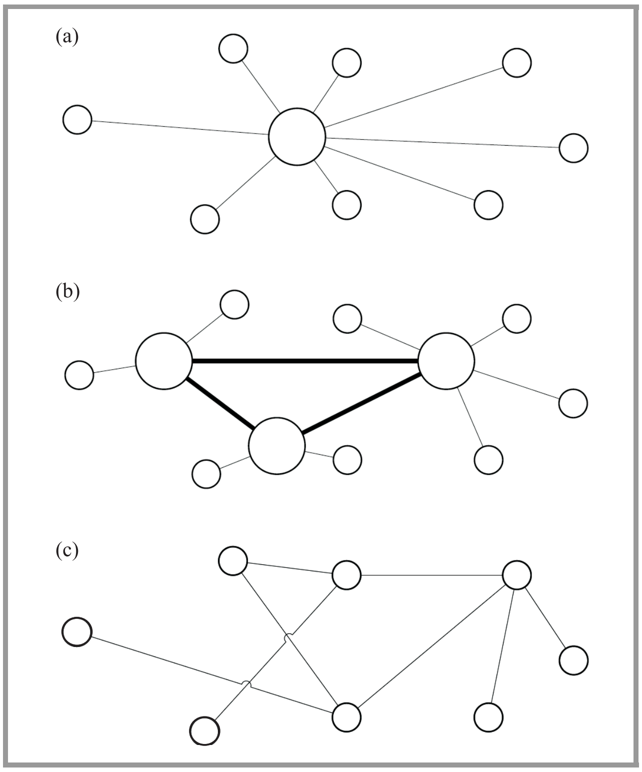


Fig. 1. Different schemes of information exchange: (a) client-server; (b) super-peer; (c) peer-to-peer.

Super-peer networks are used in Skype and FastTrack-based networks. FastTrack is believed to use a controlled flooding algorithm among the supernodes to handle overlay network updates and search requests [3]. Flooding is done also by the 0.6 protocol version of Gnutella [4]. Still, flooding the supernodes is a comparatively inefficient method to search for content in an overlay network. Existing implementations show an increased robustness to churn compared to some unstructured flooding-based overlay networks [5]. Churn in context of peer-to-peer systems denotes the process of members joining the network and leaving it. Churn may either be caused by network failures or user behavior. The startup of nodes, or *bootstrapping*, is more difficult than in pure peer-to-peer systems, as nodes need to find a supernode first. Bogus clients can obtain supernode status by fraud and cause more damage to other clients than a normal peer in an decentralized network. Users may also try to prevent to be elected a supernode to save bandwidth and computational power, increasing the load on the remaining supernodes. Every decentralized system can be transformed into a supernode network by defining the supernode's cluster as a single member of the decentralized network [6].

2.3. Structured Overlay Networks

Today peer-to-peer networks can grow to impressive size [7]. Structured overlay networks were designed to support a very large number of participants. The largest existing overlay is based on the Kademlia structured overlay and is named KAD [8].

Structured networks use a key space where peers are placed in and searching for a node in the key space then follows a (structured) routing algorithm. Each peer carries a unique identifier, defining its position in the key space. Kademlia is based on a structured peer-to-peer overlay network [9]. In Kademlia an XOR metric is introduced to define a distance between two nodes. The XOR distance is the bitwise exclusive OR on the peers' identifiers interpreted as an integer. The other important metrics used by other protocols are the prefix-based metric used by Tapestry [10], the ring metric of Chord [11] and the combined prefix/proximity metric of Pastry [12].

The routing scheme is similar for all structured peer-to-peer systems. The overlay routing is responsible for finding nodes according to their identifier (key) in the overlay. This is called *key based routing* (KBR). The routing algorithms differ but they share the principle of approaching the destination key in every routing hop and terminating at the closest node.

A simple store and retrieve functionality can be supplied by a distributed hash table (DHT) on top of the KBR. The DHT facilitates to store information into the overlay and retrieve information from the network. The application programming interface (API) is similar to a standard hash table. The idea is to attach a key to every piece of information which has to be stored in the overlay. The key is often derived by hashing the representation of the information.

The information is then routed and stored at the r nodes with the identifiers closest to the key of the information, with r as a redundancy parameter. At these locations, the information can also be found by other nodes. Any node looking for the information calculates the key from it and uses the key based routing for finding the node the data is stored at (DHT GET). The DHT is described as an integral part of Kademlia, but it is possible to deploy a DHT on top of every overlay network with key based routing.

As an example of a structured peer-to-peer routing scheme we take a closer look on Kademlia. The routing table of Kademlia is a binary tree (Fig. 2). Each leaf contains a list of nodes, the so called *buckets*. A bucket holds a fixed number (k) of references to reach other nodes. As the network may contain up to 2^b nodes, the routing table size has to be limited. In Kademlia the memory requirement for the routing table is $O(k \cdot b)$, with b as the number of bits of an identifier. A node carries a tag which defines which identifiers are contained in its subtree. In the figure, b and k are assumed to be 4, the standard key length of Kademlia is 160. A tag of $1xxx$ means that the highest value bit is 1 for the whole subtree and the other bits are unknown. The right subtree of the root carries this tag. The local node identifier is assumed to be 0000 in the depicted tree.

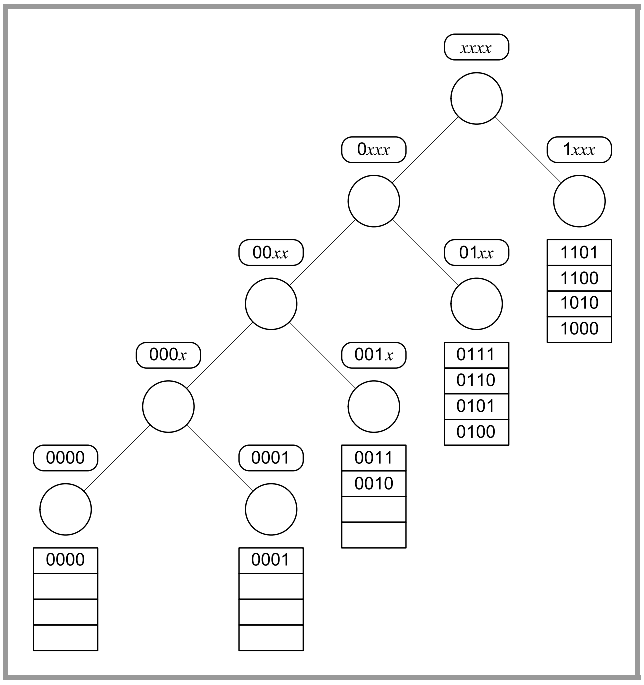


Fig. 2. Kademlia routing table.

The two subtrees below the root separate the identifier space in two halves: one subtree contains references to nodes closer to the local node than half of the maximum distance (the left side) and the other one contains references to nodes further away (the right side). The rightmost bucket holds k references to nodes which differ in their most significant bit from the identifier of the local node. The left subtree is constructed recursively with increasingly matching prefix length. This enables the local node to store more references to closer nodes than to nodes which are far away. The leftmost bucket contains only the reference to the local node, its sibling bucket may hold exactly one node which differs only in the least significant bit. It is less and less likely that the buckets to the left are filled the farther left they are. That is due to the equal distribution of the identifiers.

During a key lookup it is tried to cut the distance to the destination key at least by half. For doing so the source node XORs the own node identifier with the key to look up. The bucket with the longest shared prefix tag is selected and α nodes are picked from the bucket and the routing request is forwarded to them. Kademlia is able to parallelize its routing requests. The degree of parallelization is α and can be chosen freely. The method of picking a node from the bucket is not specified by the authors of Kademlia. A common implementation is to take the closest node to the destination key. This minimizes the hop count to the final destination. If the routing tables are reasonably filled and identifiers are equally distributed the routing algorithm terminates in $O(\log(n))$. The contacted node sends back the n_{tell} closest nodes of the requested key to the sender and adds the sender if the routing table is not already full. In a pathologic case with no lookup traffic for a long time, a stabilization interval t_{stab} is used to ping nodes from the routing table.

The Chord system is another popular peer-to-peer overlay. Chord and Kademlia share the way how node identifiers are generated. The main difference is the structure of the key space. In Chord, every node is positioned in a ring according to its identifier. The identifier next to another identifier in the ring has a numerically higher identifier, featuring a wraparound at 0. The routing table – or *finger table* – of a node contains a reference to the next node in the ring. This node is called the successor of the node. The finger table contains b references to the successor of the identifiers $(n+2^i)-1, i=0..b-1$. As in Kademlia this leads to a good knowledge of the node about its near nodes and less knowledge about far nodes. Routing can be done in a matter of binary search in the ring and runs in $O(\log(n))$ overlay hops. The distance to the destination node can be cut at least by half each routing hop if the routing table is correct.

If nodes join or leave the system without notice the Chord routing table gets outdated. A stabilization algorithm is used to repair the finger table and the successor reference. The stabilization uses a reference to the predecessor of a node. In a periodic manner the local node requests the predecessor from its successor. If the local node is not the predecessor, the successor reference is adapted to the node that is returned as a predecessor. The requested node may also adjust its predecessor reference. To be more resilient against node failures, a node may keep up to n_{succ} successor candidates in a list which are tried one after another if the first entry fails.

To stabilize the finger table, periodic search requests for the identifiers in the table are done and the found node replaces the finger reference. The stabilization uses bandwidth which may not be available for search purposes. The setting how often successor stabilization (t_{succ}) and finger table stabilization (t_{finger}) is done, is an important performance parameter.

The difference of Kademlia to other structured overlay networks is the symmetry of the XOR metric. That follows directly from the symmetry of the XOR operation. Peer A has the same distance to B as B has to A. This allows peers to learn about close nodes from incoming routing requests. It reduces the traffic necessary to maintain the overlay network. This feature makes it more promising for use in disadvantaged networks.

2.4. Performance Analysis Approaches

To analyze the performance of peer-to-peer systems, a peer-to-peer system is often simulated to isolate the influences of the underlying network and the user behavior. Experiments in real networks also exist, but for some peer-to-peer systems no widely used networks exist (e.g. Pastry). If available, often the DHT function of the system is used as a test application. The approaches to measuring a DHT's performance differ. Mostly the correctness of the algorithm is shown. Measurements in Pastry [13] were conducted as a simulation within a single Java VM, so node interaction breaks down to Java object invocation. The network model used was derived from [14]. The same network model is used in CAN's performance analysis in [15], but in contrast the node interaction has a fixed delay. Some publications [16], [17] deal with comparing different algorithms in a similar environment with different link delays, making the results somewhat comparable.

In [18] not only the algorithms, but also implementations of Chord, Pastry, Kademlia and Bamboo are measured. The analysis took place in an internet environment emulated by Linux Traffic Control. In [19] the Kademlia network is crawled and the behavior of network nodes is described. The decisive influence of its implementation on the content retrieval delays is shown in [20]. Performance measurement in peer-to-peer systems is challenging because a large number of nodes have to be set up and measured in a controlled manner. The conducted measurements show different approaches to this issue. A balance between simple setup with a precise measurement and realistic network behavior with a sufficient number of nodes has to be found.

We examine one study in further detail. In [17] the routing of Chord, Tapestry, Kademlia, Kelips and OneHop are evaluated. As Kelips uses large routing tables in size of $O(\sqrt{n})$ and Chord and OneHop are not well suited for networks with high churn rate the comparison breaks down to a comparison of Tapestry and Kademlia. The authors also identified the most important parameters and gave recommendations for the parameter values. Kademlia is able to invest bandwidth either in neighborhood consolidation or lookup correctness. The original authors of Kademlia propose a consolidation interval of one hour. The authors of the performance analysis decided to measure the system with a stabilization interval from 4 to 19 minutes. The stabilization interval of 19 minutes resulted in best behavior in terms of routing correctness and delay performance. Although identified as only a minor effective parameter by the authors, it would be interesting to investigate the effect of a consolidation interval longer than 19 minutes.

As Tapestry and Kademlia show similar success in simulations while Kademlia has the ability to learn new contacts through incoming routing requests and is also able to parallelize its requests it is considered the more promising overlay for even more difficult environments as considered in the prior analysis. The Chord overlay was not included in the comparison, so an analysis was done to compare Chord and Kademlia.

3. Comparison of Chord and Kademlia

Chord and Kademlia are compared in a simulated tactical environment to find out, which system is more suitable in a military network. Kademlia has been identified as a possible candidate in the previous section. We take the churn-optimized parameter settings from [17] as a starting point. Then we analyze the behavior of the two overlays in the presence of network errors and compare the results. We use the Chord and Kademlia implementations of the OverSim framework described in [21]. OverSim runs inside the OMNet++ network simulator [22]. The simulation includes a network and delay model as well as a model of the behavior of the nodes themselves.

The network model consists of wireless terminals equipped with IEEE 802.11b wireless LAN infrastructure mode and a fixed transmission capacity of 2 Mbit/s. There are 32 nodes per access point, forming an isolated collision domain. Each access point is attached to an IP router with a 100 Mbit/s Ethernet link. The router has a fixed delay line to every other router. The tactical network model (Fig. 3) has 4 access points and 4 routers. We used the INET extension of OverSim to simulate the full network stack from overlay down to physical layer. This model respects the increased availability of commercial of the shelf (COTS) hardware for military purposes and the tendency to use broadband radio equipment. The availability of a backbone network is anticipated as well.

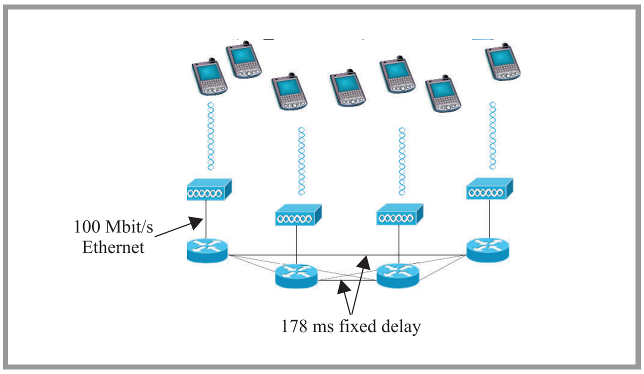


Fig. 3. The tactical network model.

Before a packet is sent, a packet error is applied according to a Bernoulli experiment with variable error probability. If the experiment yields 1, the message is tagged with an error bit. The packet is sent and network resources are consumed. The receiver silently discards the message if the error bit was set.

The delay between the routers is set to a fixed value of 178 ms. It is the mean value from the "King" data set [23], a collection of delays between servers in the internet. The delay on the wireless link is determined by the data link access and media access layer of the WLAN.

The simulated network contains 128 nodes. A tactical network may have connectivity to other tactical networks or even networks of strategic scale. The node count may rise up to thousands of peers. In the simulation the number of peers was limited by the used simulation framework and the peer and network model, not by the P2P systems. All overlay nodes are equal in capabilities and connectivity. Nodes are evenly distributed around the access point and do not move.

The nodes are dynamic in their behavior. This means that new nodes arrive and nodes leave the network over time. This behavior is called *churn* and reflects user fluctuation, either due to network failures or user behavior. Churn can be described by the arrival process of new nodes and their lifetimes. Different churn models are described in [24], [25] and [26]. Tactical users fluctuate more than the typical P2P user, reflecting roaming and network failures within the tactical domain. In [27] a distribution is proposed to model the lifetime process of a P2P system user in the internet. The lifetime of user connections follows the Weibull distribution with a mean of 164 minutes and a median of only 16 minutes. It shows a preference for short-lived connections. This is an effect which is assumed to be present in tactical networks as well. As no tactical peer-to-peer systems are known to the authors, the behavior of its users has to be estimated. Our model reflects this fact by assuming a similar Weibull distribution with the same shape but different mean lifetime. We introduce use two churn models: *normal churn* with 163 minutes mean lifetime and *intense churn* with an even shorter mean lifetime of 60 minutes.

The most important parameters of Chord and Kademlia in a churn intense environment were isolated by Li *et al.*

We took the "best" parameter set of Chord from their publication [17] to optimize for a high success ratio. The churn intense scenario in this publication is modeled as a Poisson arrival process with a mean of 1 hour. Due to the fact that we use a different churn model as described above, a different network underlay, message sizes and a reduced node count of 128, the resulting traffic production was 10 byte $s^{-1} node^{-1}$. Experiments with different parameter settings for n_{succ}, t_{finger}, t_{succ} showed that the initial parameter set already resulted in good success ratios in the tactical environment.

The newly derived parameters with the highest success ratio are shown in Table 1. The parameters for Kademlia were found by matching the traffic rate for our environment with Chord's traffic rate while maximizing the success ratio. Especially the stabilization interval could be chosen longer, as Kademlia needs stabilization only if not enough routing traffic is present. The packet error rate was varied to measure the influence on the performance of the two overlays. In every simulation run the error rate for all nodes was equal. We varied the packet error rate (PER) in steps from 0.001 to 1.

Table 1
The overlay parameters used for comparison of Chord and Kademlia in the tactical model

Parameters	Chord	Parameters	Kademlia
n_{succ}	8	n_{tell}	8
t_{finger}	120 s	k	8
t_{succ}	20 s	α	3
b	2	t_{stab}	1000

We measure the *delivery ratio* of the overlay routing process. This is the ratio of terminating routing request per total number of routing requests. The higher the ratio the more reliable the routing is. Another method to measure the correctness of the routing is to measure the *success ratio*, that is to take the ratio of successful routing by the amount of total routing request. A successful routing terminates at the node closest to a given search key. A terminating request does not necessarily find the node closest to the requested key. Successful routing requests are measured by issuing a search request on a key which is identical to a node identifier in the network. As this approach introduces a priori knowledge about the existence of certain nodes, a higher success ratio than expected is measured. A correct measurement of the success ratio would require complex distance comparison, slowing down the simulation. For these reasons we preferred to use random keys and the delivery ratio to test the lookup correctness. It has to be noticed that the success ratio of a DHT exceeds the delivery ratio of the routing by far. As the DHT may use the r closest nodes storage locations, the success ratio of the DHT mainly depends on this parameter.

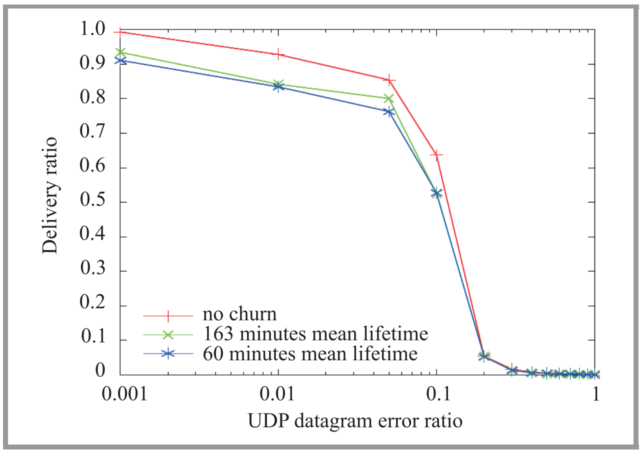


Fig. 4. Delivery ratio of Chord.

The results of the comparison between Chord and Kademlia are shown in Figs. 4 and 5. As a comparison three different levels of churn: no churn, normal churn and intense churn are depicted. The delivery ratio of both over-

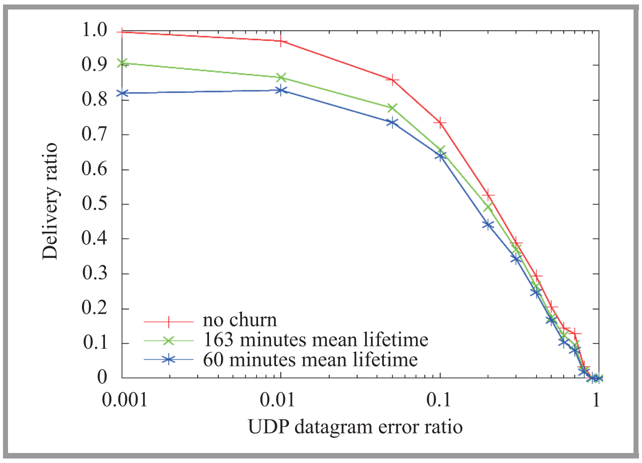


Fig. 5. Delivery ratio of Kademlia.

lay types declines with increased packet error rate. Chord achieved higher delivery ratios if a low packet error rate is present. As packet error rate increases, Kademlia shows better performance.

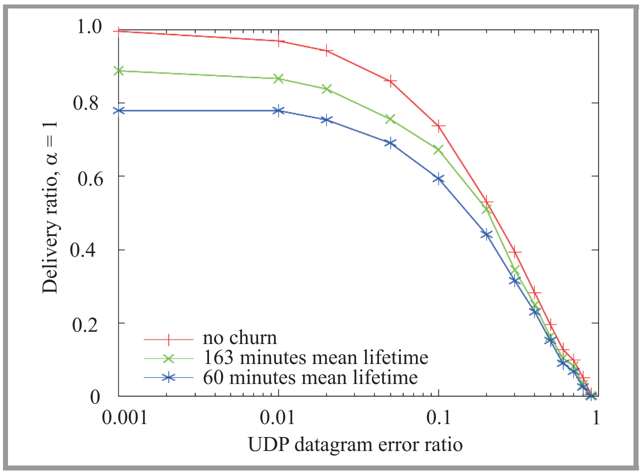


Fig. 6. Delivery ratio of Kademlia with $\alpha = 1$.

This provides rationale to prefer Kademlia in environments with high churn and high packet error rates. Chord is not able to use parallel lookups. To isolate the effect of parallel routing requests the measurement is repeated with the Kademlia parallelization parameter $\alpha = 1$, effectively disabling parallel lookups (Fig. 6). Still Kademlia is more stable when high packet error ratios are encountered. The reason for better performance in the presence of high packet error ratios is to be found in the rigidity of Chord where node failures have a more serious impact on the correctness of the routing table than in Kademlia.

4. Modification of Kademlia

We propose a change to the Kademlia routing mechanism to improve its performance in the presence of network errors and high latency. The modified Kademlia is able to incorporate signaling from lower layers or applications [28]. Our aim is to make the routing more adaptive to the underlying network structure. As the concept of proximity routing [29] requires additional messages, the proposed concept does not. It incorporates information from the routing or application layer, which can be delivered without cost in terms of additional traffic. The guarantees of the Kademlia routing, especially the completeness and the complexity properties remain untouched. A node running a modified version integrates seamlessly in running networks without modification.

Our approach is not to modify any existing routing parameters but to use a different method of choosing contacts. Although the method is also applicable to other peer-to-peer systems, the scope of this paper is limited to Kademlia. Cross-layer information is integrated into the routing decisions. The additional information is called preference value or simply the preference of a link or node.

The Kademlia routing described in Subsection 2.3 is changed in the way the sender of a lookup request selects nodes to contact. The original algorithm first selects the appropriate bucket (Fig. 7) and puts the contained routing entries into a list L of candidates. The first α candidates are then contacted and the lookup request is forwarded to them. In some situations if a bucket is very sparsely filled, entries from adjacent buckets may be used. In the modified version every contact is now augmented with a preference value. We introduce a weight factor w, which determines the influence of the preference. A factor of 1 means the next hops are determined according to the preference value and L is sorted according to the preference values. A weight of 0 represents the unmodified algorithm. Intermediate values of the weight affect the order in a continuous manner. The new sorting order is defined by:

$$m_d = weight\frac{pref_{s,d}length(L)}{max\_pref} + (1 - weight)pos_d,$$

where: s denotes the local source node making the routing decision and d a remote destination node. The original position of d the in L is pos_d. The list L is then reordered in descending order according to m_d.

The effect is shown in Fig. 7, a different node of the same bucket is preferred over a closer one. As the original version minimizes the hop count by always choosing the closest nodes the modification increases the hop count.

We test three methods to generate a preference value. The first method is to take the channel delay between the local node and the next hop i of L as preference value. After a normalization step the delay is used as preference. We call this modification *modification 1*. The second method is to take the bit error rate between the local node and the remote node $BER_{s,d}$ as a preference value, it is called *modification 2*. *Modification 3* only takes a value defined by the remote node into consideration. The node may set a low value to attract routing traffic or a high value to avoid it.

We use a simplified model of a tactical environment to be able to simulate more nodes. The simplified network model contains 1024 nodes. For modification 1 and 2 all nodes

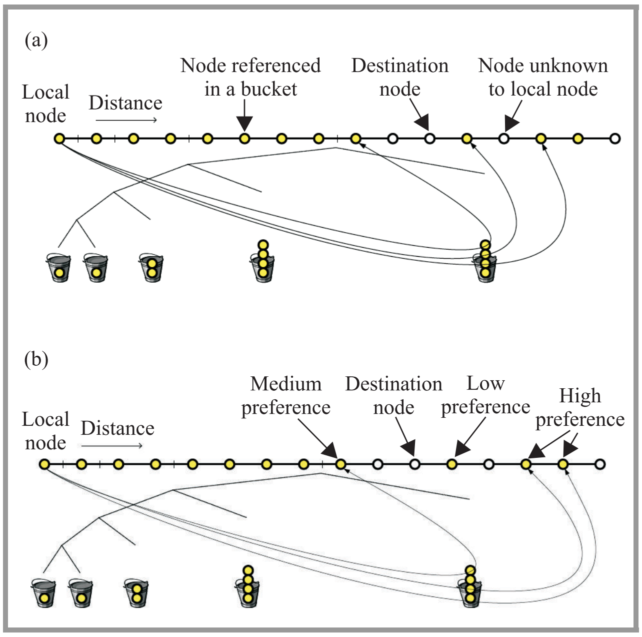


Fig. 7. Next peer selection: (a) original version; (b) changed lookup.

are equal, modification 3 introduces two different types of nodes.

The simulation time is 5 hours including a warm up time of 1 hour. Every run was repeated 10 times with different random number seeds. All nodes in the network feature a UDP network stack. Before a packet is sent to another node an error model and a delay model are applied. Before a packet is sent, a bit error may be applied according to a variable error probability. No churn is used in the simplified model. The nodes are placed equally distributed. The link delay between two nodes increases linearly with the Euclidean distance of the nodes. The maximum delay is about 7 s, this is below the message timeout value of 10 s. This simulates the effects of lower layer protocols such as multi hop propagation in a simplified manner. Lost messages get detected by the overlay 10 s after they have been send. In Fig. 8 the results of modification 1 and 2 are shown. The measurement with weight set to 0 is included as a reference to the unchanged Kademlia routing algorithm in the simplified network model. The figure depicts the time it takes to perform a DHT GET request with modification 1 and modification 2. The DHT GET latency is the duration it takes to retrieve a previously stored value from the DHT. The weight was modified to analyze the effect of the preference values to the routing. The absolute numbers of sender traffic and success ratio are of lesser importance as they are mainly dependent on the parameter settings of the overlay. It is possible to increase the success ratio for example by an increase of parallel lookups. The absolute values for the DHT latency are mainly dependent on the network model and the link delays. We focus on the relative change of the values when we modify the routing. In Fig. 8a the effects of both modifications on the DHT GET latency is shown. The preference for faster connections in modification 1 does not seem to pay off in terms of latency. The reason for the low impact of modification 1 is how Kademlia sends parallel lookups during the routing process. Since Kademlia tries to hold α lookup request in flight, the fastest response is immediately processed and the routing continues with the sending of another routing request. The parallelization elevates the effects of preferring fast nodes as the probability is high that 1 of α nodes reacts. Timeouts dominate the influence on latency. Timeouts occur if a node has failed. Sending slots are blocked for the duration of the timeout.

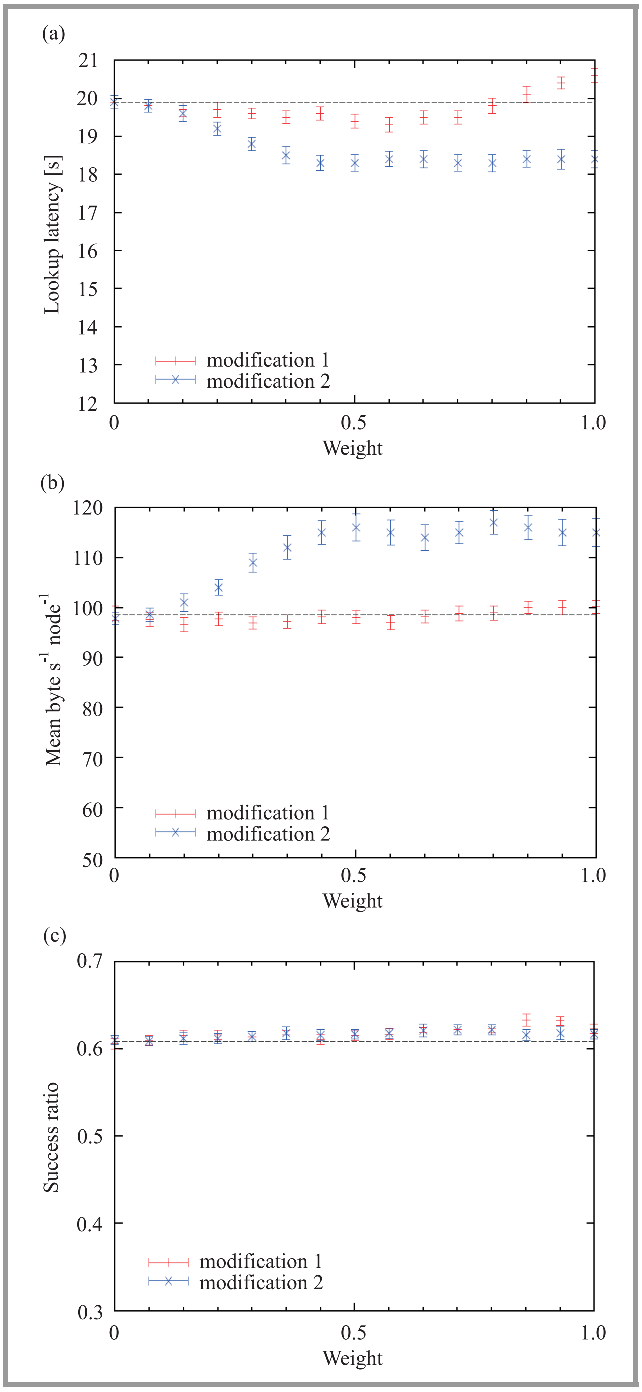


Fig. 8. Effects of modification 1 and 2: (a) DHT GET latency; (b) overlay traffic sent; (c) DHT success ratio.

If all slots are blocked the routing stalls until the first timeout occurs. Because less timeouts occur, modification 2 (which prefers low BER links) performs better in terms of reduced latency. This observation is in contrast to existing implementations as the RTT is often used to approximate the reliability of a link. In environments with high BER, preferring low BER links is better than to use faster links. Modification 2 facilitates a trade-off between delay and the amount of transmission capacity needed to maintain the peer-to-peer overlay.

Figure 8b shows the mean number of bytes sent per node per second for a fixed weight. Modification 1 does not change the amount of sender traffic as it has a too little effect on the routing decisions. Modification 2 increases the bytes sent. This is because the original Kademlia routing decision is optimized for low overlay hop count. Any change to the routing decision will increase the hop count. The higher hop count leads to an increase of the traffic of modification 2. At a weight of 0.5 the modification offers minimum latency, at this weight it causes an increase of 17% in overlay routing traffic. Our results show the possibility to exchange reduced latency for an increase in sender traffic.

The success ratio of the DHT lookups is shown in Fig. 8c. A DHT lookup was counted as successful if the value could be retrieved without timeout from 1 of the redundant storage locations ($r = 3$). The success ratio stayed constant or increased slightly with increased weight. The absolute rate of success is less important as it is always possible to trade bandwidth for increased success ratio by parameter changes. The important observation is the success ratio does not decrease as an effect of the modifications.

To test modification 3 we set every 10th node to the least preferred value. This simulates a node with limited battery capacity. As sending consumes scarce battery power, we measured the accumulated number of bytes sent out by the tagged nodes over the whole simulation duration. The results can be seen in Fig. 9. The original amount of data sent out is shown as crosses, the simulation run with modification 3 is shown in an x-shape. The amount of bytes per node is not equal for all nodes even in the unmodified scenario. Nodes join the network successively, so node 0 is the first and 255 the last. The effect is not visible if churn is applied. In Kademlia long lived nodes are preferred over newly arrived nodes if a bucket is full. This bucket eviction policy leads to the fact that node 0 features the highest traffic and node 255 the least. The low battery nodes can lower their amount of sending by up to 25% with modification 3, while still taking part in the overlay with no disadvantages. The accumulated traffic of all nodes over the whole simulation time remains nearly unchanged at 151.11 MB versus 153.38 MB with modification 3, also the success ratio remains nearly unaffected.

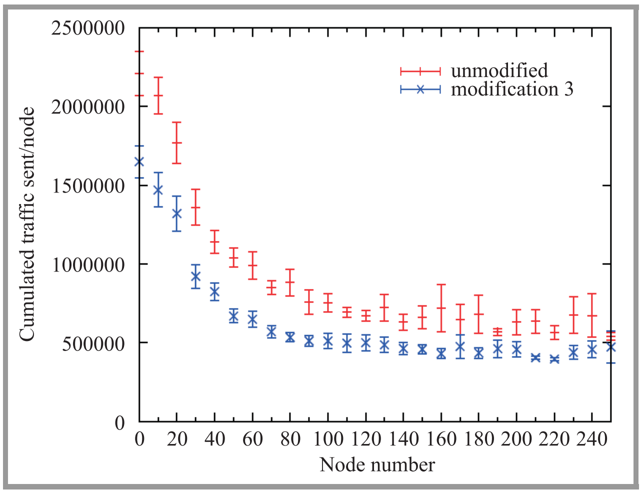


Fig. 9. Traffic shaping by modification 3.

5. Summary

We analyzed different peer-to-peer systems for their suitability in an error-prone military network. Chord and Kademlia were identified as candidates to be suitable in such networks. The candidates were compared in a simulated tactical environment. The environment features wireless and wired networks and a faithful media access simulation. We showed that Kademlia offers a higher delivery ratio than Chord in the presence of churn and high packet error rates. Then we introduced a change to Kademlia's routing algorithm to include cross-layer information. We gave three examples how to use the extension. Error rates of links are reported through the new interface (modification 2). It was shown that it is possible to reduce the number of timeouts and thereby decrease the latency of the peer-to-peer system. A cross-layer signaling of the link latency did not improve the performance of Kademlia in the considered environment because the parallelization of routing requests in Kademlia elevates the effects. Nodes which need to save battery power can use extension 3 to reduce their contribution to the overlay network. Low battery nodes remain full members of the network and suffer no disadvantages but they send significantly less traffic. The overlay can cope with a considerable percentage of disadvantaged nodes with no limitations. Our modified client may join a Kademlia network without interfering with existing clients and overlay networks. The presented results show a possibility to increase the availability of information in tactical networks. Future steps include the analysis of the peer-to-peer system with a tailored publish/subscribe capability.

References

[1] Dr. Scholl. Opennap, 2000 [Online]. Available: http://opennap.sourceforge.net/

[2] The gnutella protocol specification v0.4 [Online]. Available: http://www.stanford.edu/class/cs244b/gnutella_protocol_0.4

[3] J. Liang, R. Kumar, and K. Ross, "The kazaa overlay: a measurement study", in *Proc. 19th IEEE Ann. Comput. Commun. Worksh.*, Bonita Springs, USA, 2004.

[4] T. Klingberg and R. Manfredi, Gnutella 0.6, 2002 [Online]. Available: http://www.rfc-gnutella.sourceforge.net/src/rfc-0_6-draft.html

[5] S. Guha and N. Daswani, "An experimental study of the skype peer-to-peer voIP system", Techn. Rep., Cornell University, Dec. 2005.

[6] B. Yang and H. Garcia-Molina, "Designing a super-peer network", Techn. Rep. 2002-13, Stanford University, 2002.

[7] M. Steiner, E. W. Biersack, and T. En Najjary, "Actively monitoring peers in KAD", in *6th Int. Worksh. Peer-to-Peer Sys. IPTPS'07*, Bellevue, USA, 2007.

[8] *Why Kad Lookup Fails*, H. Schulzrinne, K. Aberer, and A. Datta, Eds. in *Proc. IEEE 9th Int. Conf. P2P Comput. 2009*, Seattle, Washington, USA, 2009.

[9] P. Maymounkov and D. Mazieres, "Kademlia: a peer-to-peer information system based on the XOR metric", in *Int. Worksh. Peer-to-Peer Sys. IPTPS, LNCS*, vol. 1, Cambridge MA, USA, 2002.

[10] B. Y. Zhao, L. Huang, J. Stribling, S. C. Rhea, A. D. Joseph, and J. Kubiatowicz, "Tapestry: a resilient global-scale overlay for service deployment", *IEEE J. Selec. Areas Commun.*, vol. 22, no. 1, pp. 41–53, 2004.

[11] I. Stoica, R. Morris, D. R. Karger, M. F. Kaashoek, and H. Balakrishnan, "Chord: a scalable peer-to-peer lookup service for internet applications", in *Proc. 1st ACM SIGCOMM*, San Francisco, USA, 2001, pp. 149–160.

[12] A. Rowstron and P. Druschel, "Pastry: scalable, decentralized object location and routing for large-scale peer-to-peer systems", in *Proc. IFIP/ACM Int. Conf. Distr. Sys. Platforms (Middleware)*, Heidelberg, Germany, 2001, pp. 329–350.

[13] A. Rowstron, A.-M. Kermarrec, M. Castro, and P. Druschel, "Scribe: the design of a large-scale event notification infrastructure", in *Proc. 3rd Int. COST264 Worksh. Networked Group Communication NGC'2001* , *Lecture Notes in Computer Science*, J. Crowcroft and M. Hofmann, Eds., vol. 2233. London: Springer, 2001, pp. 30–43.

[14] E. W. Zegura, K. L. Calvert, and S. Bhattacharjee, "How to model an internetwork", in *Proc. IEEE INFOCOM'96*, San Francisco, USA, 1996, pp. 594–602.

[15] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker, "A scalable content addressable network", Techn. Rep. TR-00-010, International Computer Science Institute, Berkeley, USA, Oct. 2000.

[16] M. Castro, M. B. Jones, A.-M. Kermarrec, A. Rowstron, M. Theimer, H. Wang, and A. Wolman, "An evaluation of scalable application-level multicast built using peer-to-peer overlays", in *Proc. IEEE INFOCOM 2003*, San Francisco, USA, 2003.

[17] J. Li, J. Stribling, R. Morris, M. F. Kaashoek, and T. M. Gil, "A performance versus cost framework for evaluating DHT design tradeoffs under churn", in *Proc. IEEE INFOCOM 2005*, Miami, USA, 2005, pp. 225–236.

[18] D. Kato and T. Kamiya, "Evaluating DHT implementations in complex environments by network emulator" in *Proc. Int. Worksh. Peer-to-Peer Sys. IPTPS'07*, Bellevue, USA, 2007.

[19] M. Steiner, T. En-Najjary, and E. W. Biersack, "Exploiting KAD: possible uses and misuses", *Comput. Commun. Rev.*, vol. 37, no. 5, pp. 65–70, 2007.

[20] M. Steiner, D. Carra, and E. W. Biersack, "Faster content access in KAD" in *Peer-to-Peer Computing*, K. Wehrle, W. Kellerer, S. K. Singhal, and R. Steinmetz, Eds. New York: IEEE Comput. Society, 2008, pp. 195–204.

[21] I. Baumgart, B. Heep, and S. Krause, "OverSim: a flexible overlay network simulation framework", in *Proc. 10th IEEE Glob. Internet Symp. GI'07 in conj. with IEEE INFOCOM 2007*, Anchorage, USA, 2007, pp. 79–84.

[22] A. Varga, "OMNeT++ discrete event simulation system", 2009 [Online]. Available: http://www.omnetpp.org

[23] K. P. Gummadi, S. Saroiu, and S. D. Gribble, "King: estimating latency between arbitrary internet end hosts", in *Proc. 2nd ACM SIGCOMM Internet Measur. Worksh. 2002*, Marseille, France, 2002.

[24] Z. Yao, D. Leonard, X. Wang, and D. Loguinov, "Modeling heterogeneous user churn and local resilience of unstructured p2p networks", in *Proc. IEEE Int. Conf. Netw. Prot. ICNP'06*, Washington, USA, 2006, pp. 32–41.

[25] R. Brunner, *A Performance Evaluation of the Kad-Protocol*. Corporate Communications Department Institut Eurécom France, 2006.

[26] K. C. Almeroth and M. H. Ammar, "Multicast group behavior in the internet's multicast backbone (mbone)", *IEEE Commun. Mag.*, vol. 35, pp. 224–229, 1997.

[27] D. Stutzbach and R. Rejaie, "Understanding churn in peer-to-peer networks", in *Proc. 6th ACM SIGCOMM Conf. Internet Measur. IMC'06*, J. M. Almeida, V. A. F. Almeida, and P. Barford, Eds., Rio de Janeiro, Brazil, 2006, pp. 189–202.

[28] T. Ginzler and M. Amanowicz, "Simulation of the impact of packet errors on the kademlia peer-to-peer routing", in *RTO Informa. Sys. Technol. Panel Symp. IST–092/RSY–022*, NATO, 2010.

[29] M. Castro, P. Druschel, Y. C. Hu, and A. Rowstron, "Exploiting network proximity in distributed hash tables", in *Proc. Int. Worksh. Future Direct. Distrib. Comput. FuDiCo*, O. Babaoglu, K. Birman, and K. Marzullo, Eds., Bertinoro (Forli), Italy, 2002, pp. 52–55.

Tobias Ginzler graduated as a computer scientist (Dipl.-Inf.) in 2006 at the University of Bonn, Germany. He then joined the research facility in Wachtberg which is now known as Fraunhofer FKIE. His topics of research at the department of communication systems include multicast key management, tactical messaging, service oriented architecture and peer-to-peer systems. In parallel he is a Ph.D. candidate at the Military University of Technology, Warsaw.

e-mail: tobias.ginzler@fkie.fraunhofer.de
Fraunhofer FKIE
Neuenahrer Str. 20
53343 Wachtberg, Germany

Marek Amanowicz – for biography, see this issue, p. 53.

Paper

Review of Distributed Beamforming

Jason Uher, Tadeusz A. Wysocki, and Beata J. Wysocki

University of Nebraska-Lincoln, Omaha, USA

Abstract—As the capabilities of individual nodes in wireless sensor networks increase, so does the opportunity to perform more complicated tasks, such as cooperative distributed beamforming to improve the range of communications and save precious battery power during the transmission. This work presents a review of the current literature focused on implementing distributed beamformers; covering the calculation of ideal beamforming weights, practical considerations such as carrier alignment, smart antennas based on distributed beamformers, and open research problems in the field of distributed beamforming.

Keywords—beamforming, distributed antenna array, smart antenna, virtual array.

1. Introduction

With recent advancements in both size and power efficient computing, the concept of the ubiquitous wireless sensor network has quickly emerged as a legitimate research topic. It is now possible to have a large network of relatively small devices distributed over a large area, all with limited means of communications, and precious little power to spare for long haul links. Significant research has been done on efficient routing algorithms, mutual information coding, and multi hop transmission schemes in an effort to reduce the amount of power required to transfer sensor data from the individual nodes in a network to a final destination where the data can be used. In an effort to further reduce power consumption, the use of distributed phased arrays has come into focus as a method for nodes to collaborate in their transmissions, saving power overall during the data transfer. By cooperating, the nodes are able to emulate a traditional fixed array of antenna elements and achieve the same gains in terms of main lobe enhancement, side lobe reduction, and null pointing to improve the intended receiver's SNR and remove the interference caused by unwanted transmitters. These arrays are called distributed smart antennas, or distributed beamformers, and have their own unique set of problems over fixed beamformers when it comes to ideal weight calculations.

Use of the term "distributed" has two distinct meanings in the sense of distributed beamforming. The first meaning indicates that the antennas of the array themselves are distributed over the receiving plane in some randomly structured fashion. This is a departure from traditional beamforming literature, which relies on a strict, uniform placement of the antenna elements to reduce the complexity of the analysis through the removal of dependence on the individual locations of nodes within the arrays. When the nodes are no longer structured so nicely, the location of each element must be considered on its own, rather than simply considering the location of the array as a whole. In this scenario, the elements are still controlled by some central source; hence the locations, phase offsets, and transmit capabilities of each node are known quantities to be taken advantage of during ideal weight calculations.

The second meaning builds on the first, implying that the elements are not only distributed in terms of location, but are also independent processing units, such as with a wireless sensor network in a field. This second scenario severely limits the quantity and quality of information available to a beamformer. In this case, methods for determining ideal complex weights must distributed in the sense that they can be carried out by each node individually without sharing significant amounts of information. If the nodes were allowed to share the total amount of information about themselves, such as through some pre-communication phase, the second scenario would collapse into the first, where ideal weights could be calculated based on the global information and disseminated through the network by a single cluster head.

Early work with systems where the global parameters for each transmitter are known, but the transmission elements are not in an organized regular array allowed for an initial insight into how arrays of unfixed elements might be approached. When the elements are distributed as in wireless sensor network, new considerations can be added to the algorithms, taking into account the need for distributed processing and synchronization. Growing from there, the capabilities of randomly distributed networks with specific distributions can be analyzed in terms of their capability with respect to steering both peaks and nulls.

The remainder of this review is organized as follows: Section 2 presents early work performed in the area of unevenly distributed and randomly distributed antenna arrays and its application to modern distributed arrays. This groundwork paved the way for virtual antenna arrays in distributed wireless nodes, which are discussed in depth in Section 3. This section is focused on the ideal calculation of beamforming weights in distributed networks to achieve the ideal beampattern for broadcasting and reception. Section 4 covers the practical aspects of beamforming in a distributed network and the methods for using the ideal weights discussed in Section 3. Finally, Section 5 gives some open research problems in the area of distributed beamforming and smart antennas.

2. Early Work in Random Antenna Arrays

Due to the nature of original phased array systems, the concept of a distributed phased array was not something that was inherently obvious. As the topic grew from sonar and acoustics into the electromagnetic domain, it was seen as a given that arrays could be placed in specific patterns as dictated by a designer, there was just no need to analyze distributed or random arrays.

There was a small body of work, however, that focused on the properties of both non-uniform and random linear arrays; usually with the intention of decreasing the number of required elements, eliminating the need for individual amplitude control hardware, or analyzing the effects of placement errors when building a physical array. The first attempts at moving away from a strict linear array was presented in [1], which introduces the idea that elements in a linear phased array need not be evenly distributed over the length of the antenna. The main goal was the reduction of the number of elements, the author demonstrated that by placing the elements at arbitrary points along the line, the designer increases the degrees of freedom in the overall design because through the addition of location. The extra degrees of freedom allow for an increase in the capabilities of the array (while necessarily increasing the complexity of design). To reduce this complexity, [2] introduced the concept of an equivalent uniformly-spaced array (EUA), which reduces the non-uniform array to an equivalently driven uniform array with a chosen spacing. Building on these original papers, other researchers continued to develop the concept of non-uniform linear arrays, and new methods for the design and optimization of phased arrays were discovered. Initial mathematical models and descriptions allowed for the general construction of arrays with desirable properties, but left out optimization of specific parameters [3]. Further development of these models led to more practical design considerations such as sidelobe reduction [4], as well as some experimental verification of the models being derived [5] .

With a strong understanding of the physical characteristics of non-uniform arrays in hand, researchers were able to build on the available models to generate theories based on random non-uniform arrays and how they behave statistically, rather than over a single iteration. Work included the analysis of general arrays [6], [7], the properties of the sidelobes [8], and even multidimensional arrays (disks and spheres) [9], [10].

3. Optimum Beamforming Weights

With the popularity of wireless sensor networks increasing steadily, there is now a need to further the analysis of arrays with truly random element spacing. Using the initial analyses from Section 2, especially those on multidimensional arrays [9], as a basis for the analysis of distributed beamforming in wireless networks allows researchers to apply the old concepts of fixed random arrays to wireless sensor networks.

Modern arrays benefit from a number of factors not available to the original body of research. First, the rapid advances in computers allow for fast, statistically significant numerical simulations, which allows potential schemes to be quickly evaluated. Second, increasing transmission capabilities allow for better control of arrays through the use of dynamic weighting. Arrays can now be weighted in software using complex weight multiplication, rather than cumbersome fixed phase and amplitude modifiers at the antenna elements themselves. With these improvements in technology, random array beamforming in wireless sensor networks bears only a slight resemblance to the previous work done on random arrays, but the initial research still provides valuable insight into the relationships between node placement and the achievable beam patterns.

The core of any modern beamformer is the complex weights used to modulate the signals at each element of the array in order to achieve the appropriate constructive and destructive interferences required for optimum results during cooperation. There are several problems unique to the calculation of ideal weights when beamforming using a distributed virtual array. The first change in weight calculation is clearly that the elements of the array are no longer in a fixed pattern, leading to extra complexity in the convergence of smart antenna algorithms. Additionally, it is possible that the elements may even be moving, introducing yet another factor in the calculation of ideal weights. This section focuses on the calculation of beamforming weights for individual nodes under these circumstances. In most cases, an ideal array of nodes is assumed; that is, the nodes have synchronized carriers and total knowledge of the array topology and source locations. When weight calculation methods deal with arrays where these assumptions do not hold, it will be specifically mentioned.

3.1. Distributed Beamformers as Wireless Relay Channels

At its heart, distributed beamforming can be modeled as a relay channel, with the transmitting nodes as sources, and the cooperative nodes as relays. It does not matter that often the sources and relays are the same node, or that every node may be able to reach the receiver on its own; the analysis is still pertinent in terms of ideal weight calculations. In this type of analysis, the definition of ideal may be flexible, meaning maximum gain at the destination, minimum power consumed, or minimum interference to unintended receivers. The benefit of this type of analysis is that little information about the array geometry or actual nodes is necessary, the network as a whole is abstracted, allowing generic analysis of performance under constraints on power consumption (weight magnitudes), available channel state information, and the number of cooperating nodes. The main drawback is that the geometry of the array, location of intended receivers, and location of interferers is abstracted into the channel state information between the nodes, which is not always available or easily estimated. In addition, knowledge of the antenna geometry may alleviate a number of constraints that are artificially introduced in dependence on the inter node channels. The analysis is beneficial, though, as a fair comparison of theoretical ideal weight calculation methods for individual nodes. Relay channels are characterized by what they do to the message from the source node, such as amplify-and-forward (AF), decode-and-forward (DF), or filter-and-forward (FF); each with different challenges in terms of the presence of noise, algorithm complexity, and required node information. The most complex of these, in terms required information and potential sources of noise is the AF case; both the DF and FF network types are special cases of the AF network with constraints placed on the types and placement of noise and the complex weights used to amplify the signal. Figure 1 shows an example of a relay network with noisy channels and individual node weights. Due to the abstract nature of relay channels, and their application across a wide variety of domains, there exists a large body of work devoted to their analysis. When applied to distributed network beamforming, several limitations in traditional relay channels are added into the problem. First, it is not typically assumed that the transmitting node is involved in the cooperation to reach the receiver, in distributed network beamforming the transmitter is usually part of a cluster, and takes part in the transmission. Second, typical solutions rely on a single constraint, the quality of the link the receiver. This has the effect of creating a maximum main lobe towards the intended target, but gives little consideration to the rest of the beampattern generated by the cooperating nodes. Additional constraints, such a minimizing the side lobes or steering nulls towards unintended receivers, add complexity to systems that already require significant relaxations to reach closed form solutions. As such, these systems may not be useful in practice due to the extreme complexity of the solutions.

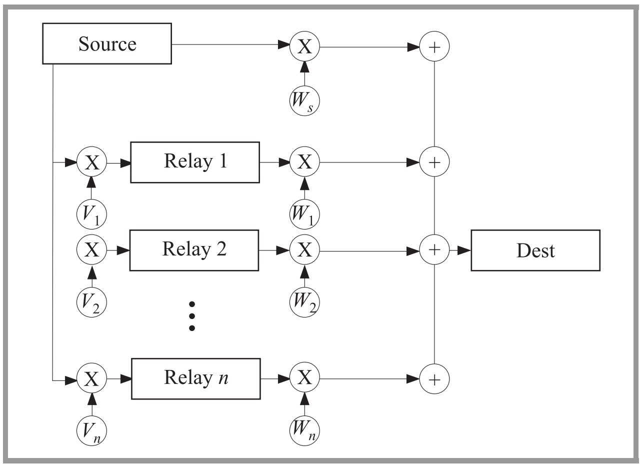


Fig. 1. Sample relay network.

Useful coverage of relay networks with respect to beamforming and beamcoding would warrant more coverage than this paper is capable of providing. As such, the remainder of the section will only provide a brief introduction on relay network methods with specific application to distributed beamforming, e.g. calculating ideal transmission weights to overcome channel effects, with a specific example of a paper with good coverage on that topic. To begin, [11] presents an excellent introduction to application of relay networks with respect to beamforming. It gives excellent examples of the workflow used when analyzing relay network beamformers. First, the constraint is chosen, in this case the SNR at the receive nodes, however any quantifiable quantity can be chosen, such as the capacity of the total link, the total power consumption, the per-node power consumption, etc. Second, an analytical derivation for the optimized value of interest is created based on the relay network model, and a method for iteratively reaching that optimum is presented. Finally, the problem is broken up such that the transmitter or receiver (or both) can calculate a single coverage parameter that leads the individual nodes to find their own optimum weights, distributing the calculation over the network. Here, basic AF and DF networks are analyzed with in terms of maximizing the SNR at the receiver based on varying degrees of channel state information, as shown in Fig. 2. For application to cognitive radio, it is possible to treat the relays such that there are multiple transmitters and receivers (either paired, or in one-to-many configurations), such as in [12]. This allows maximization of the constraint at each pair of nodes communicating through the relay. Additional constraints may be added to reduce the interference to primary users of a channel. In most cases, channel state information is a required piece of information (due to the lack of node geometry information). It is possible, however, to overcome this utilizing the relay network model, which utilizes the second order statistics of the channel, allowing for a more distributed approach.

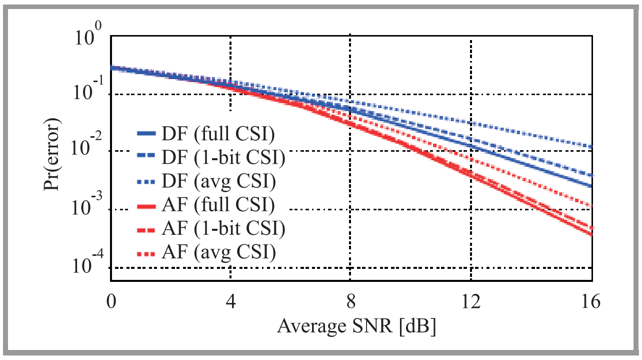


Fig. 2. Error analysis of DF and AF networks [11].

3.2. Adaptive Distributed Arrays

When a target is moving, or ideal weight calculations are not possible due to lack of information, an adaptive method may be used to home in on ideal weights by iteratively changing the phases based on the array performance. This leads to distributive smart antennas; capable of compensating for movement within the array, target or interferer as well as other changes in the channel.

In order to arrive at the ideal weights without full CSI, [8] attempts to iteratively find the optimum weighting at each relay using one bit feedback from the destination node. Two methods of iteration are introduced which randomly perturb the weights based quality information fed back from the destination node. The first method, take/reject perturbation (T/R), the weight is either perturbed or not based on the feedback; that is, if the receiver feels that the signal is sufficient, there is no perturbation. If the weights are to be changed, the perturbation new value is chosen such that

$$\tilde{a}'_k = a_k + \mu q_{k \bmod N}\,,$$

where μ is a scaling factor which affects the rate of convergence and q_k is a preset value from the perturbation set q.

Utilizing T/R allows for constant improvement in the quality of the weights, that is, the quality of the overall link in each successive iteration is at least as good, or better, than the previous iteration. However, this method will be slow to converge, as it is possible that the quality may remain constant over several iterations. In the second method, plus/minus perturbation, the next weight is perturbed twice during transmission, and the feedback bit specifies which of the two was the best. In this case, the tested perturbations are

$$\tilde{\alpha}_k^{\pm} = \frac{\alpha_k \pm \mu q_{k \bmod N}}{\|\alpha_k \pm \mu q_{k \bmod N}\|}\,.$$

The additional values allow for a faster convergence to an ideal weight vector because there are twice as many perturbations available in each iteration. However, it may also be the case that neither α^+ or α^- are better than the old weight α_k, leading to worse performance in the immediate time window. In both cases, the calculation of the perturbations needs to be normalized by the weights of the entire array. Because this would require each element to share its random weight with the entire array, a method of deterministic perturbations is introduced which allows each of the relay nodes to calculate its weights independently of the others. Figure 3 shows a BER comparison between each of these schemes under fading conditions along with a traditional ideal weight method that require full knowledge of the CSI based on gain equalization. Modifications to this scheme in [9] show that utilizing multiplicative perturbations rather than additive can give an increase in the performance of the system due to the fact that the deterministic multiplicative perturbation set Q can be chosen such that a set phase rotation is applied regardless of the current weight value. In this case, the weights are perturbed as such

$$\tilde{a}'_k = a_k + Q_{k \bmod N}\,.$$

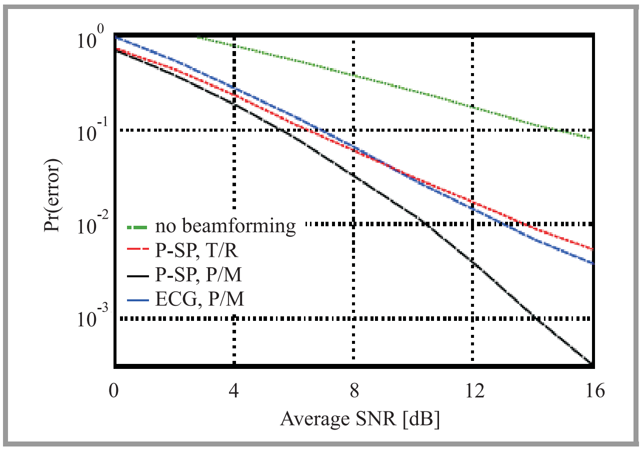


Fig. 3. Comparison of perturbation feedback schemes.

Figure 4 demonstrates the benefits of this method versus additive perturbation. Each perturbation moves the weights closer to their ideal values, as the multiplication prevents perturbation magnitude in irrelevant dimensions. In [13], the authors show that optimum beamforming weights can be found iteratively by having each node broadcast an effective cost to the other nodes in terms of its own interference. With of an idea of how its transmissions affect the other users, a single node can maximize its utility (data rate) while minimizing its interference to the other nodes. In addition, the updates of the users cost are distributed throughout many frames. In fact, when the costs are updated simultaneously, it causes oscillations and prevents the nodes from converging on an optimal solution.

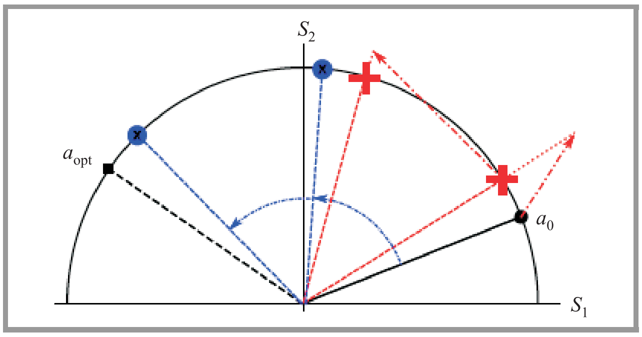


Fig. 4. Multiplicative versus additive perturbations [14].

A common relay channel technique is to use singular value decompositions (SVD) to obtain ideal beamforming weights using global CSI information. The SVD method allow the transmitter to precode data (x) sent over a MIMO channel (h) with the left decomposition of the channel (v) and the receiver to decode with the right decomposition (u), giving the received vector (y)

$$y = u^H H v x + u^H n\,.$$

The authors of [15] present an iterative method for calculating the right SVD vector, relying on blind adaptive methods for calculation of the left. This method works by treating the multiple paths as parallel SISO links with gain

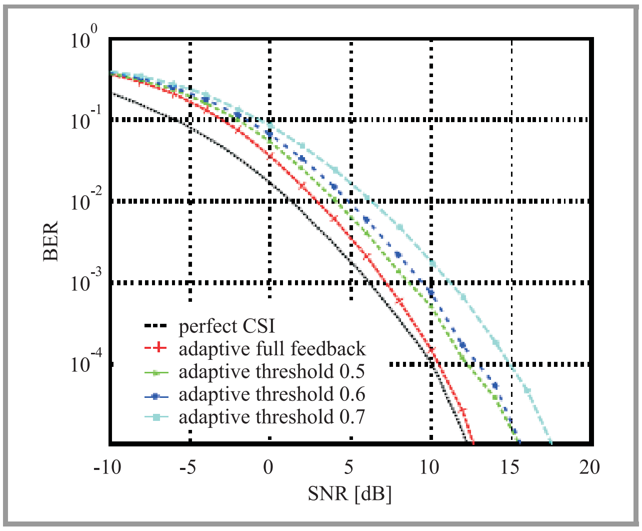


Fig. 5. Effects of feedback threshold on SVD weight calculations [15].

specified by the diagonal elements of the SVD, allowing nodes to calculate their ideal weights based on their local element. The amount of information sent over the feedback channel is reduced by using a predictor to estimate the value of the current singular vectors at each transmitting node, rather than feeding back the vector in each iteration. After each iteration of the transmission, the values of the singular vector are transmitted back to the relay if the difference between the estimated values and the calculated values exceed a set threshold, allowing for a balance between performance and the overhead in the control channel. In Fig. 5, the BER performance for varying thresholds is shown.

3.3. Distributed Weight Calculation

The greatest overhead in distributed beamforming is the sharing of locations or channel state information between the nodes to allow for weight calculations across the network. Methods for calculating these weights in a distributed fashion, sharing as little data as necessary, allow for vast improvements in the overall performance of the distributed arrays; making it one of the most important research topics in the field.

The best way to prevent the need for sharing CSI for every node is to not use CSI during the weight calculations. The authors of [16] present a system that uses the second order statistics of the individual node channels. Starting with an initial estimate of ideal weights, the individual nodes can continue to refine their own weights locally using only a parameter based on the combination of the transmissions in the uplink, which is fed back from the receiver. This idea is carried forward in [17] , where the second order statistic calculation includes multiple source transmitter pairs, adjusting the weights at the relay nodes to optimize the signal at several receivers rather than just one, that is, the beamformer adjusts to minimize the transmit power (P_T) subject to the required SNR between each transmit/receive pair

$$\min_{w} P_T \quad SNR_k \geq \gamma_k \forall k \, ,$$

where SNR_k is the SNR at each transmitter pair and γ_k is the minimum required SNR for that pair.

In this case the SNR is actually the signal to interference and noise ratio (SINR) as the signals from other pairs are treated as interference,

$$SNR_k = \frac{P_k}{Pn_k + \sum_{i \neq k} P_i} \, ,$$

where Pn_k is the noise at P_k and the sum term is the power at each of the other pairs.

Utilizing these constraints, we can find the optimal weights through the following minimization

$$\min_{w} w^H D w \quad \text{s.t.} \quad \frac{w^H R_k w}{wH(Q_k + D_k)w + \sigma^2} \geq \gamma_k \, ,$$

where R, Q, D, and σ^2 are the correlation matrix of the channels, the average of the complex paths, the diagonal values of R, and the noise variance. The authors of [18] the less common route of constraining the energy per node as well as the total system power, but still uses common information transmitted back from the receiver; namely the maximum SNR capability (Γ_{opt}) and computed scalar channel statistics (ξ, β_{opt}):

$$w_i = \min \frac{\xi l_i}{1 + \beta_{\text{opt}} + \lambda_{\text{opt}} \Gamma_{\text{opt}} N_i} \left(h_i g_i \right)^* .$$

Here the common scalar channel characteristics taken in accord with the local values of the channel (h_i, g_i, N_i) allow the local node to compute its own value without sharing weights individually. The calculation of the common statistics, in particular Γ_{opt}, takes into consideration the transmission power at individual nodes. Figure 6 shows the performance of the systems as a comparison of

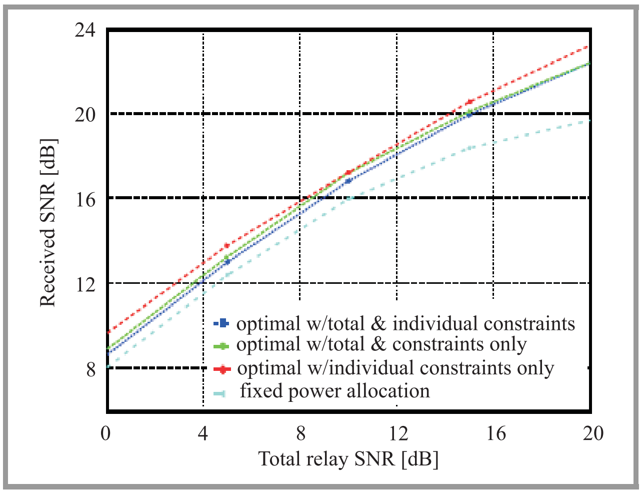


Fig. 6. Distributed beamforming constraints [19].

the SNR at the receiver dependent on the total SNR of the relay, the total transmit power divided by average noise power. It is clear to see that the per node power constraints allow for a higher SNR at the receiver when the overall SNR is low, but that when the links are of a higher quality, the total power constraint method is superior. In [14], a system for weight calculation is presented such that there will be one node whose impact is the greatest, and will transmit at full power, or at some value depending only on the local information at the node. This is based on traditional relay selection where only the best relay is chosen, and the others remain quiet to conserve battery power. However, rather than remain quiet, these other nodes can still contribute power based on their own channels. The scheme in [14] feeds back an indication to the nodes as to which of them has been selected as the optimal transmitter, the chosen node will transmit at full power as in the traditional relay selection schemes, but the rest of the nodes will also transmit with a small amount of power based on their own channels to the receiver. A distributed SNR balancing approach in [20] that finds ideal weights to balance the transmission capabilities so that the lowest SNR of the transceiver set is maximized within the constraints, that is

$$\max_{w}\ \min(SNR_1, SNR_2) \quad P_T \le \tilde{P}\,.$$

Here the lowest SNR at the two receivers is maximized subject to the total transmit power of each transceiver and all the relays

$$P_T = P_1 + P_2 + \sum_{i=1}^{N} P_{ri}\,.$$

They find that in the optimal case, it falls out that the SNRs become balanced, that is, $SNR_1 = SNR_2$. The authors go on to show that the phases of the individual relay nodes are essentially irrelevant (they are always a static linear combination of the phases of the transceivers, and therefore do not change over time), and that the ideal weights depend only on the ideal amplitude α. With these simplifications, the optimization problem can be reduced to its distributed form, which is dependent only on local information (f_i, g_i, b_i) and scalar values transmitted over a common control channel $(\xi_1, \xi_2, \xi_3, \alpha^T b)$.

$$\alpha_i = \left(\frac{|f_i|^2}{1+\xi_1} + \frac{|g_i|^2}{1+\xi_2} + \frac{\beta\xi_3}{(1+\xi_1)+(1+\xi_2)}\right)^{-1} \frac{b_i}{(\alpha^T b)}\,.$$

When an array is large, it is possible that only certain nodes will be selected to cooperate in the beamforming. Subsection 4.1 discusses this topic in length, but it is common to select nodes, which approximate a uniform array, and apply a least squares estimation to the weights to correct the non uniformities. In [21], the authors present a system that uses this method, with the weight estimation distributed over the nodes. Though the entire steering matrix is still needed to calculate the ideal weights, the processing involved with calculation of the matrix itself is be distributed over the nodes. A statistical method for distributed weight calculation is presented in [22] and discussed in the next Subsection 3.4, on statistical analysis.

3.4. Statistical Analysis

Because the nodes in a distributed array might be randomly placed, it is useful to look at the average beampattern capabilities of distributed virtual arrays. The assumption in this avenue of research is that given a large enough set of nodes, there will be some subset that is capable of performing at least as well as the mean, giving a strong set of design criteria for ubiquitous distributed networks where the number of possible nodes is very high. In [23], an initial analysis of the average beam pattern for a random array is presented. The array is uniformly distributed over a disc, and derivations for both the average and distribution of the achievable beampattern is presented. The properties of both the main lobe and first side lobe are investigated. They show that on a good distribution, the beampattern is capable of approaching a main lobe with gain N and a sidelobe with gain $1/N$ where N is the number of cooperating nodes. Specifically, the beampattern approaches

$$P(\phi) = \frac{1}{N} + \left(1 - \frac{1}{N}\right)\left|2\frac{J_1\left(4\pi R \sin\left(\frac{\phi}{2}\right)\right)}{4\pi R \sin\left(\frac{\phi}{2}\right)}\right|^2,$$

where J_1 is the 1st order Bessel function. The first $1/N$ term is the average sidelobe gain, and is the minimum of the average pattern, while the $(1 - 1/N)$ term contributes to the main lobe.

Figure 7 shows the average beampattern for a variety of scenarios with different disc sizes (R) and N_s. It can be seen that when N equals 16 and 256, the average sidelobes are equal to $1/N$ ($-10\log_{10}(16) = -12.04$ and $-10\log_{10}(256) = -24.08$ respectively). If the nodes are distributed in a non-uniform manner, the mean (and distribution) of the main lobe will clearly change. Gaussian distributed nodes have a smoother mean curve for both the main lobe [19] and side lobe [25] areas of their pattern.

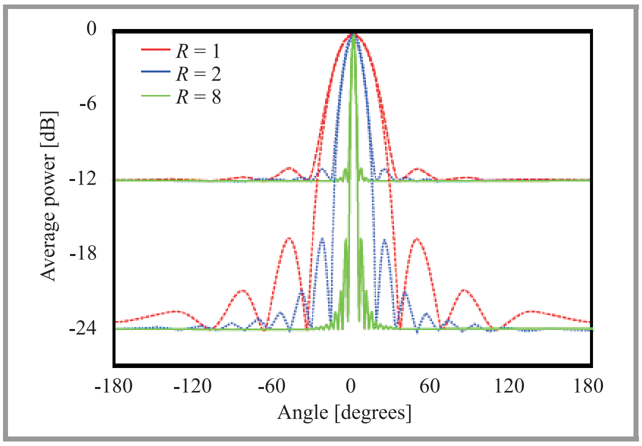


Fig. 7. Average beampattern for random networks [23].

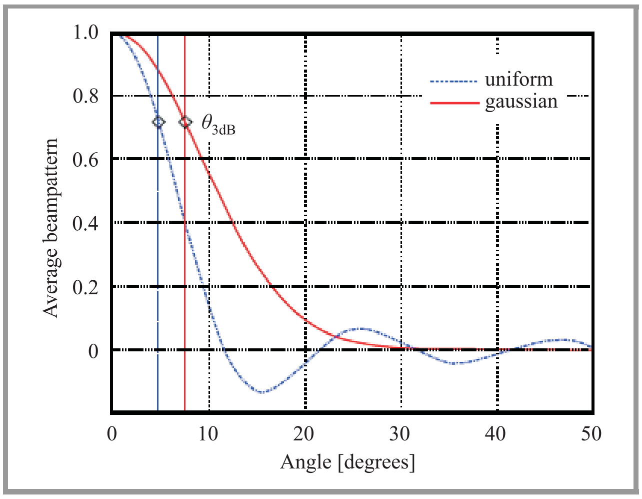


Fig. 8. Peak and sidelobe comparison [19].

This means that the Gaussian nodes have a wider main lobe, but the average pattern outside of the main lobe does not oscillate. Figure 8 shows a comparison of the average beampattern from a Gaussian and uniformly distributed random network. The smoother sidelobe achieved in the Gaussian distribution is clear to see, as is the wider main lobe. In comparison to the above equation, the average beampattern for the Gaussian network is

$$P(\phi) = \frac{1}{N} + \left(1 - \frac{1}{N}\right) \left| e^{-\frac{4\pi \sin\left(\frac{\phi}{2}\right)\sigma^2}{2}} \right|^2 .$$

In this case there are no sidelobe oscillations from the Bessel function, only a smooth approach to the minimum. In [26], the authors extend the analysis of uniform distribution to include the concept of null steering, looking at the average interference rejection capabilities of distributed nodes without knowledge of the array geometry. Approximations for the locations of the other nodes based on the their random distribution are used, and achieve similar performance results when the number of nodes is high due to the fact that the approximations are based on the values of the beamformer as that number goes to infinity. When nodes are mobile in addition to being randomly distributed, the ideal weights will change with time. In [27], the modeled beampattern from the preceding paper [23] is simplified to an ideal pattern where the gain is the maximum (N or N^2 depending on the arrangement) in the mainlobe beam width and the minimum ($1/N$ or N) everywhere else. The capacity of the network under this simplified assumption is derived and simulated as compared to a traditional pie wedge. The inclusion of the sidelobes in the approximation gives a better estimation of the capacity than the pie wedge. The authors of [28] present a method for determining the time required between updates when the nodes move with respect to a particular theoretical motion distribution. These models, based on the number of nodes and mobility distribution, can give guidelines based on the required fidelity at the receiver.

4. Practical Considerations in Distributed Beamforming

4.1. Node Selection and Placement within Virtual Arrays

The coordination of distributed nodes to find optimum weights for beamforming focuses on finding weights for every node cooperating in the solution, but it is not always prudent for every node in a network to cooperate in a given transmission. Often, a given subset of the network is capable of transmitting with the same quality as the entire network. Each extra node is using energy in the transmission, and is adding extra local overhead in the pre-transmission phase. In energy constrained wireless networks, this extra battery drain is unacceptable.

The topic of optimizing the size of a beamforming cluster is presented in [29], where the total energy for a cluster is calculated, based on the number of cooperating nodes. This paper presents an excellent introduction into the processes to take into account when selecting nodes for ideal power consumption. It is shown that there are an optimum number of cooperating nodes to obtain the minimum energy expenditure for the same quality of link. If the amount of power required to receive (P_R), transmit (P_T), and process a frame (P_{s1}) are known, then N nodes can collaborate using frames of length L_1 at rate R utilizing energy equal to

$$P_{col} = (2N-3)(P_T + P_R + P_{s1})\left(\frac{L_1}{R}\right).$$

During their responses of length L_2, the nodes will consume

$$P_{res} = (N-2)(P_T + P_R + P_{s1})\left(\frac{L_2}{R}\right).$$

Finally, the nodes will collaborate to reach the destination by sending their messages, consuming a total of

$$\begin{aligned} P_{tot} &= (P_T + P_R + P_{s1})\left[(2N-3)\left(\frac{L_1}{R}\right) + (N-2)\left(\frac{L_2}{R}\right)\right] \\ &\quad + NP_{s2}\left(\frac{L_2}{R}\right). \end{aligned}$$

Because each of the variables is known, the equation can be minimized with respect to N, giving the least amount of required power to transmit the frame. This is shown to be convex and to have one global minimum. This is because when the number of nodes is small, the amount of required energy to reach the source per node is very high, but as the number of nodes increases, the energy cost of collaboration becomes prohibitive. In [30], the authors take a different approach. They provide a method for node selection with an emphasis on carrier synchronization. Although this topic is covered in depth in Subsection 4.2, those methods focus on synchronization of the selected nodes. In this paper, the authors select nodes based on their relative phases, rather than try to synchronize the phases directly. Nodes are clustered by their carrier phases to provide synchronized transmission, optimizing the selection criterion to maximize the received energy as opposed to minimizing transmission energy. Because the phase offsets from some reference should be evenly distributed across all of the nodes, you will find that there are even numbers of nodes in each phase group, distributing the beamforming load across the entire network. In [31] the authors present a scheme that adaptively changes the number of beamformers to maintain the optimum ergodic capacity of the network. As channel information is fed back to the collaborating nodes, the overhead grows with the number of nodes. If this overhead is large enough, there will be an optimal number of nodes to participate in the beamforming process. The authors show that the capacity of a beamforming link is described by

$$C = \log_2\left(1 + P_T \sum_{i=1}^{N} L_i\right),$$

where L_i is the large scale fading factor in the channel.

Utilizing the large scale fading factor from each node, the network can decide which users should are hindering the process, and selectively coordinate only the beamformers who increase the capacity at a given point in time based on their large scale fading factors. When nodes are chosen within an array for collaboration, they must be able to communicate with one another. As shown in [24], it is often the case that the ideal beamforming nodes will be out of communications range with one another, requiring the use of relay nodes and increasing the overhead of the node synchronization phase. By appropriately selecting nodes for beamforming based on their relative locations to one another, rather than solely on their fitness in terms of the desired beam, communications between the relay nodes can be maintained. If a specific area is required for the cooperative nodes in 4-order to maintain a good beam, the nodes from the edges of the necessary area should be chosen, providing a strong mesh network around the perimeter, rather than a loose network across the entire area. The achievable beampattern of this randomly chosen group of nodes is similar to that derived in [26], however there is no longer the chance that an iteration of the random process will have nodes in the middle of the disc. As such, the new average beam pattern is equal to

$$P(\phi) \approx \frac{1}{K} + \left(1 - \frac{1}{K}\right)\left|J_0\left(4\pi R_o \sin\left(\frac{\phi}{2}\right)\right)\right|^2,$$

where R_o is the outer radius of the disc.

This perimeter selection method is expanded upon in [26], where a series of concentric circles are chosen from the center outward to provide several strongly linked group of collaborative nodes with different capabilities in terms of possible main lobes and interference rejection based on the equation for $P(\phi)$ above.

Utilizing only certain nodes within a wireless sensor network to perform as a virtual array allows for a distribution of work to help prolong the life of a power constrained group of nodes. By finding the optimum number of beamformers and shutting down the transmission beyond the point of diminishing returns and by reducing the pretransmission overheard, the overall power consumption in these networks can be reduced in an effective manner.

4.2. Carrier Synchronization

With traditional beamforming systems, when each of the array elements are controlled by a single source, the carrier signal of each of the elements is assumed to be of the same frequency and phase, so the ideal weights calculated can make the necessary phase adjustments from the same baseline. Obviously, it is very difficult to assume that a distributed array of independent sensors in an array would have the same carrier phase across the whole network just by chance. Accommodations for the differences in the carrier must be made, either through direct synchronization of the carriers, or through changes in the calculated weights at each node. Though this problem is very important, indeed at the heart, of distributed virtual arrays; it is also important in a variety of applications in the wireless sensor network domain, including certain sensing applications, distributed space-time block coding, inter node relays, and timing applications. As such, there is a wealth of information regarding the carrier synchronization of nodes in a local distributed wireless sensor network. Because of this, a brief review of the carrier synchronization literature with specific applications to distributed virtual antenna arrays is presented.

The initial literature on distributed beamforming focused on the calculation of ideal weights for a non-uniform array, leaving the carrier synchronization of the nodes as a problem for the future. Paper [32] gives a general introduction to a two node beamformer which automatically adjusts the phase between the two collaborating transmitters so their transmitted symbols sum constructively at a receiver. A system is presented which allows the two nodes to synchronize via master slave architecture, and to "precode" their transmissions with a measured channel response, what we would call beamforming weights. This is a good general model for carrier synchronization in a distributed beamforming system, and is a theme that is often repeated. In [33], a test bed was built to monitor the performance of distributed acoustic beamforming for locating sources of noise. This initial, practical implementation showed the potential success for distributed beamforming without perfectly synchronized carriers, the authors of [34] and [35] sought to provide an analytical estimation of the limitations that distributed beamforming systems might see from unsynchronized carriers. When a reference carrier is transmitted from a master node, the effects of different phase differences are summed in the received carrier

$$u_i(t) = \cos\left(2\pi f_0 t + \theta_o + \theta_e(i)\right) + n_i(t),$$

where θ_o is the static offset between the carriers, θ_e is the error in the phase due to transmission and placement errors and n is the transmission noise.

Both analytical and numerical results were presented for a master-slave architecture in which cooperating nodes lock their carriers to a master based on the received carrier $u_i(t)$. Each of the slave nodes is then able to adjust its own carrier based on this received waveform and the known distance between the two nodes. Guidelines are also presented as to the limitations of such a system when there are estimation errors in the received carrier and measured inter nodal distances. Paper [36] introduces a system in which a copy of the carrier is transmitted continuously from a each of receiving nodes, who merge their carriers on this distributed reference, requiring a significant amount of additional hardware at each cooperating node. Improvements to [36] are implemented in [37], where the number of required transmitting beacons was reduced (along with hardware complexity). This reduction comes in the form of carrier transmissions over time slots, effectively turning the carrier synchronization beacons into TDMA users who share the channel to broadcast their carrier information. With only a single bit a feedback from the receiver, an iterative approach to carrier synchronization can be carried out [38]. The authors introduce a system where the nodes are synchronized through the difference of groups of phases. In each iteration, nodes are assigned randomly to one of to groups, which transmit their data sequentially. The aggregate phase difference is calculated and transmitted to the second group, who update their own phases by this correction factor. After each iteration, the total synchronization of the entire set is closer than it was before. In comparison to the individual random perturbation scheme from [39], the pair wise updating method converges much more quickly, though it requires extra feedback information. By restricting the random search space, the new algorithm will converge quickly, but will not necessarily be able to guess the correct phase in a single iteration. However, because such feat has a very small probability anyway, so restricting the space leads to an overall improvement. Analysis and simulations include the performance gains when the search space (possible offsets) is restricted to a particular probability distribution, as well as the implementation of the algorithm on software-defined radios. Rather than attempt to synchronize the carriers of cooperative nodes at all, [40] presents a method in which the unsynchronized carriers do not matter. By having the cooperating nodes simply repeat each symbol several times, there will be a point when the carriers constructively interfere naturally, which can be detected at the receiver. Analysis of the number of repetitions required based on the number of cooperating nodes is presented, and numerical results show that the probability of alignment, and hence the number of repetitions required, reaches a steady state point for a specified number of cooperating nodes.

Though carrier synchronization is necessary in order to perform optimum beamforming in distributed networks, the preceding papers have shown that the problem is not as daunting as it seems. Through various combinations of data sharing between nodes, feedback from the end sinks, and statistical analysis of the networks the carrier synchronization problem is not insurmountable. Further research into optimum methods for carrier synchronization can only improve the quality of beamforming in distributed networks, but the problem is well defined.

5. Future Work in Distributed Smart Arrays

Though the topic of distributed smart arrays has been studied from a high level to ensure that appropriate complex weights for individual nodes can be calculated optimally, the high level approach leaves significant gaps in the path toward utilizing DB in a non-coherent application like a wireless sensor network. Before DBF can be used in systems such as wireless sensor networks, personal area networks, and even mobile phone networks, further research into what might be physically capable by these networks is needed. The topics below present a brief cross section of some of the open problems in distributed beamforming that are available to researchers in the wireless communications field.

5.1. Achievable Beampatterns

In the current body of distributed smart array work, most attention is focused on the maximum achievable extension of the main lobe, and methods for steering that main lobe towards an intended receiver while directing nulls in the directions of interferers. A virtual array created from a wireless sensor network will have a significant number of elements available to it. This should allow for a beampattern that is capable of multiple beams and nulls. The capabilities of randomly distributed arrays in terms of amplifying and nullifying multiple incoming sources is a topic that would have strong applications, as a network rarely needs to interact with a single sink, or a single source of interference.

5.2. Cognitive Radio

As the ubiquity of sensor networks grows, so will the chance of utilizing bandwidth in an area where it is already assigned to some other entity. Cognitive radio attempts to diminish or eliminate interference with the primary of a particular channel and insert secondary communications into the spaces between primary transmissions. A virtual array should be able to find an ideal set of weights that can be used to eliminate interference at the primary user's location allowing the array to communicate at the same time as the primary user. Further investigation into the applicability of distributed phased arrays for the purposes of cognitive radio is essential to allowing wide spread operation of sensor networks in densely populated areas.

5.3. Heterogeneous Node Types

If distributed arrays are to be expanded beyond the scope of wireless sensor networks and be applied in other network types such as a personal area network (PAN) or a group of military units, the assumption of identical cooperating nodes must be dropped. For example, a deployed group of soldiers may form a network between themselves, a local radio relay, and support vehicles. Each of these types of units will have a different antenna type on their equipment and different capabilities in terms of constraints on maximum transmit power, which is something not considered in the current body of literature. Further investigation into the cooperation of multiple node types can lead to distributed smart antennas in networks where it is currently considered impractical, and research into the methods for optimizing these heterogeneous networks will allow them to operate longer, possibly even longer than a similarly sized homogeneous network. Additionally, investigation into the situations where a homogeneous vs. heterogeneous group of contributors would perform better will lead to an advantage when designing systems that use distributed smart antennas. For example, it may be beneficial for a designated cluster head in a network to have a different antenna type than the cooperating radiators to optimize the radiation pattern of the virtual array.

5.4. Number of Cooperators

The current body of research into distributed smart arrays focuses on networks of either many, many nodes that all cooperate at once, or limited groups that take turns cooperating. Little attention is paid to the number of nodes in a beamforming array; for example, it is highly likely that there is a rule of diminishing returns in distributed arrays, where adding extra cooperating nodes will not provide an adequate performance gains to justify their power expenditure. Research into the optimum number of nodes to balance the power usage with the desired beampattern will allow networks to prolong their battery life by limiting the amount of unnecessary power expended. Additionally, research into adaptive algorithms capable of dynamically adjusting the number of cooperating nodes required to reach a receiver could maximize the efficiency.

5.5. Channel Estimation

Though the topic of channel estimation between two wireless radios is well researched, there may be benefits to be derived from estimating several channels in a distributed array environment. Given a sufficient distance between the information sink and the network, the channels between the sink and each node should be similar. It may be possible to find new ways to estimate the channel between the sink and each node in a distributed fashion, such as finding an average channel for the virtual array, and then deriving the individual channels from the average based on the known topology of the network.

5.6. Carrier Synchronization

By far, the largest roadblock to distributed antenna arrays is synchronization of the cooperating nodes. There has already been headway into this area, but further research into not only methods of synchronization, but the effects of unsynchronized carriers, is necessary to ensure that the theoretical research into distributed arrays can be applied to its fullest.

References

[1] H. Unz, "Linear arrays with arbitrarily distributed elements", *IRE Trans. Anten. Propag.*, vol. 8, no. 2, pp. 222–223. 1960.

[2] S. Sandler, "Some equivalences between equally and unequally spaced arrays", *IRE Trans. Anten. Propag.*, vol. 8, no. 5, pp. 496–500, 1960.

[3] A. Ishimaru, "Theory of unequally-spaced arrays", *IRE Trans. Anten. Propag.*, vol. 10, no. 6, pp. 691–702, 1962.

[4] Y. Lo and R. Simcoe, "An experiment on antenna arrays with randomly spaced elements", *IEEE Trans. Anten. Propag.*, vol. 15, no. 2, pp. 231–235, 1967.

[5] Y. Lo and S. Lee, "Sidelobe level of nonuniformly spaced antenna arrays", *IEEE Trans. Anten. Propag.*, vol. 13, no. 5, pp. 817–818, 1965.

[6] A. Maffett, "Array factors with nonuniform spacing parameter", *IRE Trans. Anten. Propag.*, vol. 10, no. 2, pp. 131–136, 1962.

[7] Y. Lo, "A mathematical theory of antenna arrays with randomly spaced elements", *IEEE Trans. Anten. Propag.*, vol. 12, no. 3, pp. 257–268, 1964.

[8] M. Donvito and S. Kassam, "Characterization of the random array peak sidelobe", *IEEE Trans. Anten. Propag.*, vol. 27, no. 3, pp. 379–385, 1979.

[9] A. Panicali and Y. Lo, "A probabilistic approach to large circular and spherical arrays", *IEEE Trans. Anten. Propag.*, vol. 17, no. 4, pp. 514–522, 1969.

[10] P. Fertl, A. Hottinen, and G. Matz, "A multiplicative weight perturbation scheme for distributed beamforming in wireless relay networks with 1-bit feedback", in *Proc. IEEE Int. Conf. Acoust., Speech and Sig. Proces. ICASSP 2009*, Taipei, Taiwan, 2009, pp. 2625–2628.

[11] M. Abdallah and H. Papadopoulos, "Beamforming algorithms for information relaying in wireless sensor networks", *IEEE Trans. Sig. Proces.*, vol. 56, no. 10, pp. 4772–4784, 2008.

[12] Y. Zhang, X. Li, and M. Amin, "Distributed beamforming in multiuser cooperative wireless networks", in *Proc. Fourth Int. Conf. Commun. Netw. ChinaCOM 2009*, Xi'an, China, 2009, pp. 1–5.

[13] C. Shi, R. Berry, and M. Honig, "Local interference pricing for distributed beamforming in MIMO networks", in *Proc. IEEE Military Commun. Conf. MILCOM 2009*, Boston, USA, 2009, pp. 1–6.

[14] Y. Jing and H. Jafarkhani, "Beamforming in wireless relay networks", in *Proc. Inf. Theory Appl. Worksh.*, Petersburg, Russia, 2008, pp. 142–150.

[15] C. Tsinos and K. Berberidis, "An adaptive beamforming scheme for cooperative wireless networks", in *Proc. 16th Int. Conf. Dig. Sig. Proces.*, Santorini, Greece, 2009 , pp. 1–6.

[16] A. El-Keyi and B. Champagne, "Collaborative uplink transmit beamforming with robustness against channel estimation errors", *IEEE Trans. Veh. Technol.*, vol. 58, no. 1, pp. 126–139, 2009.

[17] S. Fazeli-Dehkordy, S. Shahbazpanahi, and S. Gazor, "Multiple peer-to-peer communnications using a network of relays", *IEEE Trans. Sig. Proces.*, vol. 57, no. 8, pp. 3053–3062, 2009.

[18] G. Zheng, K.-K. Wong, A. Paulraj, and B Ottersten, "Robust and distributed beamforming", in *Proc. Int. Conf. Wir. Commun. Sig. Proces. WCSP 2009*, Nanjing, China, pp. 1–5.

[19] M. Ahmed and S. Vorobyov, "Node selection for sidelobe control in collaborative beamforming for wireless sensor networks", in *Proc. IEEE 10th Worksh. Sig. Proces. Adv. Wir. Commun. SPAWC'09*, Perugia, Italy, pp. 519–52.

[20] V. Havary-Nassab, S. Shahbazpanahi, A. Grami, and Z.-Q. Luo, "Distributed beamforming for relay networks based on second-order statistics of the channel state information", *IEEE Trans. Sig. Proces.*, vol. 56, no. 9, pp. 4306–4316, 2008.

[21] N. Papalexidis, T. Walker, C. Gkionis, M. Tummala, and J. McEachen, "A distributed approach to beamforming in a wireless sensor network", in *Conf. Rec. 41st Con. Asilomar Conf. Sig. Syst. Comp. ACSSC 2007*, Pacific Grove, USA, 2007, pp. 606–610.

[22] K. Zarifi, S. Affes, and A. Ghrayeb, "Collaborative null-steering beamforming for uniformly distributed wireless sensor networks", *IEEE Trans. Sig. Proces.*, vol. 58, no. 3, pp. 1889–1903, 2010.

[23] H. Ochiai, P. Mitran, H. Poor, and V. Tarokh, "Collaborative beamforming for distributed wireless ad hoc sensor networks", *IEEE Trans. Sig. Proces.*, vol. 53, no. 11, pp. 4110–4124, 2005.

[24] K. Zarifi, A. Ghrayeb, and S. Affes, "Distributed beamforming for wireless sensor networks with improved graph connectivity and energy efficiency", *IEEE Trans. Sig. Proces.*, vol. 58, no. 3, pp. 1904–1921, 2010.

[25] M. Ahmed and S. Vorobyov, "Performance characteristics of collaborative beamforming for wireless sensor networks with Gaussian distributed sensor nodes", in *Proc. IEEE Conf. Acoust. Speech Sig. Proces. ICASSP 2008*, Las Vegas, USA, pp. 3249–3252.

[26] K. Zarifi, S. Affes, and A. Ghrayeb, "Distributed Proces. techniques for beamforming in wireless sensor networks", in *Proc. 3rd Int. Conf. Sig. Circ. Syst. SCS 2009*, Jerba, Tunisia, pp. 1–5.

[27] R. Harrington, "Correction to sidelobe reduction by nonuniform element spacing", *IRE Trans. Anten. Propag.*, vol. 9, no. 6, pp. 576–576, 1961.

[28] W. Lintz, J. McEachen, and M. Tummala, "Sensor beamforming with distributed mobile elements in a wireless sensor network", in *Proc. Canadian Conf. Elec. Comp. Eng. CCECE'09*, St. John's, Newfounland and Labrada, Canada, 2009, pp. 323–328.

[29] P. Vincent, M. Tummala, and J. McEachen, "Optimizing the size of an antenna array", in *Proc. 14th Asilomar Conf. Sig. Syst. Comp. ACSSC'06*, Pacifik Grove, USA, pp. 2281–2284.

[30] J. C. Chen, K. Yao, T. L. Tung, C. W. Reed, and D. Chen, "Source localization and tracking of a wideband source using a randomly distributed beamforming sensor array", *Int. J. High Perform. Comput. Appl.*, vol. 16, no. 3, pp. 259–272, 2002.

[31] N. Zhang, G. Kang, Y. Guo, P. Zhang, and X. Gui, "Adaptive transmitted distributed antenna selection strategy in distributed antenna systems with limited feedback beamforming", *Electron. Lett.*, vol. 45, no. 21, pp. 1079–1081, 2009.

[32] Y. Tu and G. Pottie, "Coherent cooperative transmission from multiple adjacent antennas to a distant stationary antenna through AWGN channels", in *Proc. IEEE Veh. Technol. Conf. VTC Spring 2002*, Birmingham, USA, 2002.

[33] J. Chen, K. Yao, and R. Hudson, "Source localization and beamforming", *IEEE Sig. Proces. Mag.*, vol. 19, no. 2, pp. 30–39, 2002.

[34] G. Barriac, R. Mudumbai, and U. Madhow, "Distributed beamforming for information transfer in sensor networks", in *Proc. Third Int. Symp. Inf. Proces. Sensor Netw. IPSN 2004*, Berkeley, USA, 2004, pp. 81–88.

[35] R. Mudumbai, D. Brown, U. Madhow, and H. Poor, "Distributed transmit beamforming: challenges and recent progress", *IEEE Commun. Mag.*, vol. 47, no. 2, pp. 102–110, 2009.

[36] D. Brown, G. Prince, and J. McNeill, "A method for carrier frequency and phase synchronization of two autonomous cooperative transmitters", in *Proc. IEEE 6th Worksh. Sig. Proces. Adv. Wir. Commun.*, New York, USA, 2005, pp. 260–264.

[37] I. Ozil and D. Brown, "Time-slotted round-trip carrier synchronization", in *Conf. Record 41st Asilomar Conf. Sig. Syst. Comp. ACSSC 2007*, Pacific Grove, USA, 2007, pp. 1781–1785.

[38] P. Jeevan, S. Pollin, A Bahai, and P Varaiya, "Pairwise algorithm for distributed transmit beamforming", in *Proc. IEEE Int. Conf. Commun. ICC'08*, Beijing, China, 2008, pp. 4245–4249.

[39] R. Mudumbai, J. Hespanha, U. Madhow, and G. Barriac, "Distributed transmit beamforming using feedback control", *IEEE Trans. Infor. Theory*, vol. 56, no. 1, pp. 411–426, 2010.

[40] A. Bletsas, A. Lippman, and J. Sahalos, "Simple, zero-feedback, collaborative beamforming for emergency radio", in *Proc. 6th Int. Symp. Wir. Commun. Syst. ISWCS 2009*, Siena, Italy, 2009, pp. 657–661.

Jason Uher holds the M.Sc. degree in electrical and computer engineering from Georgia Tech and is currently pursuing his Ph.D. at the University of Nebraska-Lincoln with a focus in wireless communications. His interests include MIMO block coding, polarization diversity, and heterogeneous beamforming in random antenna arrays.

e-mail: jasonuher@gmail.com
University of Nebraska-Lincoln
Omaha, USA

Tadeusz A. Wysocki (SM'98) received the M.Sc.Eng. degree with the highest distinction in telecommunications from the Academy of Technology and Agriculture, Bydgoszcz, Poland, in 1981 and the Ph.D. degree and the D.Sc. degree in telecommunications from the Warsaw University of Technology, Warsaw, Poland, in 1984 and 1990. Since fall 2007, he has been with the University of Nebraska-Lincoln as Professor of Computer and Electronics Engineering at the Peter Kiewit Institute, Omaha, NE. The main areas of his research interests include space-time signal processing, diversity combining, indoor propagation of microwaves, as well as protocols for wireless ad hoc networks.

e-mail: twysocki@mail.unomaha.edu
University of Nebraska-Lincoln
Omaha, USA

Beata J. Wysocki received the M.Eng. degree in electrical engineering from the Warsaw University of Technology, Warsaw, Poland, in 1990 and the Ph.D. degree from the Australian Telecommunications Research Institute at Curtin University of Technology, Perth, Australia, in 2000. Her research interests include space-time signal processing, sequence design for direct sequence (DS) code-division multiple-access (CDMA) data networks, and optimization of ultra-wide-band (UWB) communication systems.

e-mail: beatajoanna33@yahoo.com
University of Nebraska-Lincoln
Omaha, USA

Information for Authors

Journal of Telecommunications and Information Technology **(JTIT)** is published quarterly. It comprises original contributions, dealing with a wide range of topics related to telecommunications and information technology. **All papers are subject to peer review**. Topics presented in the JTIT report primary and/or experimental research results, which advance the base of scientific and technological knowledge about telecommunications and information technology.

JTIT is dedicated to publishing research results which advance the level of current research or add to the understanding of problems related to modulation and signal design, wireless communications, optical communications and photonic systems, voice communications devices, image and signal processing, transmission systems, network architecture, coding and communication theory, as well as information technology.

Suitable research-related papers should hold the potential to advance the technological base of telecommunications and information technology. Tutorial and review papers are published only by invitation.

Manuscript. TEX and LATEX are preferable, standard Microsoft Word format (.doc) is acceptable. The author's JTIT LATEX style file is available:
http://www.nit.eu/for-authors

Papers published should contain up to 10 printed pages in LATEX author's style (Word processor one printed page corresponds approximately to 6000 characters).

The manuscript should include an abstract about 150–200 words long and the relevant keywords. The abstract should contain statement of the problem, assumptions and methodology, results and conclusion or discussion on the importance of the results. Abstracts must not include mathematical expressions or bibliographic references.

Keywords should not repeat the title of the manuscript. About four keywords or phrases in alphabetical order should be used, separated by commas.

The original files accompanied with pdf file should be submitted by e-mail: redakcja@itl.waw.pl

Figures, tables and photographs. Original figures should be submitted. Drawings in Corel Draw and PostScript formats are preferred. Figure captions should be placed below the figures and can not be included as a part of the figure. Each figure should be submitted as a separated graphic file, in .cdr, .eps, .ps, .png or .tif format. Tables and figures should be numbered consecutively with Arabic numerals.

Each photograph with minimum 300 dpi resolution should be delivered in electronic formats (TIFF, JPG or PNG) as a separated file.

References. All references should be marked in the text by Arabic numerals in square brackets and listed at the end of the paper in order of their appearance in the text, including exclusively publications cited inside. Samples of correct formats for various types of references are presented below:

[1] Y. Namihira, "Relationship between nonlinear effective area and mode field diameter for dispersion shifted fibres", *Electron. Lett.*, vol. 30, no. 3, pp. 262–264, 1994.

[2] C. Kittel, *Introduction to Solid State Physics*. New York: Wiley, 1986.

[3] S. Demri and E. Orłowska, "Informational representability: Abstract models versus concrete models", in *Fuzzy Sets, Logics and Knowledge-Based Reasoning*, D. Dubois and H. Prade, Eds. Dordrecht: Kluwer, 1999, pp. 301–314.

Biographies and photographs of authors. A brief professional author's biography of up to 200 words and a photo of each author should be included with the manuscript.

Galley proofs. Authors should return proofs as a list of corrections as soon as possible. In other cases, the article will be proof-read against manuscript by the editor and printed without the author's corrections. Remarks to the errata should be provided within one week after receiving the offprint.

Copyright. Manuscript submitted to JTIT should not be published or simultaneously submitted for publication elsewhere. By submitting a manuscript, the author(s) agree to automatically transfer the copyright for their article to the publisher, if and when the article is accepted for publication. The copyright comprises the exclusive rights to reproduce and distribute the article, including reprints and all translation rights. No part of the present JTIT should not be reproduced in any form nor transmitted or translated into a machine language without prior written consent of the publisher.
For copyright form see: http://www.nit.eu/for-authors

A copy of the JTIT is provided to each author of paper published.

Journal of Telecommunications and Information Technology has entered into an electronic licencing relationship with EBSCO Publishing, the world's most prolific aggregator of full text journals, magazines and other sources. The text of *Journal of Telecommunications and Information Technology* can be found on EBSCO Publishing's databases. For more information on EBSCO Publishing, please visit www.epnet.com.

(Contents Continued from Front Cover)

Editorial Office

National Institute
of Telecommunications
Szachowa st 1
04-894 Warsaw, Poland

tel. +48 22 512 81 83
fax: +48 22 512 84 00
e-mail: redakcja@itl.waw.pl
http://www.nit.eu

INSTYTUT ŁĄCZNOŚCI
PANSTWOWY INSTYTUT BADAWCZY