Infocommunications Journal

A PUBLICATION OF THE SCIENTIFIC ASSOCIATION FOR INFOCOMMUNICATIONS (HTE)

March 2015 Volume VII Number 1 ISSN 2061-2079

Technically Co-Sponsored by IEEE COMMUNICATIONS SOCIETY

hte

IEEE HUNGARY SECTION

Indexing information

Infocommunications Journal is covered by Inspec, Compendex and Scopus.

Infocommunications Journal

Technically co-sponsored by IEEE Communications Society and IEEE Hungary Section

Supporters

GÁBOR BÓDI – president, National Council for Telecommunications and Informatics (NHIT)
GÁBOR MAGYAR – president, Scientic Association for Infocommunications (HTE)

Editorial Oce (Subscription and Advertisements)**:**
Scientic Association for Infocommunications
H-1051 Budapest, Bajcsy-Zsilinszky str. 12, Room: 502
Phone: +36 1 353 1027, Fax: +36 1 353 0451
E-mail: info@hte.hu • Web: www.hte.hu

Articles can be sent also to the following address:
Budapest University of Technology and Economics
Department of Networked Systems and Services
Te l.: +36 1 463 3261, Fax: +36 1 463 3263
E-mail: szabo@hit.bme.hu

Subscription rates for foreign subscribers: 4 issues 50 USD, single copies 15 USD + postage

Publisher: PÉTER NAGY • Manager: ANDRÁS DANKÓ

HU ISSN 2061-2079 • Layout: MATT DTP Bt. • Printed by: FOM Media

TC-linearisation of tweakable polynomials

Josef Bárta, Michal Hojsík

***Abstract*—Based on the Cube Attack by Itai Dinur and Adi Shamir and another, in the essence similar, method we devised a new polynomial linearisation technique, which proved to be more powerful, than the Cube Attack alone. Moreover, we present detailed description with formal proof not only of our findings, but also of the Cube Attack. Finally, we demonstrate the results of our efforts on a Trivium variant that is reduced in key and initialisation vector bit count. We managed to linearise polynomials representing a keystream bit output after up to 621 initialisation rounds using purely techniques described in this paper, compared to 581 initialisation rounds with original attack.**

***Index Terms*—Cube Attacks, cryptanalysis, stream ciphers, lightweight cryptography, Boolean functions, linearisation, tweakable polynomials**

I. INTRODUCTION

IN this paper we present a detailed description of Adi Shamir's Cube Attack and then to devise a generalisation, which could help push the boundaries of usability of the Cube Attack. Other important target of ours are of course the polynomials as such. Therefore we decided to actually compute the polynomial expression of the state and keystream bits of Trivium reduced in the number of the bits used as variables and then to analyse them without having to do any guessing. More specifically, we wanted to assess, whether the polynomials are linearisable using the techniques devised by us and whether they are any more effective than the original Cube Attack.

In the next section there is described the basic notation we use throughout the paper. Thereafter we "translate" the Cube Attack into our notation and we present its detailed description. In further sections we describe in the same manner another technique that can be used for attack in a similar way to the Cube Attack, which we simplify into two easier, but nonetheless effective techniques. The details about analysis of the polynomials and a description of the cryptosystem they represent can be found in the second to last section.

II. TC-LINEARISATION OF TWEAKABLE POLYNOMIALS

In this section we describe the theory behind the Cube Attack. Further in the text we define a technique, which is in itself very simple, but in its full variant could prove to be a very powerful way of linearising polynomials, if it was not for the computational complexity of the algorithm the equivalent condition yields. Nevertheless, we also present two simple variants, one of which proves in the next section to be quite powerful, when teamed up with the Cube Attack.

Supported by grant VF20102015006

Manuscript submitted on 11 September 2014, accepted 20 February 2015.
The authors are with Department of Algebra, Faculty of Mathematics and Physics, Charles University in Prague, 186 75 Praha 8, Sokolovska 83, Czech Republic, josef@barta.cz, hojsik@karlin.mff.cuni.cz

A. Tweakable polynomials

In this section, we introduce some notation and define the classes of polynomials we will be working with.

Throughout this paper, we denote by $[n]$ the set $\{0, 1, \ldots, n-1\}$ for any $n \in \mathbb{N}$. The set of all Boolean functions in n variables is denoted by $\mathcal{B}_n$, i.e. $\mathcal{B}_n = \{f : \mathbb{F}_2^n \to \mathbb{F}_2\}$.

Algebraic normal form (ANF) of a Boolean function f is its representation as a polynomial $f(x_0, \ldots, x_{n-1}) \in \mathbb{F}_2[x_0, \ldots, x_{n-1}]$ such that none of its monomials contain any variable in degree greater than one. For each Boolean function, there exist a unique algebraic normal form.

For $I \subseteq [n]$, we will use x_I to denote the monomial $\prod_{i \in I} x_i$. So for every Boolean function $f \in \mathcal{B}_n$ there exists a unique set $\mathcal{I} \subseteq \mathcal{P}([n])$ such that $f(x_0, \ldots, x_{n-1}) = \sum_{I \in \mathcal{I}} x_I$. We will write $x_I \in f$ if $I \in \mathcal{I}$.

Definition II-A.1. *Let $m, n \in \mathbb{N}$. We define set of secret variables $X = \{x_i; i \in [n]\}$ and set of public variables $Y = \{y_j; j \in [m]\}$.*

Later on, the secret variables will represent the secret key, while the public variables will represent the initialisation vector of a stream cipher, which is public and can be potentially set by an attacker.

In the rest of the paper we will use the ANF representation of Boolean functions and use the notation $\mathcal{B}[X]$, $\mathcal{B}[Y]$, $\mathcal{B}[X, Y]$ for Boolean functions (polynomials) in variables X, Y or $X \cup Y$ respectively.

Definition II-A.2. *We call a polynomial p tweakable, if $p \in \mathcal{B}[X, Y]$ and fully tweakable, if $p \in \mathcal{B}[Y]$.*

B. Basic Cube Attack

This section describes the basic principles of the Cube Attack in the same way it was done in [1]. For demonstration purposes we use fully tweakable polynomials.

Definition II-B.1. *[1] Let $p \in \mathcal{B}[Y]$ be a polynomial and $J \subseteq [m]$ a variable index subset. A superpoly of J in p is a polynomial $p_{S(J)} \in B[Y]$ such that*

$$p(Y) = y_J \cdot p_{S(J)}(Y) + q_J(Y) \qquad \text{(II-B.1)}$$

where $q = \sum_{J \not\subseteq J'} b_{J'} y_{J'}$, $b_{J'} \in \mathbb{F}_2$. We call y_J a maxterm, if the superpoly $p_{S(J)}$ is a linear, non-constant polynomial.

Note II-B.2. *The superpoly $p_{S(J)}$ does not contain any variables indexed by J.*

Example II-B.3. *Let $p \in \mathcal{B}[Y]$ be a polynomial*

$$p = y_0y_3y_4y_5 + y_1y_3y_4y_5 + y_0y_3y_5 + y_0y_2y_3y_4 + y_0y_1 + y_1y_2y_4 + y_2y_3y_4 + y_3y_4 + y_3y_5 + y_3 + y_4 \ .$$

We can factor out the monomial $y_J = y_3y_4y_5$ so we get

$$p = \underbrace{y_3y_4y_5}_{y_J} \cdot \overbrace{(y_0 + y_1)}^{p_{S(J)}(Y)} + \underbrace{+ y_0y_3y_5 + y_0y_2y_3y_4 + y_0y_1 + y_1y_2y_4 + y_2y_3y_4 + y_3y_4 + y_3y_5 + y_3 + y_4}_{q_J(Y) \text{ continued}}$$

In this case, y_J is a maxterm of J in p.

Definition II-B.4. *For an index subset $J \subseteq [m]$, $|J| = k$ we define a summation cube C_J as the set of k-tuples of variables $y_j : j \in J$ where all possible combinations of values of variables y_j are assigned. We can also understand C_J as a vector space $\mathbb{F}_2^k$ with information about indices of the variables. Hence we set $dim(C_J) = k$.*

Definition II-B.5. *[1] For every polynomial $p \in \mathcal{B}[Y]$ and for any k-dimensional summation cube C_J, $J \subseteq [m]$ we define $p_J := \sum_{v \in C_J} p_{|v}$ where $p_{|v}$ is a derived polynomial with $m - k$ variables $\{y_j : j \in [m] \setminus J\}$ and the variables indexed with J are assigned values from the k-tuple v.*

Now we can present a vital property of the superpoly of J in p, which is the main theorem in [1].

Proposition II-B.6. *[1] For any polynomial $p \in \mathcal{B}[Y]$ and variable subset J, $p_J = p_{S(J)}$.*

For our purposes, from now on, we shall call this technique of summing (partial) evaluations of a polynomial the **C-linearisation of fully tweakable polynomials**.

C. C-linearisation of tweakable polynomials

In this section we describe the C-linearisation (cube attack) on tweakable polynomials. We present a clear description of what makes a polynomial C-linearisable. In [1] this part was skipped, for they dealt with black box polynomials which demand a different approach than polynomials the explicit representation of which is known.

Definition II-C.1. *We call a polynomial $p \in \mathcal{B}[X, Y]$ C-linearisable, if there exists $J \subseteq [m]$ such that $p_J(X, Y = (1, ..., 1))$ is linear.*

For purposes of C-linearisation we present the following grouping of monomials: Let $p \in \mathcal{B}[X, Y]$ be a tweakable polynomial. Then we can write

$$p = \sum_{(I,J) \in \mathcal{I}} x_I y_J = \sum_{l \in L_p} l + \sum_{b \in B_p} b + \sum_{h \in H_p} h$$

where $\mathcal{I} \subseteq \mathcal{P}([n]) \times \mathcal{P}([m])$ and $B_p = \{x_I y_J \in p : |I| = 0\}$, $L_p = \{x_I y_J \in p : |I| = 1\}$ and $H_p = \{x_I y_J \in p : |I| > 1\}$.

The set B_p contains all monomials consisting purely of public variables and the free monomial. L_p contains all monomials consisting of exactly one secret and any number and combination of public variables. H_p consists of monomials with two or more secret variables. We can plainly see that $L_p \cup B_p \cup H_p$ contains all monomials of p.

Before we present the condition which describes precisely a C-linearisable polynomial, we present a simple lemma about C-linearisability:

Lemma II-C.2. *Let $p \in B[X, Y]$ be a tweakable polynomial. If $L_p = \emptyset$, then p is not C-linearisable.*

Proof. If there is no monomial that is linear in secret variables, there is definitely no monomial y_J, $J \subseteq [m]$, such that $p_{S(J)}$ is linear in secret variables. □

Now we can propose an equivalent definition of a C-linearisable tweakable polynomial:

Proposition II-C.3. *A tweakable polynomial $p \in \mathcal{B}[X, Y]$ is C-linearisable if and only if*

$$\exists x_i y_J \in L_p, \ : (\forall y_{J'} x_I \in H_p : y_J \not| y_{J'})$$

Proof. We shall prove the first implication by contradiction, the second directly:
"⇒": For contradiction, we assume that p is C-linearisable and $\forall y_J x_i \in L_p \exists y_{J'} x_I \in H_p : y_J | y_{J'}$. This implies that for every choice of J will in the superpoly $p_{S(J)}$ remain a monomial that is non-linear in secret variables, i.e. the superpoly will contain $\frac{y_{J'}}{y_J} x_I$ and $|I| \geq 2$ as $y_{J'} x_I \in H_p$. Thus the contradiction.
"⇐": We assume that $\exists y_J x_i \in L_p \ \forall y_{J'} x_I \in H_p : y_J \not| y_{J'}$. That implies y_J is a maxterm, which yields a superpoly $p_{S(J)}$ that is linear in secret variables. □

Example II-C.4. *In this example we rewrite the polynomial from previous example into the notation of the tweakable polynomials with distinguished secret and public variables. We shall have $m = n = 3$. So let $p \in \mathcal{B}[X, Y]$ be a tweakable polynomial:*

$$p = x_0y_0y_1y_2 + x_1y_0y_1y_2 + x_0y_0y_2 + x_0x_1y_0y_1 + x_1x_2y_1 + x_2y_0y_1 + y_0y_1 + y_0y_2 + y_0 + y_2$$

We can factor out $y_I = y_0y_1y_2$, so we obtain

$$p = y_0y_1y_2 \cdot (x_0 + x_1) + x_0y_0y_2 + x_0x_1y_0y_1 + x_1x_2y_1 + x_2y_0y_1 + y_0y_1 + y_0y_2 + y_0 + y_2$$

where $x_0 + x_1$ is the linear superpoly of $I = \{0, 1, 2\}$ in p and y_I is a maxterm.

D. T-linearisation of tweakable polynomials

Now we present T-linearisation, a technique we devised and describe to aid the C-linearisation to be as effective as possible when linearising a polynomial.

Definition II-D.1. *We call a polynomial $p \in \mathcal{B}[X, Y]$ T-linearisable, if*

$$\exists J \subseteq [m] : (\exists v \in C_J : p_{|v} \text{ is linear in secret variables})$$

In other words there exists a (partial) evaluation of the polynomial in public variables that results in $p_{|v}$ being linear in secret variables.

Proposition II-D.2. *Let $p \in \mathcal{B}[X, Y]$. If*

$$\exists l \in L_p,\ l = x_i y_J :\ (\forall h \in H_p\ \exists j \in [m] \setminus J : y_j | h)$$

then p is T-linearisable. Specially, we say that p is T1-linearisable.

Proof. If there is such l that the condition holds, then every $h \in H_p$ can be eliminated by setting respective $y_j = 0$ while the secret part of l will be kept in the polynomial by setting the public variables indexed by a smallest $J' \subseteq J$ such that $x_i y_{J'} \in L_p$ to one and the remaining ones to zero. □

We recall the polynomial from previous example to demonstrate the T1-linearisation:

Example II-D.3. *Let*

$$p = x_0y_0y_1y_2 + x_1y_0y_1y_2 + x_0y_0y_2 + x_0x_1y_0y_1 + x_1x_2y_1 + x_2y_0y_1 + y_0y_1 + y_0y_2 + y_0 + y_2$$

We can linearise this polynomial in secret variables by setting $y_1 = 0$, which gives us a non-constant polynomial, that is linear in secret variables

$$p_{|y_1=0} = x_0y_0y_2 + y_0y_2 + y_0 + y_2$$

and finally, by setting $y_0 = y_2 = 1$ we get

$$p_{|y_1=0, y_0=y_2=1} = x_0 + 1$$

which is a linear polynomial in secret variables only.

Corrolary II-D.4. *C-linearisability does not imply T1-linearisability.*

Proof. Consider polynomial $p = x_0y_0y_1 + x_0x_1y_1$. This polynomial is clearly not T1-linearisable, but it is obviously C-linearisable using $J = \{0, 1\}$. □

Clearly T1-linearisability is not a necessary condition for T-linearisability. Consider $p \in \mathcal{B}[X, Y]$,

$$p = x_0x_1y_0 + x_0x_1 + x_2 = (y_0 + 1)x_0x_1 + x_2.$$

This polynomial obviously is T-linearisable by setting $y_0 = 1$, but due to the monomial x_0x_1 T1-linearisation does not work here.

Definition II-D.5. *For any index subset $I \subseteq [n]$ and polynomial $p \in B[X, Y]$ we define the set of public monomials of p relative to I as $E_p(I) = \{y_J : x_I y_J \in p\}$ and the tweaking polynomial $p_{E_p(I)} \in \mathcal{B}[Y]$ as $p_{E_p(I)} = \sum_{y_J \in E_p(I)} y_J$.*

Using this we can express any tweakable polynomial $p \in \mathcal{B}[X, Y]$ as $p = \sum_{I \subseteq [n]} p_{E_p(I)} x_I$. In other words, $p_{E_p(I)}$ is the coefficient of x_I in p if seen as $p \in \mathcal{B}[Y][X]$.

Clearly, if we want to obtain a linear polynomial in X, we need to evaluate all $B[Y]$ coefficients of $x_I,\ |I| \geq 2$ to zero. In the following proposition we use this to present an equivalent definition of T-linearisation.

Proposition II-D.6. *Tweakable polynomial $p \in \mathcal{B}[X, Y]$ is T-linearisable if and only if*

$$\exists J \subseteq [m] :$$
$$(\exists v \in C_J : [\forall I \subseteq [n],\ |I| \geq 2 : p_{E_p(I)|v} = 0$$
$$\wedge\ (\exists I' \subseteq [n],\ |I'| = 1 : p_{E(I')|v} \neq 0)])$$

In other words, a tweakable polynomial is T-linearisable, if there is a solution to the system of polynomial equations yielded by $p_{E_p(I)}$'s, where $|I| \geq 2$ for that some of the $p_{E_p(I')},\ |I'| = 1$ does not evaluate to zero.

Proof. We prove the forward implication by contradiction, the backward directly.
"⇒": Let's assume that p is T-linearisable and

$$\forall a \in \mathbb{F}_2^m\ [(\exists I \subseteq [n], |I| \geq 2 : p_{E_p(I)}(a) = 1)$$
$$\vee\ (\forall I' \subseteq [n], |I'| = 1 : p_{E(I')}(a) = 0)]$$

That means that after any partial evaluation in public variables there either remains some monomial that is not linear in secret variables or the resulting polynomial is a constant. Hence the contradiction.
"⇐": If there exists such J and $v \in C_J$, then we can eliminate all the monomials that are non-linear in secret variables by partial evaluation in v and there is at least one monomial, that is linear in secret variables that remains in the polynomial after the partial evaluation. So $p_{|v}$ is a polynomial that is linear in secret variables, hence p is T-linearisable. □

This proposition yields a very compelling way of T-linearising a tweakable polynomial. First, we find the solutions for the system of polynomial equations defined by the $p_{E_p(I)}$'s for $|I| \geq 2$. Then we choose those solutions, for which there exist I' such that $|I'| = 1$ and $p_{E_p(I')}$ is non-zero after the partial evaluation. Naturally, this may be ineffective or even impossible, as shown in following example:

Example II-D.7. *Because our usual example polynomial is, as demonstrated, T-linearisable, for purposes of this example, we present a different polynomial, $q \in \mathcal{B}[X, Y]$, $q = x_0 + x_0x_1y_0 + x_1x_2y_0 + x_1x_2$. Obviously,*

$$q = x_0 + x_0x_1 \cdot y_0 + x_1x_2 \cdot (y_0 + 1)$$

which yields an equation system

$$y_0 + 1 = 0$$
$$y_0 = 0$$

which has no solution.

Because solving a system of polynomial equations over $\mathbb{F}_2$ in general is computationally ineffective, we present a simpler version, which we might be able to solve in a more efficient manner (given that the system of equations actually has a solution, otherwise we just conclude that there is none).

Corrolary II-D.8. *Let $p \in \mathcal{B}[X, Y]$. If*

$$\forall x_I y_J \in H_p : |J| \leq 1 \wedge$$
$$\wedge\ [\exists a \in \mathbb{F}_2^m : (\forall I' \subseteq [n], |I'| \geq 2 : p_{E(I')}(a) = 0) \wedge (\exists i \in [n] : p_{E(\{i\})}(a) = 1)]$$

then p is T-linearisable. Specially, we say that p is T2-linearisable.

This means, that a polynomial is T2-linearisable, if is T-linearisable and the tweaking polynomials for all I's, such that $|I| \geq 2$, are linear or constant.

Corrolary II-D.9. *T2-linearisability does not imply T1-linearisability.*

Proof. Polynomial $p = x_0y_0y_1 + x_0x_1y_0 + x_0x_1 = x_0y_0y_1 + x_0x_1(y_0 + 1)$ is clearly T2-linearisable (set $y_0 = y_1 = 1$), but not T1-linearisable. □

Corrolary II-D.10. *T1-linearisability does not imply T2-linearisability*

Proof. Polynomial $p = x_0y_0 + x_0x_1y_0y_1$ is clearly T1-linearisable (set $y_1 = 0$), but not T2-linearisable. □

E. TC-linearisation of tweakable polynomials

In this section we present TC-linearisation, our generalisation of Shamir's Cube Attack's C-linearisation.

In order to proceed to the definition of a TC-linearisable polynomial, we first define more general version of a maxterm.

Definition II-E.1. *Let $J \subseteq [m]$ be an index subset. We call the monomial y_J a T1-/T2-/T-maxterm, if the superpoly of J in p is a T1-/T2-/T-linearisable polynomial.*

Definition II-E.2. *Let $p \in \mathcal{B}[X, Y]$ be a tweakable polynomial. Then p is TC1-/TC2-/TC-linearisable if and only if there exists $J \subseteq [m]$ such that y_J is a T1-/T2-/T-maxterm respectively.*

Note, that TC1-/TC2-/TC-linearisation with $J = \emptyset$ equals T1-/T2-/T-linearisation since $p_{S(\emptyset)} = p$.

Example II-E.3. *In this example we use the same polynomial $p \in \mathcal{B}[X, Y]$ as previously:*

$$p = x_0y_0y_1y_2 + x_1y_0y_1y_2 + x_0y_0y_2 + x_0x_1y_0y_1 + x_1x_2y_1 + x_2y_0y_1 + y_0y_1 + y_0y_2 + y_0 + y_2$$

This polynomial is TC-linearisable, because we can either set $y_0 = y_2 = 1,\ y_1 = 0$ and obtain a non-constant linear polynomial $x_0 + 1$ by T-linearisation only or get for example $x_0 + x_1$ by summing over the cube defined by $J = \{0, 1, 2\}$. There is also the possibility of using a combination of both, as is made possible using TC-linearisation:

$$p = y_0 \cdot (x_0y_1y_2 + x_1y_1y_2 + x_0y_2 + x_0x_1y_1 + x_2y_1 + y_1 + y_2 + 1) + x_1x_2y_1 + y_2$$

and then we set $y_1 = 0 \wedge y_2 = 1$ to get x_0 as our linear polynomial. In this particular case it would of course be more efficient to use T-linearisation only, because x_0 is

Note II-E.4. *We call a tweakable polynomial $p \in \mathcal{B}[X, Y]$ TC1-/TC2-/TC-linearisable, if we can derive a polynomial that is linear in secret variables from it by using partial evaluation as described in the equivalent definitions of T1-/T2-/T-linearisation and cube summation presented as C-linearisation combined.*

Clearly, if a tweakable polynomial does not contain any monomials linear in secret variables, then we can not apply any of the presented techniques:

Corrolary II-E.5. *Let $p \in \mathcal{B}[X, Y]$ be a tweakable polynomial. If $L_p = \emptyset$, then p is not T-, C- or TC-linearisable.*

III. LINEARISING TRIVIUM KEYSTREAM POLYNOMIALS

With the previous one dealing with the underlying theory, this section describes the experimental part of the paper. At first we describe the cryptosystem we will be attacking and after that we present the attack itself. Since we wanted only to test our linearisation methods, we have not attempted the key-recovery, which is the aim of the full attack. In other words we show, how far we got with the techniques devised by us and compare them to the Cube Attack.

A. Trivium

Before we present the attack itself, we describe the cryptosystem we are about to attack. It is a reduced variation of a stream cipher Trivium [4]. At first we describe the original cryptosystem and after that we describe reduced variation we will attack.

Trivium is a very simple stream cipher with three non-linear feedback registers, 80 bit key and 80 bit initialisation vector (IV). The cipher produces a keystream $\{z_i\},\ z_i \in \mathbb{F}_2$ which is added to the plaintext to produce the ciphertext. The detailed description is to be found in [4].

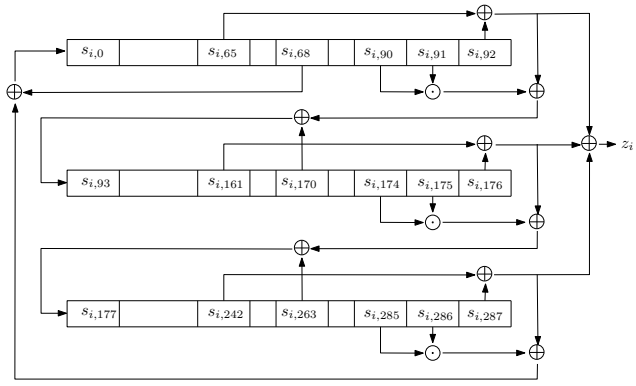


Fig. 1. Trivium cipher scheme

For our endeavour we had to reduce the original cryptosystem. Trivium-8 is a version of Trivium with shortened key and IV to 8 bits each. Aside from the shortened key and IV, it is the same Trivium as described above, so it has a 288 bit inner state. The length of the key and IV was chosen to be 8, i.e. $K = (k_0, \ldots, k_7)$ and $IV = (IV_0, \ldots, IV_7)$, since this should be enough to make the polynomials $p_i(X, Y)$ reasonably complex (interesting) while maintaining them small enough for the computation to be feasible with a standard PC.

In order to demonstrate the capabilities of the presented techniques we will assume that the keystream generation starts without any initialisation rounds (there are 1152 initialisation rounds in Trivium).

B. Attack description

As already mentioned, we will use the described linearisation methods to attack Trivium-8, a Trivium reduced in key and IV bits.

We assume that an attacker has access to Trivium-8 with fixed unknown key K. He can repeatedly choose the IV and

obtain the keystream $z_i = p_i(K, IV)$ (chosen IV attack). His goal is to find K by solving the respective equation systems.

For this purpose we need to compute the polynomials $g_{i,j}(X, Y)$ for all inner state bits and use them to compute the polynomials $p_i(X, Y)$ expressing the keystream bits. To obtain the actual values of z_i, we need an implementation of Trivium-8 that computes the keystream bits from the K and IV.

Attack on black-box polynomials, i.e. polynomials explicit representation of which is unknown to us, is out of scope of this paper and subject of future work.

1) Attack using T-linearisation: This part is pretty straightforward. Assume that $IV \in \mathbb{F}_2^m$ and $i \geq 0$ such that $p_i(X, IV)$ is linear in X. Then we get the value of $z_i = p_i(K, IV)$ and form a linear equation $p_i(X, IV) = z_i$.

If the J from the definition of T-linearisation is such that $J \neq [m]$, we add to the values of respective $v \in C_J$ values for the remaining public variables arbitrarily.

2) Attack using C-linearisation: Assume that for an $J \subseteq [m]$ the y_J is a maxterm of $p_i(X, Y)$ so by assigning ones to all y_j, $j \notin J$ we obtain a linear superpoly. We set $U = \{(u_0, \ldots, u_{m-1}) \in \mathbb{F}_2^m; u_j = 1 \ \forall j \notin J\}$ and obtain $z_i(a) = p_i(K, u)$ for all $u \in U$. The equation obtained by C-linearisation is then $p_{S(J)}(X, v) = \sum_{u \in U} z_i(u)$ where v is an element of U. Note that $p_{S(J)}(X, Y)$ can depend only on public variables with indices not in J, hence the equation does not depend on the particular choice of $v \in U$.

3) Attack using TC-linearisation: In C-linearisation, we set all public variables with indices not in J to one. In TC-linearisation, we assume that there exists $w \in C_{[m]\setminus J}$, such that $p_{S(J)|w}$ is linear. I.e. when summing over all k-tuples from the cube C_J, we have to extend each k-tuple with the T-linearising bits for partial evaluation, that remain fixed during the whole summing.

C. Experimental results

In this section we present results we got when we applied the presented linearisation techniques on Trivium-8 as presented in respective section.

Since solving a system of linear equations is simple, in our experiments we shall concentrate on whether we can actually obtain any linear, non-constant polynomials, from which we could build such.

For polynomial multiplication to build the representation we used a variant of algorithm from [2].

1) C-linearisation: The C-linearisation proved to be, as expected, very effective. We could use it to linearise the polynomials representing keystream bits with indices up to 609. It is important to note, that by not all keystream polynomials up to the 610^{th} are C-linearisable. Moreover, it is only effective up to the 582^{nd} keystream bit, because after that the linearising cube has dimension 8. making the linearisation uneffective.

2) T1-linearisation: T1-linearisation did not prove to be especially effective. We could T1-linearise polynomials representing up to the 361^{st} keystream bit.

This is actually a result we expected: If T1-linearisation would be effective on the polynomials, even if in relatively early stages of the initialisation, it would mean, that there would be no monomial $x_I y_\emptyset$, $I \subseteq [n]$. This is highly improbable though, because it would mean that none of a set of 2^n monomials would be present, which happens with probability of 2^{-2^n} if dealing with a random polynomial, which we assume the polynomials in the later stages of initialisation to be.

3) T2-linearisation: In spite of T2-linearisation using an approach that significantly differs from that used in T1-linearisation, we managed to linearise keystream polynomials only up to p_{361}. It is easy to see, why this technique was not any more successful: it demands, that for the polynomial $p \in \mathcal{B}[X, Y]$ that we are attempting to linearise every monomial in H_p has degree at most one in public variables. In fact then, this is a surprisingly good result.

4) TC1-linearisation: With TC1-linearisation, the situation is a bit more complicated. It is at least as effective as C-linearisation, but it could at some point prove to be able to reduce the cube and therefore the complexity of the linearisation. This happened only when the polynomial was T1-linearisable, so this technique turned out to be a bit disappointing.

5) TC2-linearisation: As we already hinted, TC2-linearisation is the technique, that really does improve the C-linearisation. We managed to linearise polynomials representing the keystream of Trivium-8 with indices up to 622, which is slightly more, than we managed using C-linearisation (609).

In the figure below, we present our results graphically. The dashed line denotes, where there are only such C-linearisable polynomials that are linearisable with a cube of dimension 8 only. Clearly, we consider the results achieved with TC2-linearisation to be a great success.

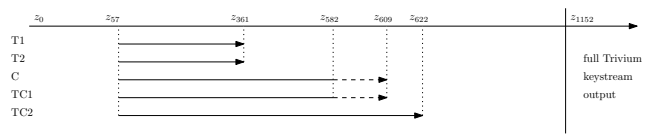


Fig. 2. Range of effectiveness of linearisation techniques

IV. CONCLUSION

In this paper we presented a detailed description of Cube Attack devised by Adi Shamir et. al. and its generalisation that proved to be slightly more effective when tested on key- and IV-reduced Trivium variant. However, none of these techniques is advanced enough to linearise keystream polynomials after full 1152 initialisation rounds.

REFERENCES

[1] Itai Dinur and Adi Shamir: Cube Attacks on Tweakable Black Box Polynomials, Cryptology ePrint Archive, Report 2008/385, 2008.
[2] Subhabrata Samajder and Palash Sarkar: Fast Multiplication of the Algebraic Normal Forms of Two Boolean Functions, International Workshop on Coding and Cryptography 2013, Bergen (Norway), 2013.
[3] Claude Carlet: Boolean Functions for Cryptography and Error Correcting Codes, chapter of the monography "Boolean Models and Methods in Mathematics, Computer Science, and Engineering", Cambridge University Press, 2006.
[4] Christophe de Cannière and Bart Preneel: Trivium, eSTREAM: the ECRYPT Stream Cipher Project, 2005.

Josef Bárta Josef Bárta has received bachelor degree in mathematics from Charles University in Prague and has been accepted to continue his studies at the Royal Holloway University of London.

His research interests include symmetric cryptanalysis, lightweight cryptography, authentication protocols and smart cards. He is currently working as a software engineer and preparing for his studies at the Royal Holloway.

Michal Hojsík Michal Hojsík has received master degree in mathematics from Charles University in Prague and PhD in computer science from University of Bergen, Norway.

His primary research interests are block ciphers and stream ciphers and lately also lightweight cryptography and authentication schemes. He is currently working as a cryptographic engineer.

CALL FOR PAPERS

Special Issue on Advanced wireless and mobile technologies and services

We have been witnessing a rapid development of wireless and mobile technologies and services during the past two decades. 4G mobile services are penetrating and mobile access is becoming an increasingly important way for accessing the Internet and it is expected to become the dominant one. The progress continues. 5G mobile systems are underway. Although many of the new technologies have already been incorporated in practical systems, there is still enough room for research and experimentation, in particular in the areas of cognitive radio, self-organizing networks, M2M communications, cross-layer optimization, just to name a few.

Topics of interest include but are not limited to:

- Cross-layer issues in wireless networks
- Cognitive radio for wireless communications
- QoS and resource allocation in wireless networks
- Mobile/wireless networks modeling and simulation
- Localization and positioning in wireless scenarios
- Topology control, self-organizing wireless networks
- Tools for modeling and analysis of wireless systems
- Personal wireless communications beyond 5G
- Software defined wireless networks and re-configurability
- M2M communications and the Internet of Things
- Storage, smart caching, and cloud for wireless
- Wireless social networks, participatory computing
- Molecular and nano-scale wireless communications
- New disruptive concepts for wireless systems

Selected papers from the European Wireless 2015 conference, *http://ew2015.european-wireless.org* will be invited to submit extended journal versions of their papers to this Special Issue, but high quality papers are welcome from open call too. Submissions will be peer reviewed according to the journal policy and international standards. Instructions for authors can be found on the journal website: *www.infocommunications.hu*.

Deadline for submission of manuscripts: June 30, 2015.

Tentative publication date: end of September, 2015.

Guest Editors:

SÁNDOR IMRE [M'93] is Professor and Head of Dept. of Networked Systems and Sevices at the Budapest University of Technology (BME). He obtained Dr. Univ. degree in in probability theory and statistics 1996, Ph.D. degree in 1999 and DSc degree from the Hungarian Academy of Sciences in 2007. He is Chairman of Telecommunication Scientific Committee of Hungarian Academy of Sciences. He participates on the Editorial Board of two journals: Infocommunications Journal and Hungarian Telecommunications. He was invited to join the Mobile Innovation Centre as R&D director in 2005. His research interests include mobile and wireless systems, quantum computing and communications. Especially he has contributions on different wireless access technologies, mobility protocols, security and privacy, reconfigurable systems, quantum computing based algorithms and protocols.

HASSAN CHARAF received his PhD in 1998. He is an Associate Professor and fellow at the Department of Automation and Applied Informatics at the Budapest University of Technology and Economics. He is the head of the IT group. As an outstanding figure in teaching, research and development, he is in key positions at several organizations at the university. His research fields are: distributed systems, cloud computing, multiplatform application development methods, software modeling and data technologies.

Network Coding Based Caching for Near Real-Time Streaming Media

Csaba Simon, Markosz Maliosz

***Abstract*— During crowded events streaming services generate high demands in the wireless access networks. In this paper we present a solution to offload the access network in case of such a streaming service. We detail the streaming service itself, and our offload solution based on local caching and network coding. We introduce a model that allows us to analyze our proposal, we implement it in a simulation environment and assess it. Finally we discuss the consequences of several design decisions we made during our work.**

***Index Terms*—multimedia applications, network communication, network coding, caching**

I. INTRODUCTION

The Internet traffic is dominated by streaming multimedia content as users demand higher quality video and ubiquitously available services. With the advent of high performance smart handheld devices the users expect that their usual services received on their desktops are available on these smart devices, too. Thus users can access advanced services from new places where they start to use their devices on regular basis. On turn, these new situations generate new demands: once the users get used to the new scenario, they start to require new, adapted services.

A typical scenario is a crowded event, where even a few years ago users could not use their mobile devices due to network congestion. E.g., it was common that during New Year's Eve calls were blocked and only SMS-es went through the overloaded networks. Similarly, sporting events at remote areas required a careful design and temporary increase in mobile access capacity to serve the increased demand. This motivated us to offload access networks during crowded events for a new streaming service, specific to this environment.

Users attend crowded events for the live experience, which combines the feeling of "being there" with the potential of rich social interactions among fellow users with similar interests. Nevertheless, until recently the participation at such events forced the attendants to stop following the online (e.g., live commentaries, additional info) and broadcasted (e.g., TV) content. The solution that offers both experiences, live attendance and online information stream, comes with the introduction of the so called "second screen".

Manuscript submitted September 29, 2014, revised February 22, 2015.

The authors are with the Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics, Budapest, Hungary, 1117 Budapest, Magyar Tudosok krt. 2. (e-mail: {simon|maliosz}@tmit.bme.hu)

Second screen originally refers to the use of an online device (e.g., smartphone, iPad) that doubles the screen of a device offering "linear" program (e.g., TV, projector). We extend the meaning of this term, calling second screen any online device that offers additional content associated with a live event, attended by the user of the device. Current access networks are hard pressed to provide the required QoS, because attendees of live events continue using their smart devices as second screens (to consume more and more multimedia content). In this environment, shortage of available capacity seems to perpetuate at least until the mid-2020s, when 5G technologies will mature. The focus of the operators is on assuring the basic service, not to mention any new service with additional bandwidth demand. Therefore offloading the wireless access currently is very important for the operators, and it will be so for the coming decade.

The data to be distributed in such an environment is not only the real-time, live multimedia stream, but also extra, add-on content, which has less strict delay constraints, and is related to recent events (e.g., replays, statistical analysis of the game, etc.). Still, their importance is higher soon after the original event happened (e.g., a goal right after it was scored), that is why we call them near real-time events. We propose a novel streaming service specific to this environment that can be offered on top of classical streaming media services, consisting of replayed live scenes. At the core of our solution to offload the access network delivering this service is the distributed local caching of the data, made reliable and versatile by the introduction of network coding techniques. To best of our knowledge, network coding was not proposed before to support such caching solutions (also see section III-C). The motivation behind such novel add-on services are not only recognized by recent research projects [1], but also attract major players from the streaming live event distribution industry [2].

In the next section we present related work that we relied on in our research. Then we present several scenario variants for our proposal and introduce the novel near real-time data delivering service that can be offered on top of classical streaming media services. In section IV we present a model that will allow us to analyze its behavior, and we evaluate it in section V. Finally we conclude our paper.

II. COMMUNICATION IN THE LOCAL WIRELESS DOMAIN

The support of new service types in such a challenging environment as crowded spaces needs a complex approach, which relies on results from several research areas of communications. In this section we briefly introduce the main aspects that influence the most our proposal and are referenced in later Sections during the definition of the model of our proposal. Specifically, the wireless technologies define the limitations of direct inter-node communication, while network coding ensures the flexibility and robustness of local data distribution.

A. Offloading the Local Wireless Access

Smartphones have several wireless interfaces that can be used to achieve direct communication. The natural choice is WiFi, with its adhoc variant. WiFi adhoc was very popular among researchers in the laptop era. Lots of mobile ad hoc protocols (MANET) were prototyped and investigated using such connections. Unfortunately this technology is not supported anymore by the vendors, although some Android smartphone models still can be tweaked to work in adhoc mode. The main advantage of the ad hoc mode is that it is very flexible.

Officially the replacement technology of the WiFi ad hoc mode is the WiFi Direct [3]. Nevertheless, the latter comes with some limitations, but these do not affect our scenarios. Both technologies have a different problem, too: the interfaces can work only in one WiFi mode only. Nevertheless, for modelling purposes we can use the WiFi Direct interface, as most of the community is familiar with this technology. Note that in real life deployments the local networking connection might be one of the UMTS/LTE technologies (e.g., using femtocells). Then the WiFi interface of the smartphone is available to support our service.

A different option might be the new variants of Bluetooth. The advantage is that typically this interface is not used, but it has lower capacity and it is harder to set up a link.

Finally we have to mention the promising new LTE variant, the LTE Direct [4] (or LTE D2D), which offers direct connectivity in a non-WiFi band, but it will cost more, as operates in a licensed band.

B. Network Coding

Network coding is a technique that, in contrast to channel coding, "allows and encourages the mixing of data at intermediate network nodes", instead of just encoding messages in a redundant way, allowing the network to have a maximum flow of information achieving a larger throughput [5]. With network coding, information transmitted from a source can be received by the receivers, but it can also be inferred or decoded. Intermediate nodes are still able to forward information but if it is the case, the node can combine different received streams of information into just one and transmit it to its outgoing nodes.

The fixed version of network coding uses simpler coding techniques (e.g., bitwise XOR-ing the packets of the involved stream(s)), making the flow always encodable and decodable, however this advantage comes with the downside of having to define the structure and the number of participants of the entire network previously. There is an alternative that follows a random behavior [6], where nodes assign coefficients to each packet randomly and according to the finite field used, there is a probability of these coefficients being decodable. Using random network coding all nodes are independent and randomized, without the need of any knowledge of the rest of the network. Intermediate nodes build a linear combination of incoming messages that then transmit on each on their outgoing links. Differently than the fixed method (e.g., bitwise XOR-ing), this combination uses independently and randomly chosen coefficients over a finite field. By allowing this kind of local encoding under a sufficiently large Galois field (i.e. finite field), the received coded blocks are decodable with a very high probability at the sink peers, on the order of the inverse of the size of the finite field [7].

The first practical wireless network coding scheme designed to deal with inter-flow traffic is also based on Random Linear Network Coding (RLNC) [8]. It exploits the shared nature of wireless medium and combines available data chunks with ones overheard from neighbors to restore the original information. Although it significantly improves network throughput, it is limited to situations where multiple streams cross the same network segment. A different approach uses RLNC to encode data within the same flow [9]. This intra-flow network coding improves the performance of the network over wireless links.

We divide the stream in generations. Only packets from a generation are linearly combined (encoded), thus at the receiving end enough linearly independent packets from a generation should be collected to be able to decode the content. For the streaming service it acts as a time window, because all packets must arrive before the playout of that particular encoded sequence can start. Any encoded packet belonging to a generation is useful (until we receive enough of them), the packets are not ordered, which simplifies the timing at the receiver side.

Network Coding was mostly proposed to be used in information distribution [10], multicast [11] and data averaging [12], which assures the usage in distributed sensor network data collection as well.

III. CACHING OF STREAMING MEDIA AT THE END NODES

A. Crowded Event Scenarios

We have identified three different scenarios for crowded events. All three scenarios offer a solid business model to build on and attract dedicated users who have the motivation to be actively involved in the content consumption process. E.g., they are interested in the details of the performers, want to know previous stories about the protagonists, etc. This offers a good audience for our proposed service.

The first scenario is the open air city festival, where attendees have access to multiple scenes and several selling and catering locations within a geographically limited area. Note that such events might become very congested, especially around sites of interest. We will refer to this scenario as the "festival" one.

A similarly crowded scenario is offered by the stadiums ("stadium" scenario). The main difference is that in the

stadiums the participants are bound to their seats and typically these events last for only a few hours. Therefore the people are not moving as much as in the previous scenario.

Combination of the previous two scenarios are the open air sporting events (bicycle tour, triathlon, etc). In such scenarios the attendance is scattered along the track, but usually they form small groups of people at interesting or spectacular portions of the track. Due to the largest such cycling event, (Tour de France), we will refer to this scenario as the "Tour" scenario.

Note that we expect different user behaviors and topologies in each of these scenarios, as detailed later in sub-section V-C.

B. Streaming Media Services

The traffic volume of streaming media had exceeded that of any other traffic type, including peer-to-peer or web access and researchers tried to reduce its bandwidth demand by various methods, including caching. There is a vast available literature in this field. In [17] we highlighted the most important ones. The reader interested in further details of streaming video caching is directed to a thorough overview of this field [14].

Our proposal is an additional service to extend the original streaming service. Currently replays are broadcasted within the original content, there is no possibility to watch them on-demand. In this paper we focus only on the service offering replays close to the moment in time when it originally happened (e.g., 100 minutes), which happens in a near real time fashion (in the worst case). In lots of cases the users want to re-watch the missed content, too. But in such cases only some special moments are of high interest, as the user then wants to resume and follow the original live content. Therefore recording the whole stream and playing it with a constant lag is not an option. In this case it is better to cache such short sequences from the live stream and make it available for instant replay. This caching service is detailed in the next sub-section.

C. Caching of the Replayed Content

In this section we introduce our solution for near real-time media streaming that also offloads the wireless access.

The load in the wireless access is decreased by the use of network coding and stretching the lifetime of the network encoded packets somewhere in the network distributed in end devices or well-placed points in the distribution/access domain. The cache distribution is implemented primarily on the user's devices.

This scenario is depicted in Fig. 1. The original live streaming media is distributed by the AP_i Access Points. The nodes receiving the data encode it and cache it locally (dark gray stars $-S_1$, …). Any time a node (light gray star) wants a replay, they will have the data chunks readily available in their local mesh network.

As already mentioned, the nodes should organize themselves to find the caches, this can be done using techniques used by peer-to-peer applications [13]. The advantage is that the neighbor list maintenance can be supported by the APs with limited extra costs in terms of bandwidth usage (e.g., neighbouring APs can exchange the list of connected devices that act as caches and broadcast that list periodically for all).

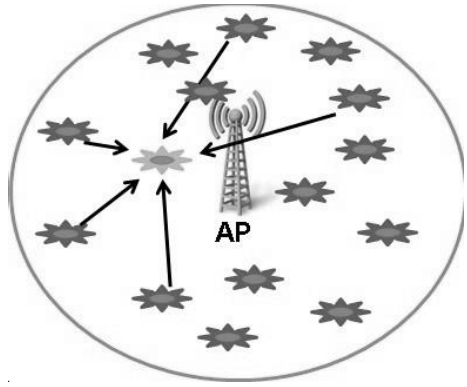


Fig. 1. Media streaming with caching

Also note that the direct node-to-node control traffic is not affecting directly the AP (for details also see sub-section V-A). In our scenario each node is attached to one AP and additionally it can communicate with several fellow nodes from its. We do not analyze the technological details, but two possible alternatives for such local communications to be largely adopted in the coming years are the WiFi Direct [3] and LTE Direct [4].

In our solution we use additional wireless sessions to receive the data from the caches, which run in parallel with the AP-to-node sessions. This might increase the number of collisions; some of these collisions are hidden by lower layers, others will result in packet losses. At higher layers (networking and above) these packet losses are perceived as a lower bandwidth. We measured the effect of this parallel communication, as described in section 5.1 and we used the results in our simulations. Thus during the evaluation of our model we take into account the impact of the addition of cache-to-node sessions through the modified data transfer rates.

IV. MODELING THE CACHE BASED STREAMING SERVICE

A. Network Coding Based Caching

From our model's perspective the important thing is to have a direct node-to-node communication in parallel with the AP-to-node connection. However we also have to take into account that a given node cannot communicate directly with every neighbor, the degree of connectivity is upper bounded. Currently the most used file distribution is BitTorrent, and the most used p2p streaming application (SopCast) is based on the same principles [21]. In BitTorrent only 5 slots are available for uploading data. The goal of this limitation is to protect the uploaders (i.e., the cache) from overload.

We consider that the potential coverage areas of the APs will be much larger than the area they accept connection requests from. Therefore, if there is a node somewhere at the border of two (or more) neighbouring AP coverage areas, then it can be identified by the respective APs. Based on that they can provide this node with the necessary information about cached content within its reach (but it will not "spam" this node with the list of far distant caches). The request statistics for a given replay or generally, for this type of service and similar statistics specific to streaming services can be computed in the background

[15][16]. Note that caching nodes can also report the read statistics to their APs. The details of the practical implementation of such services (e.g., the description of a protocol implementing it) are out of the scope of this paper. In our model we consider that these data are available, and it can be obtained from the AP a given node is connected to.

In our model the nodes that actually execute the network coding task are the same ones that work as caches. Due to the nature of RLNC we do not have to previously configure them. Therefore in our model we do not have to dedicate special attention to network coding aspects, nor to the selection process of the coding nodes. During the construction of our model we just had to focus on the caching aspects, the selection of the network coding nodes implicitly resulted from it.

Time dependencies are also not addressed directly, because the focus is on collecting the packets of a generation. Once collected enough packets, the content is readily available, implicitly meaning that the timing of content distribution conforms to the requirement of the service. If this requires slightly more time, then it can be considered as a slightly longer buffering time, which will not affect the (near real-time) service, because it has less strict delay requirements then a real-time one.

In the simplest model we search for the number of caching nodes to serve those nodes that are attached to the same AP, but we do not limit the cache size. Note that this solution does not allow that a node attached to AP_i to request data from a caching node attached to AP_j. In order to deduce the minimum number of overall caching points in the network, we can formulate an Integer Linear Program (ILP), as presented in the following sub-section. Then, in section IV-C we refine our model, aiming to minimize the size of the cache, at the same time letting more nodes to step in as caches, and we formulate an ILP for this case, too.

B. Minimizing the Number of Caching Nodes

We have a set of $\{AP_i\}$ Access Points, but for our model we should rather focus on the nodes. The nodes $\{v_1, ..., v_N\}$ and the direct links between them $\{ e_{ij} \}$ can be considered the vertices and the edges of a graph G. We define the connectivity matrix $\{a_{ij}\}$, where a matrix element is *1* if there is a direct connection between nodes i and j, *0* otherwise. Note that this can be obtained by recoding the original content locally with pseudo random coefficients.

Our goal is to determine the minimal number of caching nodes, a subset of G. c_i is the total capacity of the direct link from node i (we consider that the total incoming and outgoing capacities are equal). In order to decode the original content, we need at least a full generation of encoded packets (the size of a generation is noted with g_w). We introduce two binary variables; x_{ij} denotes a direct link between nodes v_i and v_j, $x_{ij} = 1$ if v_i sends the cached content to v_j, $x_{ij} = 0$ otherwise. Similarly, $u_i = 1$, if node i is a cache, $u_i = 0$ otherwise.

Our optimization problem is:

$$\text{minimize} \sum u_i \tag{1}$$

subject to

$$\sum_j x_{ij} \le c_i \quad \forall i \tag{2}$$

$$\sum_i x_{ij} \le c_j \quad \forall j \tag{3}$$

$$\sum_i x_{ij} \ge g_w \ \forall i, \forall j \tag{4}$$

$$x_{ij} \le a_{ij} \quad \forall i \ \forall j \tag{5}$$

$$x_{ij} \le u_i \quad \forall j \tag{6}$$

$$u_i, x_{ij} \in \{0,1\} \quad \forall i, \forall j \tag{7-8}$$

Equations (2) and (3) ensure that the caches and the regular nodes cannot exceed their total link capacities. Equation (4) ensures that all demand is served. Equation (5) assures that x_{ij} can be greater than 0 only if there is a direct physical connection between the nodes and eq. (6) that x_{ij} is greater than *0* only if the connection is originating from a cache.

The generic form of this kind of optimization problem is known as the geometric set cover problem, and has been continuously researched in the last decades. Based on the earlier research results it is hard to solve [18][19], and the most versions of the problem are still considered to be NP-hard [20]. Also, [25] states that 0-1 Integer Linear Programming (ILP) is NP-complete.

C. Minimizing the Cache Size

In the previous two sub-sections we analyzed the ways to minimize the number of caching nodes, but we did not restrict the size of the cache. In this subsection we try to minimize the size of the cache, but we do not restrict the number of caching nodes. Note that every node is a potential cache node, because every node plays the streaming content (we exclude those nodes that do not follow the video stream). Let us keep the same notations we introduced earlier in this section. We note the number of chunks stored at node i with k_i, and the number of actually downloaded encoded chunks from i to j with h_{ij} .

Now the objective is to

$$\text{minimize} \sum k_i \tag{9}$$

subject to

$$\sum_j x_{ij} \le c_i \quad \forall i \tag{10}$$

$$\sum_i x_{ij} \le c_j \quad \forall j \tag{11}$$

$$x_{ij} \le a_{ij} \quad \forall i \ \forall j \tag{12}$$

$$\sum_i h_{ij} \ge g_w \quad \forall j \tag{13}$$

$$0 =< h_{ij} <= x_{ij} M \quad \forall i, \forall j \tag{14}$$

$$k_i >= h_{ij} >= 0 \quad \forall i, \forall j \tag{15}$$

$$x_{ij} \in \{0,1\} \quad \forall i, \forall j \tag{16}$$

$$k_i, h_{ij} \in \mathrm{N} \quad \forall i, \forall j \tag{17}$$

We have (10), (11), (12) and (13), since conditions (1), (2), (5) and (8) are valid in this scenario, too. Eq. (13) shows the actually downloaded number of chunks for one node, v_j should at least be equal with g_w. If we select M sufficiently large (e.g., $M \geq g_w$), then eq. (14) states that downloads are possible only from selected caches. Eq. (15) says that the number of downloaded packets from a given cache is upper bounded by the content available at that cache.

V. EVALUATION OF THE MODEL

In this section we evaluate the proposed model and discuss the particularities of the proposed scenarios.

A. Offloading the Wireless Access

We evaluated the effect of the additional service on the original AP capacity. In this experiment we considered that the AP is using WiFi, while the nodes for their direct communication (i.e., cache access) use such technologies that uses the same frequency band (e.g., WiFi Direct or Bluetooth). We measured the impact of WiFi Direct on WiFi, when 5 to 20 streaming devices are connected to the AP and we had pairs of nodes testing WiFi Direct connections with iPerf. We found that if the two technologies run on different channels, the total capacity is relatively less affected when larger number of direct node-to-node pairs communicate. We used this value in our simulator to represent the effect of direct node-to-node communication on overall network load (e.g., packet losses due to collisions). Note that the combination of LTE Direct with WiFi APs yields better results in the favor of the distributed caching solutions. The co-existence of LTE-based streaming and LTE Direct was not assessed in this paper, as we focus on local wireless technologies. In this sub-section we compare three scenarios. The original one is when the replays are sent by the AP, which increases the load linearly with the number of requests.

The alternative solution is when caching is implemented without network coding. In this case the packets are sent directly from node to node, without directly consuming AP bandwidth. In this case there is a large control traffic overhead required to organize the download of the content. This case resembles the pure peer-to-peer streaming solutions, where control overhead in terms of number of packets is reported to vary between 5% and 20%, with the larger values for the leading streaming peer-to-peer application, SopCast [21][22]. The size of the control packets is one order of magnitude lower than the size of data packets, but p2p streaming applications contact many other peers, not only those they are downloading from. Note that based on our measurements, this 20% packet overhead is a conservative value, because at the beginning of downloads (starting to watch a replay) or when a seeder has to be replaced (churn event), the control traffic exceeds 2/3 of the total packet counts. As a consequence, we used a 15% overhead in terms of bandwidth (on the direct node to node links). In [21] they calculated with minimum 10% signalling overhead, SopCast having larger overheads.

The third case is the proposed network coding based caching. For RLNC based distribution in [23] the authors calculated with 5% overhead, but our scenario is simpler, because the infrastructure takes over some parts of the discovery and maintenance job and the peers are within direct layer 2 contact. Therefore we calculated with a minimal overhead (we used a 3% value in our simulations), as the requester also does not have to deal with the uniqueness of the segments.

The result of the evaluation is shown in Fig. 2, with linear trends fitted on all three data series. We simulated 50 requesting nodes the most, because a WiFi AP will not serve more than 100 streams, and out of this maximum number of connected nodes only a fraction of them will access the cache at the same time. It can be seen that both distributed caching solutions significantly offload the network, and scales well with the growing replay demands. Also we can see that the proposed network coding based caching solution outperforms both alternatives.

B. Optimization of the Caches

We have built a simulator to test the scenarios given in section IV in the different network conditions. We applied the graph libraries of the lemon tool [26] and the glpk and gurobi public ILP solver tools [27][28]. We generated the connectivity matrix considering that the network nodes were uniformly distributed. We have generated several different networks and averaged the solutions to get the presented results. We limited the upload capacity to 5 units.

The size of the encoded chunks should be less than the size of an UDP packet, somewhere around 1kB. The number of data

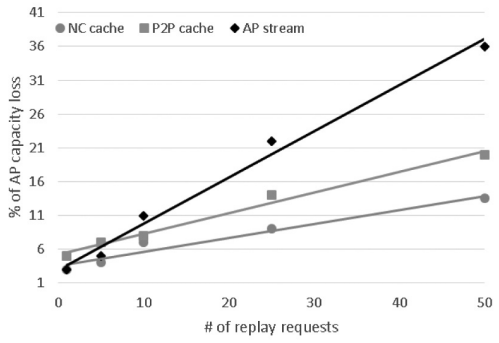


Fig. 2. Evaluation of AP offloading efficiency

chunks encoding the same generation (g_w) should be of orders of tens, eventually few hundreds.

These results give the theoretical bounds as a result. However, in practically feasible implementations, due to the distributed implementation, we can just approximate this result at the best. We have proposed heuristic algorithms for both optimization problems in [17], but in this paper we will analyse the ILP models only.

Because of the complexity of the problem, in order to allow the ILP solver to find the optimum, we used max. 100 nodes (N)

in the system, such as the diameter of the network is 5 hops in a dense scenario. As explained earlier, the number of nodes connected to a given AP should be less than the maximum capacity, in order to maintain the required QoS, so the above parameters would mean the deployment of 10 APs. We also investigated a sparse scenario, where nodes are farther away, and the maximum number of APs is 15. The number of nodes demanding the service was set to 20% of N, a worst case upper bound. The g_w parameter is also downsized to allow the solver to complete, and we set it for $g_w=10$, meaning that a node can not get all its chunks at once from a single cache.

TABLE I
OPTIMAL CACHE PARAMETERS FROM THE OPTIMIZATION RESULTS OF ILP MODELS

	N = 10	*N = 50*	*N = 100*
Total nr. of caches (dense)	*2*	*7*	*13*
Total nr. of caches (sparse)	*3*	*10*	*15*
Average cache sizes (dense)	*1*	*3*	*4*
Average cache sizes (sparse)	*2*	*4*	*5*

Table I presents the average values resulted from 5 successful runs for each scenario. For the first optimization problem we show the total number of caches in the network. For the second optimization problem we show the average cache size / each node in the network.

We can see that when we optimize on the number of caches, we get larger numbers compared to the situation when we would evenly assign the maximum number of requesting nodes to each cache. This occurs due to the randomness of the topology: some caches might not serve nodes at full speed, as the requesting node is out of its contact area. As the network grows in size, the requests are distributed more evenly on average, thus relatively fewer caches suffice. For the second optimization problem we see that the average number of chunks to be stored at nodes remains low.

C. Particularities of Different Networking Topologies

The three scenarios introduced in section III-A have direct implications on the underlying networking topologies. Actually, the difference is made by the users (participants at the event), whose actions and movement is constrained in different manners. Obviously, any categorization of such behavior simplifies the scenario and in the case of real life deployment the operator or service provider should conduct its own assessment on the user group it wants to serve. With this remark in mind we still can model with good accuracy the behavior of the users, which has direct effect on the networking topology and mobility of their handheld smart devices. In the following we will focus on the nodes, even if the decision on their position and mobility is taken by their respective owners.

We have analysed the requirements in each scenario and we recommended the appropriate caching method for each, as follows (also see Table II).

TABLE II
CACHING METHODS FOR DIFFERENT SCENARIOS

	Stadium	*Festival*	*Tour*
XOR based coding			**X**
Minimal nr. of caches	**X**		
Minimal cache size		**X**	

In the Tour scenario the nodes are partitioned in separate groups along the track of the competition. Practically this results in smaller, isolated groups of nodes. Additionally, once a viewer joins a group, she/he will stick to that group. We have in this group lower number of nodes and higher group stability, so we considered that in this particular scenario we should rethink the priorities based on which we selected the network coding method implemented in the caches. Note that the RLNC variant of network coding offers us scalability (we can bring in many coding nodes if more caches are required) and flexibility (we do not have to define in advance the roles among caches). Nevertheless, in this tour scenario it is worth considering the fixed network coding, which trades flexibility for simplicity. Practically this requires in addition the definition of the nodes that have to act as caches, encoding nodes and for each cache the nodes that are linked to them.

The lightest technique for fixed network coding is the bitwise XOR [5][7]. In order to confer some flexibility and robustness to the caches using XOR based coding we propose the following solution. First, several nodes should agree on serving as caches. Then they should divide the roles, some of them storing packets without encoding, while at least one of them should store bitwise XOR-ed packets. Once constructed this caching group, any requester should choose any combination of those caches in order to be able to replay the stream.

In the stadium scenario we have static nodes, bound to the seats and we have a high node density. But due to the large number of nodes we might use the advantages of statistical multiplexing of more sources compared to the tour scenario. Under such conditions we should try to use the RLNC based method and minimize the number of caches. In order to avoid the battery drain, periodically we should change the caching roles of the nodes, similarly to the top peer rotation mechanism in SopCast.

Compared to the previous two scenarios, in the festival scenario we have much higher node mobility, because the participants can walk within the festival area. Also, we have large number of nodes. This leads us to the use of RLNC based solution, minimizing the cache size. This also spreads content as much as possible, and the nodes are not forced to rely only on few caches (which would happen if we minimize the number of caches instead).

D. Discussions on the Caching Details

In the case of RLNC, one question is to set the size of the Galois field. In our earlier work we used GF(2^8). This means that for each packet we should attach a 1 byte coefficient, which would result in large overhead. Given that there is no need to encode the same packet multiple times (which would imply the encoding of the coefficients, too) we can apply the workaround proposed in [23]. It suffices to embed only the seed to be used to generate the series of random coefficients. Additionally we have to take care to use the same pseudo random number generator at each node. This trick allows us to reduce the coefficient related overhead to a mere four bytes, and this value remains constant, whatever the number of encoded segments might be.

In our model we did not include the effect of the distance between the wireless source and destination, although in some wireless technologies this might change the coverage area of the source. Also we did not consider the possibility of overhearing [24] (which might be considered as an implicit multicast packet distribution) that would further increase the efficiency of our proposal.

Note that when we minimize the number of caches we might gain a collateral advantage. Because the cache will operate at higher loads, it will be more efficient, since it avoids idle periods (in terms of data transfer), under which it still has to keep its wireless interface active, waiting for newer requests. Therefore from p.o.v. of green networking the first optimization problem corresponds to the maximization of a naïve green networking model. Nevertheless, the details of this relation should be further investigated (e.g., the effect of the receiver's distance from the source).

VI. CONCLUSION

In crowded events several scenarios are possible where streaming media based services are required. Due to their high bandwidth demand, these applications heavily stress the local access networks. In such cases any extra service results in dramatical QoS degradation. One possibility to support such services is to offload the access network by local, distributed caching mechanisms. We have proposed such a solution and built a model to investigate the behavior of our proposal. We found that it is more advantageous than a simple distributed caching solution and discussed the particularities of the proposed scenarios.

Our proposal allows the design and deployment of added value services for future large events at lower infrastructure costs. In our future work we plan to investigate the integration of caching and peer-to-peer mechanisms for the real time streaming media distribution, expecting that the application supporting near real-time services brings further advantages to the service providers.

ACKNOWLEDGMENT

Csaba Simon's research work was supported by the European Union and the State of Hungary, co-financed by the European Social Fund in the framework of TÁMOP 4.2.4.A/1-11-1-2012-0001 National Program of Excellence (NKP). The authors thank their colleagues Krisztián Németh and Attila Kőrösi for assistance in ILP modeling and simulation.

REFERENCES

[1] Cs. Simon, "In-network caching of media streams in access networks", (in Hungarian), National Program for Excellence (NKP) Newsletter, pp. 4., June 2014
[2] Cs. Simon, "Real-time Streaming Support in Crowds", EIT ICTLabs Partner Event – Future Networking Solution Workshop, April 2014
[3] Wi-Fi Alliance, "Wi-Fi Direct" - available from: http://www.wi-fi.org/discover-and-learn/wi-fi-direct
[4] Qualcomm White Paper, "LTE Direct Overview", http://www.qualcomm.com/media/documents/lte-direct-overview, July, 2013.
[5] R. Koetter, M. Medard, "An algebraic approach to network coding", IEEE Trans. on Networking, October 2003
[6] Li B, et al., "Random network coding in peer-to-peer networks: From theory to practice", 2011
[7] Gajic B, Riihijrvi J and Mhnen P., "Performance evaluation of network coding: Effects of topology and network traffic for linear and xor coding", Journal of Communication, vol. 4(11), pp. 885-893, 2009
[8] S. Chachulski and S. Katti, "Trading structure for randomness in wireless opportunistic routing", in Proc. of ACM SIGCOMM 2007
[9] S. Katti, D. Katabi, W. Hu, H. Rahul, and M. Medard, "The importance of being opportunistic: Practical network coding for wireless environments", in Proc. 43rd Annual Allerton Conference on Communication, Control, and Computing, 2005
[10] Gkantsidis, Ch., and Pablo R., "Network coding for large scale content distribution.", in Proc. of IEEE INFOCOM 2005. Vol. 4. IEEE, 2005
[11] D. Traskov, Lenz, J., Ratnakar, N. and Médard, M., "Asynchronous Network Coded Multicast", in Proc. of ICC Communication Theory Symposium, 2010
[12] X. Zhang, G. Neglia, J. Kurose, "Network Coding in Disruption Tolerant Networks", Network Coding: Fundamentals and Applications Elsevier Science (Ed.) 2011
[13] Zs. Zalatnay, Cs. Simon, M. Maliosz, B. Terza, "Managing streaming services in a distributed testbed", (accepted for publication) MACRO 2015, March 2015
[14] Li, B., Wang, Z., Liu, J., & Zhu, W., "Two decades of internet video streaming: A retrospective view", ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP), 9(1s), 33., 2013
[15] G. Szabo and B. A. Huberman, "Predicting the popularity of online content," Communications of the ACM, vol. 53, no. 8, pp. 80–88, 2010
[16] Jaleel, A., Theobald, K. B., Steely Jr, S. C., & Emer, J., "High performance cache replacement using re-reference interval prediction (RRIP)", In Proc. of ACM SIGARCH Computer Architecture News, Vol. 38, No. 3, pp. 60-71, 2010
[17] Cs. Simon, M. Markosz, B. Baranyai, "Network based caching for near real-time streaming video", (to appear in) Acta Universitatis Sapiantiae – Electrical and Mechanical Engineering, 1/2014.
[18] U. Feige, "A threshold of ln n for approximating set cover", J. Assoc. Comput. Mach., 45:634–652, 1998
[19] C. Lund and M. Yannakakis, "On the hardness of approximating minimization problems", J. Assoc. Comput. Mach., 41(5):960–981, 1994
[20] Har-Peled, S., Lee, M.,"Weighted geometric set cover problems revisited", Journal of Computational Geometry, 3(1), 65-85., 2012
[21] Hei, X., Liang, C., Liang, J., Liu, Y., & Ross, K. W., "A measurement study of a large-scale P2P IPTV system", *IEEE Transactions on Multimedia*, *9*(8), pp. 1672-1687., 2007
[22] Silverston, T., Fourmaux, O., " Measuring P2P IPTV Systems", In *Proceedings of NOSSDAV* (Vol. 7)., 2007
[23] Liu, Z., Wu, C., Li, B., Zhao, S., "UUSee: large-scale operational on-demand streaming with random network coding", IEEE *INFOCOM, 2010* (pp. 1-9). 2010
[24] D.E. Lucani, M. Médard, M. Stojanovic, "Systematic network coding for time-division duplexing", in Proc. IEEE Symposium on Information Theory Proceedings – ISIT, 2010
[25] Karp R. M., "Reducibility Among Combinatorial Problems", in Proc. Sympos. Complexity of Computer Computations, IBM Thomas J. Watson Res. Center, Yorktown Heights, N.Y. New York: Plenum, p.85-103. 1972
[26] Homepage of Lemon library - http://lemon.cs.elte.hu/trac/lemon
[27] GNU Linear Programming Kit homepage - https://www.gnu.org/software/glpk/
[28] GUROBI Optimization Libraries homepage - http://www.gurobi.com/

Csaba Simon received his PhD degree in computer science in the field of infocommunication systems at Budapest University of Technology and Economics (BME), Hungary.
He is an assistant professor in the High-Speed Networks Laboratory at the Department of Telecommunication and Media Informatics, BME. His research interests include QoS of IP networks, IP network architectures, network and service management. He participated in numerous national and international projects in the fields of network resource management, mobility management and smart content delivery.
He is a member of Scientific Association for Infocommunications, Hungary (HTE).

Markosz Maliosz received his MSc (1998) and PhD (2006) degrees in computer science in the field of infocommunication systems at Budapest University of Technology and Economics (BME), Hungary.
He is an assistant professor in the High-Speed Networks Laboratory at the Department of Telecommunication and Media Informatics, BME. His research interests include virtual, cloud and sensor networking along with optimization techniques. He participated in numerous national and international projects in the fields of network resource management, multimedia and smart content delivery.
He is a member of Scientific Association for Infocommunications, Hungary (HTE).

Call for Papers

Prospective authors are invited to submit original research papers for publication in the upcoming issues of our Infocommunications Journal.

Topics of interests include the following areas:

- Data and network security
- Digital broadcasting
- Infocommunication services
- Internet technologies and applications
- Media informatics
- Multimedia systems
- Optical communications
- Society-related issues
- Space communications
- Telecommunication software
- Telecom. economy and regulation
- Testbeds and research infrastructures
- Wireless and mobile communications

Theoretical and experimentation research results achieved within the framework of European ICT projects are particularly welcome. From time to time we publish special issues and feature topics so please follow the announcements. Propo-sals for new special issues and feature topics are welcome.

Our journal is currently published quarterly and the editors try to keep the review and decision process as short as possible to ensure a timely publication of the paper, if accepted. As for manuscript preparation and submission, please follow the guidelines published on our website

http://www.infocommunications.hu/for_our_authors.

Authors are requested to send their manuscripts via electronic mail (preferably) or on a CD by regular mail to the Editor-in-Chief:

Csaba A. Szabo
Dept. of Networked Systems and Services,
Budapest University of Technology and Economics
2 Magyar Tudosok krt., Budapest 1117 Hungary
e-mail: szabo@hit.bme.hu

Call for Proposals of Special Issues

Infocommunications Journal welcomes proposals for Special Issues – collections of papers dedicated to a particular topic of interest for the readers of the journal.

A Special Issue can be based on a recent high quality workshop or conference or papers can be collected from open call. Invited papers can be part of the special issue as well.

A Special Issue can fill in a whole issue, in which case the number of papers is expected to be 8 to 10, or it can be a Mini – Special Issue. In the latter case, at least 3, preferably 4 papers are required.

Proposals for special issues should include:

– contact information
(name, e-mail, title, affiliation and address);
– resume(s) of the proposer(s), with a representative list of recent publications and related experience
(Editorial Board memberships, Guest Editorships, or roles in relevant conferences' program committees);
– the proposed title for the special issue;
– the way the special issue will be compiled
(contributions solicited from a technical event, or to be collected from call for this special issue;
– intent to include invited papers should be also indicated, if possible with the names of professionals who are planned to be invited;
– scope and motivation and description of the special issue;
– guest editors (if different from the proposers) with detailed contact information and resumes.

Proposals should be sent to the Editor-in-Chief:

Csaba A. Szabo
Dept. of Networked Systems and Services,
Budapest University of Technology and Economics
2 Magyar Tudosok krt., Budapest 1117 Hungary
e-mail: szabo@hit.bme.hu

New Key Agreement Techniques for Sensor Networks

Abhishek Parakh and Subhash Kak

***Abstract*—We propose two computationally efficient key agreement algorithms. The schemes are ideally suited for computationally constrained environments such as sensor networks. The first proposed technique is general and uses matrix factorization. We provide constructive algorithms to implement the scheme. The second algorithm uses commutative property of matrices to distribute keys and provides two different keys per node pair. Both the algorithms are practical in terms of implementation, security provided and linear in computational complexity.**

***Index Terms*—Key distribution, sensor networks, matrix factorization**

I. INTRODUCTION

Sensor networks are becoming increasingly popular for applications such as patient health monitoring, detection of border crossings, bridge stress monitoring, signal relay points in battlefields and so on. In many of these applications sensors need to communicate securely to either relay data to base station or perform distributed computations. Therefore, encryption/decryption keys need to be distributed among the sensors.

Key distribution in sensor is particularly challenging because sensors have very limited computational power and transmission ranges. While in recent years the memory capacity for sensors has grown, they still cannot hold large number of keys for pair-wise communication. The key distribution challenge is further complicated by the fact that most sensors are deployed at random. As a result, we do not know a priori which sensors are going to be neighbors of other sensors that is within communication range of each other.

In general, for any key distribution scheme two techniques can be adopted - either install each node with pairwise symmetric keys before deployment or let sensors perform a public key exchange.

Installing pairwise symmetric keys is not a practical solution as it requires large storage capacity and does not allow for dynamic networking where nodes leave and new nodes join. This may happen because old sensors stop working and need to be replaced with new ones or the batteries run out.

If we consider a network to have N nodes, then a pair-wise symmetric key distribution would require each node to store $N-1$ unique keys (because of lack of a priori knowledge of sensor's neighbors). If AES is used as the encryption algorithm, this would require $(N-1)\cdot 128$ bits of storage as it is typical to have 10,000 sensors deployed in a network. If we allow for multiple sensor to use the same key, then we can reduce the number of keys installed on a given sensor, but that also means that once deployed there is a chance a sensor may not share a key with some of its neighbors. Therefore, if a sensor wished to communicate with a neighbor with which it does not share a key (or is out of its communication range), then link encryption (hop-by-hop) is used. In link encryption, assume sensor a wants to communicate with sensor d with which it does not share a key (or d is out of its communication range). If a shares a key with node b which in turn shares a key with node d, then a can send b a message such as $E_{k_{ab}}(m)$; where k_{ab} is a key shared between a and b. Node b upon receiving this message, first decrypts it and then re-encrypts it with key k_{bd} that it shares with node d and send it to d. This latter approach requires multiple encryption/decryptions along the way as well as a path finding and routing algorithm.

Eschenauer and Gligor [1] introduced the above approach where they assumed limited memory capacity and limited communication range for sensor networks. Further, they assumed random deployment of sensors, i.e. a sensor's neighbors were not known before deployment. As a result, after deployment the sensors performed a neighbor discovery in which they determined who their neighbors are and with which one of them they share keys. Then the sensors performed a path discovery to those sensors with which they do not share keys. Once a path was discovered, messages were sent using link-encryption. Although, the scheme proposed in [1] is very general and applicable to most scenarios, in practise one does have some knowledge of sensor neighborhood before deployment. Hence, EG requires the storage of larger number of keys on each sensor than may be required in a given scenario. Further, the path finding and routing protocols in a distributed sensor network are not trivial, especially when the number of neighbors one shares keys with are only a fraction of the number of neighbors actually in communication range.

Du et al. [2] assume deployment knowledge to reduce the number of keys stored per node. A gaussian probability distribution function is assumed with every sensor having a high probability of being deployed at a specific coordinate in a grid. However, such a scheme is not applicable to mobile nodes. Chan et al. [3] proposed a q-composite scheme that is similar to the EG scheme but requires that the nodes share q keys from the key ring instead of just one key and then final key to be used for encryption is computed as a function of these q shared keys.

In [4] it is assumed that mobile sensors handle the load of key distribution while static sensors only require minimal resources for key management. A bootstrapping technique is

Manuscript received December 10, 2014, revised March 10, 2015

A. Parakh is with the Nebraska University Center for Information Assurance, University of Nebraska, Omaha, NE 68182, e-mail: aparakh@unomaha.edu

S. Kak is with the School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK 74078

proposed in [5] that enables sensors to compute keys, once the network is deployed, based on network density, nodes memory and transmission range. A new post-deployment pairwise key distribution scheme for two-tier sensor networks is considered in [6] and a polynomial based key generating model is used for generating pair-wise keys to be shared with neighboring nodes.

The second approach, for key distribution, is that of using public key algorithms. In such algorithms, each sensor is installed with a public key and a corresponding private key. After deployment, the sensors broadcast their public keys to the neighboring sensors. The sending node can then encrypt all communication with the receiving node's public key. A number of public key algorithms have been implemented on sensors that claim to provide practical solutions, however, they consume many times more power than secret key encryption algorithms [7]. A hybrid scheme where public keys are used to exchange secret keys and the data encryption and transfer takes place using secret key algorithms is probably of a greater practical use as it reduces power consumption. One such hybrid approach is explored in [8] where the entire sensor network is divided into clusters managed by cluster heads. These cluster heads implement public key cryptography and aggregate data while individual sensors only use symmetric keys for encryption. A central key management server is used to establish keys in [9]. A hash chain based key distribution mechanism is discussed in [10].

Blom [11] discussed key exchange techniques based on the use of (n, k) linear codes with threshold property. A slightly modified version of Blom's algorithm was used by Du et al. [12] for establishing multiple shared keys between nodes by essentially executing Blom's scheme multiple times. Blundo et. al proposed a key distribution scheme [13] based on bivariate symmetric polynomial. Another scheme using LU Composition integrated with Elliptic Curve Diffie-Hellman has been proposed in [14]. Techniques based on polynomial interpolation and the idea of secret sharing are discussed in [15], [16] but have slightly higher computational cost compared to the proposed scheme. Similar polynomial based scheme for a two-tier network is proposed in [17].

In this paper we discuss an approach the bridges the gap between secret key and public key algorithms and enables sensors to establish shared secret keys with each other after deployment. In this approach each sensor is pre-installed with a small amount of seed information that can be exchanged with a neighbor to agree on a secret key. Any secret key encryption algorithm can be used to encrypt data thereon. Therefore, the proposed approach provides a number of advantages:

1) Pre-deployment of encryption keys is not required (key are generated after deployment).
2) Key agreement only has linear complexity.
3) A given node can share keys with all its neighbors resulting in larger connectivity within the network. This in turn leads to shorter path lengths compared to other methods where nodes share keys with only a fraction of its neighbors.
4) It allows for dynamic networks with nodes leaving and joining.
5) It assumes no pre-deployment knowledge of node locations and hence is general.

The proposed algorithm is based on matrix operations, where the computationally expensive pre-processing is pushed to pre-deployment phase and can be done at a base station.

In the following section, we present the proposed algorithm. In subsections II-A and II-B we discuss the size of matrices used and the complexity of the proposed algorithm. Section III discusses the security of the proposed algorithm for different size of matrices used and subsection III-C discusses the resilience against node capture. Section IV presents some constructive algorithms for the proposed scheme. Section V presents the second algorithm with commuting matrices and section VI concludes the paper.

II. PROPOSED ALGORITHM

Our aim is to provide alternatives to the use of public key algorithms for the establishment of shared secret keys. To achieve this we store a small amount of information, pre-deployment, on all the sensors. Once deployed, the sensors exchange a part of the pre-installed information with their neighbors to generate a shared secret key. This is very similar to what happens in Diffie-Hellman key exchange. However, here we do not require any exponentiation operation and the security of the scheme does not rely on the difficulty of mathematical operations (for example the security of Diffie-Hellman depends on the difficult of finding logarithms in finite fields). Since, the generation of the session key takes place after deployment, any sensor can perform a key exchange with any other sensor within its communication range. Larger connectivity between neighbors provides shorter path lengths through the network.

In the proposed algorithm all computations are performed modulo a large prime p. The proposed method is based on matrix factorization. It is further assumed that there are N sensors in the field and the deployment is done at random. The proposed algorithm consists of two phases - the pre-deployment phase and the key agreement phase.

The pre-deployment phase is performed by a base station. This phase essentially involves the factorization of a symmetric matrix. The symmetric matrix consists of random numbers from a finite field. These random numbers are the actual keys that will be used, therefore an appropriately large finite field must be used (usually on the order of 128 bits - if AES is being used). The base station performs the pre-deployment computations as follows.

Pre-deployment Phase (at base station):

1) Randomly choose a symmetric matrix K with elements in Z_p.
2) Find two matrices X and Y such that $XY = K$.
3) Randomly assign r^{th} row of X and r^{th} column of Y to each sensor node.

While distributing rows and columns, if node i receives the r^{th} row of matrix X, it also receives the r^{th} column of matrix Y. Here r is an integer chosen at random with uniform

probability from $[1, q]$, where matrix K is a symmetric matrix of size $q \times q$.

After every sensor is installed with a row-column pair from X and Y, the sensors can be deployed in the field using any mode of deployment. Once deployed, each sensor probes its neighborhood to discover the neighbors and then agree on a symmetric key as follows.

Key Agreement Phase:

When any two nodes, i and j, wish to agree on an encryption key, they exchange their columns of Y (in plaintext) and compute a common key as follows,

$$\text{Node } i \text{ computes: } K_{ij} = row_i(X) \cdot col_j(Y)$$

$$\text{and node } j \text{ computes: } K_{ji} = row_j(X) \cdot col_i(Y)$$

As matrix K is symmetric, $K_{ij} = K_{ji}$. Since a node in the network has only one row and one column installed on it, the notation $row_i(X)$ denotes the row of X that was stored on node i. This must not be confused with row i of X. Similarly, $col_j(Y)$ is the column of Y assigned to node j.

Fig. 1 show the pictorial representation of matrices X and Y and fig. 2 illustrates a sensor network in which all the nodes within the communication range with each other can share an encryption key. Only a few nodes with their communication ranges are shown.

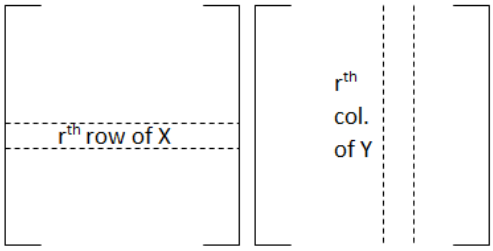


Fig. 1. A node gets the r^{th} row and column of matrices X and Y.

A. Size of Matrices

Assume a network with N nodes. If we wanted every node-pair in the network to share a randomly and uniformly chosen key, then there would exist $\frac{N(N-1)}{2}$ independent keys. In other words, if an eavesdropper is able to determine the key being used for the link between nodes i and j, then he gains no advantage in determining the key being used on any other link. For this to be true, the symmetric key matrix K needs to be of size $N \times N$ since the upper (or lower) triangle of the matrix contains $\frac{N(N-1)}{2}$ elements (not including the diagonal elements that only form "self-keys").

A $N \times N$ key matrix can be factored into X and Y in different ways where the size of matrix X is $N \times m$ and the size of matrix Y is $m \times N$. As a result, every node in the network can receive a unique row-column pair during the pre-deployment phase. A simple row-column pair distribution algorithm would give node i, the i^{th} row of X and i^{th} row of Y. As a result, the storage required on each node is $2m$ integers.

In general if reuse of keys is allowed, matrix K may be of size $q \times q$, where $q \leq N$, and matrices X and Y are of sizes $q \times m$ and $m \times q$, respectively. In this case, step 3 of the algorithm randomly assigns rows and columns, where a row-column pair may go to more than one sensor node. This implies that some of the node pairs may share the same encryption key. More precisely, a $q \times q$ matrix has $\frac{q(q-1)}{2}$ random and independent numbers. Therefore, it is expected that any given key will be shared by $\frac{N(N-1)}{q(q-1)}$ links and each row-column pair may go to $\frac{N}{q}$ nodes.

B. Linear Computational Complexity of Key Generation

The key generation operation for a given link involves the multiplication of one row of X with a column of Y. If we assume the size of X is $q \times m$ and the size of Y is $m \times q$, then computing a key requires m multiplications and $m - 1$ additions. Further this is dependent on the size of matrices X and Y which in turn depends on the desired level of security. In the worst case X and Y are of size $N \times N$ and hence key generation requires N multiplications and $N - 1$ additions.

III. Security of the Proposed Scheme

A. Size of K is $N \times N$

Assume that matrix K is of size $N \times N$ and therefore X and Y are of sizes $N \times m$ and $m \times N$ respectively. The upper triangle (including the diagonal) of matrix K has $\frac{N(N+1)}{2}$ elements all of which are generated from $2(N \times m)$ elements. For each key there are p possibilities and it is clear that in the absence of any knowledge of the elements of X and Y, all the p possibilities for every key remain equally likely.

However, since the columns of Y are being transmitted in plain text, an eavesdropper can record these columns. Further, if the eavesdropper is able to listen to N distinct transmissions, of columns of Y, from N distinct nodes, then there remain $N!$ possibilities to arrange these columns in matrix Y. However, this gives no information about matrix X which has $N \times m$ elements in it.

Now, if the adversary tries to guess the values of elements in matrix X, then every new row of X gives the adversary decreasing amount of information. This is because the key matrix K is symmetric. As a result, the first row of X when multiplied with Y gives $N - 1$ unique keys, the second row of X when multiplied with Y gives $N - 2$ unique keys and so on. However, to determine all the keys this way, the adversary will have to determine all the rows of X, i.e. $N \times m$ values from Z_p.

Two possible choices remain for the eavesdropper:

- If $m > N$ then it is more difficult to determine X than it is to determine the elements of K directly. Therefore, capturing N columns of Y gives no advantage.
- However if $m < N$, then determining the first row of X (m values) will result in $N - 1$ keys, determining the second row X (another m values) will result in $N - 2$ keys and so on. Consequently, an adversary can stop

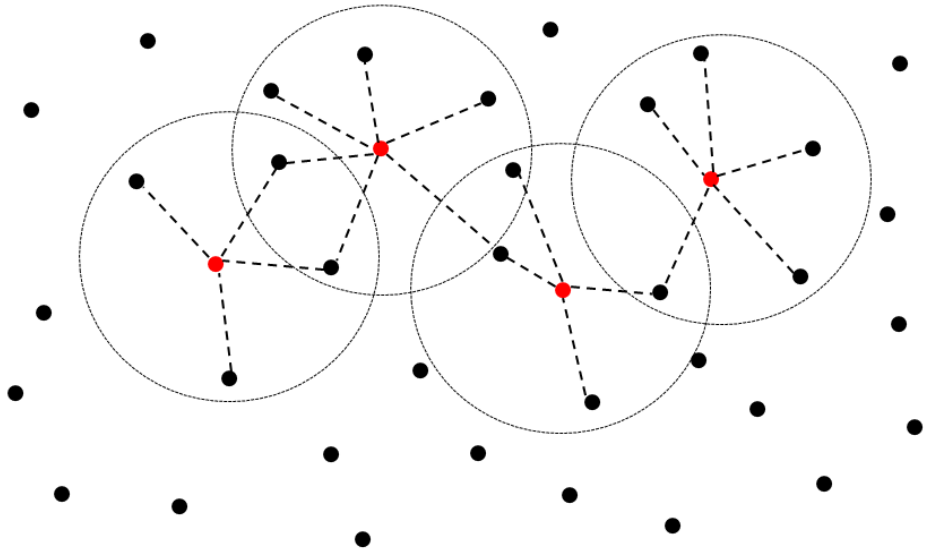


Fig. 2. Illustration of a network. We've shown the communication ranges with dotted circles for a few nodes (shown in red at the center of the circles). The dashed lines between nodes represents that a key agreement will take place between these nodes since they are within communication range of the red nodes. Similarly every node will have its own communication range and it will agree on a key with all nodes in its communication range. Nodes can join and leave the network at will.

determining the rows of matrix X when determining m values of that row gives fewer keys than m. At this point it would be beneficial for the adversary to directly guess the remaining values in the key matrix K.

This is in addition, however, to determining which of the $N!$ ways are the columns of matrix Y arranged and for each arrangement different possible keys exist.

Moreover, such an attack may be impractical because it requires an adversary to listen to every transmission of Y that takes place in the *entire* network.

B. Size of K is $q \times q$, $q < N$

Assume that matrix K is of size $q \times q$, $q < N$, then the matrices X and Y are of size $q \times m$ and $m \times q$. Since, there are fewer than N rows and columns, the rows-columns pairs pre-loaded on the sensors can be pre-loaded randomly with repetition. As a result, some of the links in the network will share the same encryption key.

The probability that any two nodes will share the same key may be computed as follows. Note there are two possible events that can occur.

1) Two node pairs are assigned the same K_{ij} from the key matrix.
   This may happen because $q < N$ and same keys are used more than once. Since there are a total of $\frac{q(q-1)}{2}$ possible keys, the probability that two node pairs will be assigned the same key is: $\frac{1}{\frac{q(q-1)}{2}} = \frac{2}{q(q-1)}$.
2) Two node pairs are assigned different K_{ij}s from the key matrix.
   This happens with a probability of $1 - \frac{2}{q(q-1)}$. However, since each K_{ij} is randomly and uniformly picked from Z_p, any two elements of matrix K will be equal with probability $\frac{1}{p}$.

Therefore, the total probability that any two node pairs will receive the same encryption key is given by:

$$(1 - \frac{2}{q(q-1)}) \cdot \frac{1}{p} + \frac{2}{q(q-1)} \cdot 1$$

where $q > 1$. When $q = 1$ the probability of repetition of the same key is 1 and refers to a master key system.

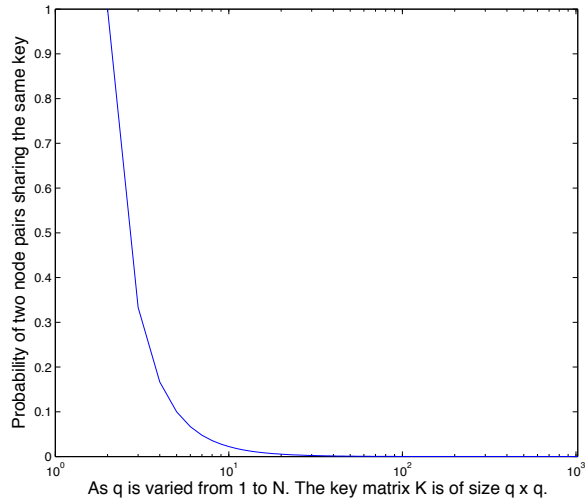


Fig. 3. Probability of two node pairs sharing the same key as q is varied from 2 to N. Size of the key matrix K is $q \times q$.

Figure 3 shows how the probability of sharing the same key between two node pairs decreases as size of the key matrix K is increased. We see that at size of $q = 200$ the probability of

sharing a key decreases to less than 0.0001. The network size is fixed at $N = 1024$.

C. Resilience Against Node Compromise

Compromise of nodes leads to greater loss of information than eavesdropping. However, unlike Blom's scheme [11], the proposed algorithm does not possess a threshold property. Hence, the degradation of network characteristics is graceful with compromise of nodes. Further, depending on how the factors of K are computed the scheme can be adapted to possess a threshold property if desired.

If the size of K is $N \times N$ and each node receives a unique row-column pair then if an attacker was to compromise m nodes, he will be able to construct: $(N-1)+(N-2)+\ldots+(N-m) = m \cdot N - \frac{m(m+1)}{2}$ keys (elements) from matrix K.

If the size of K is $q \times q$, $q < N$ then the probability of construction of all the keys upon L node compromises will be based on the retrieved L row-column pairs. However, since $q < N$ not all of these row-column pairs are distinct. The minimum number of nodes that an adversary will need to compromise in order to retrieve q distinct row-column pairs is q nodes. Consequently, for L nodes compromised the probability that the adversary will see q distinct row-column pairs at least once is computed as follows:

Suppose in L compromises the adversary does not see all the q distinct row-column pairs, i.e. at least one pair is missing. In other words, we can compute all the ways in which L choices (with repetition) can be made from $q-1$ possible pairs. Out of total q pairs there is $\binom{q}{(q-1)}$ ways to choose $q-1$ pairs. From each of these possibilities, we can make L random choices with repetition in $(q-1)^L \cdot \binom{q}{(q-1)}$ ways. However, we need to subtract the double counted possibilities which is $q^2 - 2q$. As a result, the total number of ways, an adversary will fail to see all possible row-column pairs in L compromises is given by $(q-1)^L \cdot \binom{q}{(q-1)} - (q^2 - 2q)$. This is out of the total number of possible ways all choices can be made, i.e. q^L. The probability that an adversary will successfully see all the pairs in L compromises is then given by,

$$1 - \frac{(q-1)^L \cdot \binom{q}{(q-1)} - (q^2 - 2q)}{q^L}$$

IV. Constructive Algorithms to Determine X and Y

Although in general X and Y may be chosen by trial and error, their determination becomes easier if one of the following methods is used. The examples below present some of the different methods to construct X and Y.

1. One method would be to choose Y random and non-singular and compute $X = K \cdot Y^{-1}$.

For example, assume a symmetric matrix K of size 3×3 and we work mod 11.

Let $K = \begin{bmatrix} 3 & 4 & 6 \\ 4 & 5 & 2 \\ 6 & 2 & 1 \end{bmatrix}$ and $Y = \begin{bmatrix} 3 & 4 & 8 \\ 1 & 5 & 2 \\ 9 & 10 & 4 \end{bmatrix}$

Then $Y^{-1} = \begin{bmatrix} 0 & 6 & 8 \\ 2 & 4 & 5 \\ 6 & 4 & 0 \end{bmatrix}$ and

$X = \begin{bmatrix} 0 & 3 & 0 \\ 0 & 8 & 2 \\ 10 & 4 & 3 \end{bmatrix}$

2. Another example of construction of X and Y where they are smaller than the size of K is as follows (again working mod 11).

Let $X = \begin{bmatrix} 1 & 3 & 1 \\ 3 & 4 & 1 \\ 9 & 9 & 1 \\ 4 & 6 & 1 \\ 5 & 1 & 1 \end{bmatrix}$ and

$Y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 5 & 3 & 2 & 4 \\ 1 & 3 & 9 & 4 & 5 \end{bmatrix}$ then

$K = \begin{bmatrix} 5 & 8 & 8 & 0 & 7 \\ 8 & 4 & 2 & 4 & 2 \\ 8 & 2 & 1 & 9 & 6 \\ 0 & 4 & 9 & 9 & 0 \\ 7 & 2 & 6 & 0 & 3 \end{bmatrix}$

Here every node stores $2m = 2 \cdot 3 = 6$ elements from Z_p.

3. LU factorization may be used.

4. Computing powers of matrices. Assume that matrix K is diagonalizable; then $K = M^{-1}AM$ where M is a matrix whose columns are the eigenvectors of K and A is a diagonal matrix of eigenvalues of K. With such a factorization the algebra on K reduces to the algebra on the elements of the diagonal matrix A. For example $K^r = M^{-1}A^rM$ and since A is a diagonal matrix $A^r = (a_1^r, a_2^r, \ldots, a_q^r)$ where a_i are the diagonal elements of A. Then we may factor K as follows:

1) Randomly choose a symmetric diagonalizable matrix K with elements from the finite field Z_p.
2) Randomly choose a element r from the field and compute $X = K^{\frac{1}{r}}$ and $Y = K^{1-\frac{1}{r}}$.

If K is diagonalizable then the r^{th} root can be computed as discussed above.

In this method, if an adversary captures all the columns of Y and figures out the layout of the captured columns of Y in the matrix, then to compute X, he will have to compute the $(r-1)^{th}$ root of the matrix Y. However, not knowing the value of r which was randomly and uniformly chosen from Z_p there are $p \cdot (r-1)$ possible choices for X.

V. Using Commuting Matrices

If one was to use commuting matrices the requirement of matrix K being symmetric can be eliminated. This would require every node to store some additional information that provides every node pair two different keys that are used for communication depending on which node initiates the communication. These keys may be hashed together to generate another key that is used for encryption thus further improving the security of the system. The algorithm works as follows:

Pre-deployment

1) Choose two $q \times q$ matrices X and Y such that $XY = YX$ and Y is symmetric.
2) Randomly pick r from a uniform distribution over $[1, q]$.
3) Assign node i the r^{th} row and column of X and the r^{th} column of Y.

Assume two nodes i and j wish to agree on a key then the key agreement proceeds as follows,

1) Node i sends its i^{th} column of Y to node j.
2) Node j sends its j^{th} column of Y to node i.
3) Node i computes $K_{ij} = row_i(X) * col_j(Y)$ and node j computes $K_{ij} = col'_i(Y) * col_j(X)$.
4) Node i computes $K_{ji} = col'_j(Y) * col_i(X)$ and node j computes $K_{ji} = row_j(X) * col_i(Y)$.

Where $col'_i(Y)$ is the column of Y transposed. Figure 4 illustrates the predistribution of rows and columns for node i.

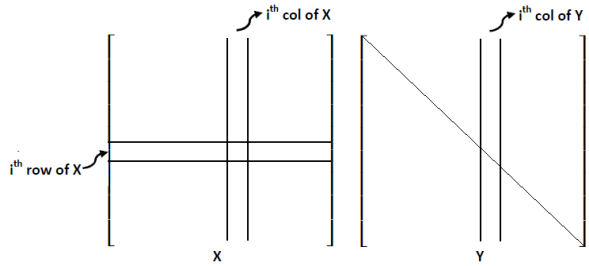


Fig. 4. Predistribution of rows and columns for node i for the case when $XY = YX$ and Y is symmetric. The line across the diagonal of matrix Y indicates the symmetry.

Finding Commuting Matrices X and Y:

Matrix diagonalization may be used to find two such matrices using the following steps:

1) Choose a diagonalizable symmetric matrix Y at random.
2) Diagonalize Y such that $Y = M^{-1}D_yM$, where D_y is a diagonal matrix with eigenvalues of Y.
3) Randomly pick a diagonal matrix D_x and compute $X = M^{-1}D_xM$.

The above algorithm generates two matrices that commute with each other as seen below,

$$XY = M^{-1}D_xMM^{-1}D_yM = M^{-1}D_xD_yM$$

and

$$YX = M^{-1}D_yMM^{-1}D_xM = M^{-1}D_yD_xM$$

Since D_x and D_y are diagonal matrices $D_xD_y = D_yD_x$. If an eavesdropper is able to listen to q distinct columns of Y being transmitted and also determine which out of $q!$ ways they are to be arranged then he can diagonalize Y and retrieve M. In order to reconstruct all the keys in the network he then has to guess the values in the diagonal matrix D_x and there are p possible values for every eigenvalue in D_x.

Unlike the previous algorithm, using commuting matrices requires X and Y to be square matrices. However, the size of the matrices may be $q \times q$ where $q \leq N$. The advantage of using commuting matrices is that every node pair now shares two keys that may be used for encryption in following ways:

1) For fast encryption and low computational overhead, encryption keys are used to seed random number generators. Then to encrypt data the sequence of random bytes generated (from the RNG) is XORed with the bytes of the data and transmitted. However, if the same seed is used in the RNG for data sent both ways, i.e. from node i to j and vice versa, then it can lead to a two-time-pad attack. As a result, the two encryption keys generated: K_{ij} can be used to seed the RNG for encrypting data sent from node i to j and K_{ji} can be used for data sent from j to i.
2) If two-time-pad is of no concern then we could $K = Hash(K_{ij}||K_{ji})$ as the common encryption key between nodes i and j; where the keys are hashed in a predecided order. This also provides an additional layer of indirection, using hash functions, for an attacker doing cryptanalysis on captured encrypted data.
3) Or the two keys could simply be concatenated $K_{ij}||K_{ji}$ to increase the key length.

Using commuting matrices increases network resilience as the attacker would need to determine the entire K matrix rather than only half of it as is the case when K was symmetric. Recall that by using commutativity we have eliminated the requirement of K being symmetric. This is in contrast with other methods that either implement Blom's scheme multiple times to increase the number of keys shared (example to share two keys, implement Blom's scheme twice) [12] or Chan et al.'s scheme [3], called q-composite scheme, that shares at least q keys between nodes to increase network resilience by decreasing the key pool size.

The computational complexity, when using commutative matrices, remains linear as it requires multiplication of a row and a column for each key.

VI. CONCLUSIONS

We have proposed two algorithms for key agreement between nodes in a network. The first algorithm factors a symmetric matrix K into two factors X and Y and the second algorithm chooses matrix X and Y such that they commute and Y is symmetric. In the latter method, matrix K need not be symmetric. We have provided several constructive algorithms to choose X and Y in order to decrease the computational load during pre-deployment phase. Key generation only requires the multiplication of a row and a column of a matrix and is linear in complexity. The commutative method provides two keys per node pair that can be used in different ways depending on security requirements and encryption algorithm used.

REFERENCES

[1] L. Eschenauer and V. D. Gligor, "A key-management scheme for distributed sensor networks," in *Proceedings of the CCS'02*. New York, NY, USA: ACM, 2002, pp. 41–47.
[2] W. Du, J. Deng, Y. Han, S. Chen, and P. Varshney, "A key management scheme for wireless sensor networks using deployment knowledge," in *INFOCOM 2004*, vol. 1, March 2004.
[3] H. Chan, A. Perrig, and D. Song, "Random key predistribution schemes for sensor networks," in *Proceedings of 2003 Symposium on Security and Privacy*, may 2003, pp. 197–213.

[4] B. Tas and A. Tosun, "Mobile assisted key distribution in wireless sensor networks," in *2011 IEEE International Conference on Communications (ICC)*, June 2011, pp. 1–6.
[5] F. Liu, X. Cheng, L. Ma, and K. Xing, "SBK: A self-configuring framework for bootstrapping keys in sensor networks," *IEEE Transactions on Mobile Computing*, vol. 7, no. 7, pp. 858–868, July 2008.
[6] K. Gagneja, "Pairwise key distribution scheme for two-tier sensor networks," in *2014 International Conference on Computing, Networking and Communications (ICNC)*, Feb 2014, pp. 1081–1085.
[7] M. A. Simplício, Jr., P. S. L. M. Barreto, C. B. Margi, and T. C. M. B. Carvalho, "A survey on key management mechanisms for distributed wireless sensor networks," *Computer Networks*, vol. 54, no. 15, pp. 2591–2612, Oct. 2010.
[8] S. Ruj and K. Sakurai, "Secure and privacy preserving hierarchical wireless sensor networks using hybrid key management technique," in *2013 IEEE Global Communications Conference (GLOBECOM)*, Dec 2013, pp. 402–407.
[9] S. Sanyal, "Spike: A novel session key management protocol with time-varying secure cluster formation in wireless sensor networks," in *2013 Eleventh Annual International Conference on Privacy, Security and Trust (PST)*, July 2013, pp. 151–160.
[10] W. Bechkit, Y. Challal, and A. Bouabdallah, "A new class of hash-chain based key pre-distribution schemes for WSN," *Computer Communications*, vol. 36, no. 3, pp. 243–255, Feb. 2013.
[11] R. Blom, "An optimal class of symmetric key generation systems," in *Proceedings of the EUROCRYPT 84 Workshop on Advances in Cryptology: Theory and Application of Cryptographic Techniques*. Springer-Verlag New York, Inc., 1985, pp. 335–338.
[12] W. Du, J. Deng, Y. Han, P. Varshney, J. Katz, and A. Khalili, "A pairwise key predistribution scheme for wireless sensor networks," *ACM Transactions on Information and System Security*, vol. 8, pp. 228–258, May 2005.
[13] C. Blundo, A. D. Santis, A. Herzberg, S. Kutten, U. Vaccaro, and M. Yung, "Perfectly-secure key distribution for dynamic conferences," in *Proceedings of the 12th Annual International Cryptology Conference on Advances in Cryptology*, ser. CRYPTO '92. London, UK: Springer-Verlag, 1993, pp. 471–486.
[14] E. K. Wang, L. C. Hui, and S.M.Yiu, "A new key establishment scheme for wireless sensor networks," *International Journal of Network Security & Its Applications*, vol. 1, no. 2, pp. 17–27, July 2009.
[15] W. Chunying, L. Shundong, and Z. Yiying, "Key management scheme based on secret sharing for wireless sensor network," in *Emerging Intelligent Data and Web Technologies (EIDWT), 2013 Fourth International Conference on*, Sept 2013, pp. 574–578.
[16] M. Bertier, A. Mostefaoui, and G. Tredan, "Low-cost secret-sharing in sensor networks," in *2010 IEEE 12th International Symposium on High-Assurance Systems Engineering (HASE)*, Nov 2010, pp. 1–9.
[17] K. Gagneja, "Pairwise key distribution scheme for two-tier sensor networks," in *2014 International Conference on Computing, Networking and Communications (ICNC)*, Feb 2014, pp. 1081–1085.

Abhishek Parakh is an Assistant Professor of Information Assurance at the College of Information Science and Technology at University of Nebraska, Omaha. He received his Ph.D. in Computer Science from Oklahoma State University. He is also a member of Nebraska University Center for Information Assurance, a National Security Agency (NSA) designated Center for Academic Excellence in Information Assurance Education - Cyber Defense (CAE-IA/CD). His research interests include cryptographic engineering, distributed systems security, resource constrained encryption systems, protocol development, quantum cryptography and SCADA systems security. He has over 35 publications and has been funded by NSF and NASA.

Subhash Kak is Regents Professor in the School of Electrical and Computer Engineering at Oklahoma State University in Stillwater. Prior to joining Oklahoma State University, he served for many years as the Delaune Distinguished Professor of Electrical and Computer Engineering at Louisiana State University in Baton Rouge.

He is the author of several books that include The Nature of Physical Reality (New York, Peter Lang, 1986) and The Architecture of Knowledge (New Delhi, CRC, 2004). His areas of interest include data security, quantum computing, information theory, neural networks, and history of science. His technical research is in the fields of data security and cryptography, information theory, neural networks, and quantum information and he has also written on archaeoastronomy and art. Amongst his awards are British Council Fellow (1976), Science Academy Medal of the Indian National Science Academy (1977), Kothari Prize (1977), UNESCO Tokten Award (1986), Goyal Prize (1998), National Fellow of the Indian Institute of Advanced Study (2001), and Distinguished Alumnus of IIT Delhi (2002).

Investigation of Semiconductor Optical Amplifier Direct Modulation Bandwidth

Eszter Udvary, *Member, IEEE*

***Abstract*—In-line and reflective semiconductor optical amplifiers can be advantageously utilized as external modulators because they can perform simultaneously two functions (amplification and modulation). When the SOA is used in next generation access networks the modulation bandwidth is a crucial property, because it will limit the application. In the paper theoretical and experimental investigations are presented with the result of improved modulation bandwidth. The experimental results confirm that by proper adjustment of the operation point and environmental conditions a significant improvement in the modulation bandwidth is achieved. Mainly the bias current and the level of the incident optical power determine the carrier lifetime. The modulation bandwidth can be doubled by this approach. The theoretical and simulation results represent the effect of the device parameters, like device length and electrode setup. However there is a trade-off between modulation efficiency and modulation bandwidth, which demands circumspect design of the device and system parameters.**

***Index Terms*—Semiconductor Optical Amplifier, intensity modulation, optical modulation, access network, optical communication**

I. INTRODUCTION

Reflective and in-line semiconductor optical amplifiers (RSOAs and SOAs) are looks like key components for next-generation access networks (NGANs), since they allow for directly-modulated colorless transceivers. Next generation access networks will probably gain great advantages in exploiting the potentialities of wavelength division multiplexing (WDM). It is true for both WDM based 60 GHz Radio over Fibre systems (60G-RoF) [1] and WDM Passive Optical Networks (WDM-PONs) [2]. The unit at the user side should be colorless in these systems. It can be obtained by the injection technique performed on Fabry-Perot lasers [1] and reflective semiconductor optical amplifiers (RSOA) [3] or even in Self-seeded RSOA [4]. Optical amplifiers are preferred devices for many solutions. RSOAs and SOAs have already demonstrated their multifunctional capability by combining optical amplification with either modulation, gating, photodetection, dispersion compensation, linearization, commutation, wavelength conversion, signal regeneration, etc. [5, 6]. The SOA offers a semiconductor based, small size, integrable, low cost optical intensity modulator. It requires relative low bias current and modulation signal. The detected electrical power is high because of the optical gain in contrary to the optical insertion loss of other modulators. Additionally it has better linearity than Mach-Zehnder modulator [7]. Contrary, the amplifier adds optical noise to the signal and the operation of SOA based modulator depends on system and device parameters [8].

In the RSOA-based WDM PON, the upstream and downstream signals are propagating within the same fiber due to the single-fiber loopback configuration. On the other hand, the signal distribution of RoF systems can be realized by point-to-point, point-to-multipoint, bus, ring and open loop topologies, where in-line device structure is more powerful [5].

However modulation bandwidth of semiconductor optical amplifier is usually limited to around 1-3 GHz [14]. Increasing the modulation bandwidth is still a challenge. In this paper, the modulation response of a semiconductor laser amplifier will be analyzed, i.e., the frequency dependence of the amplitude modulation imposed on an injected CW optical beam when the bias current is modulated. The paper is organized as follows. Section II overviews the bandwidth enhancement methods. Section III discusses the optimal environmental parameters; the expected results are validated by experimental work. Section IV represents the modulation improvement applying optimal device structure and finally Section V concludes the paper.

II. MODULATION BANDWIDTH ENHANCEMENT METHODS

Various techniques have already been demonstrated to overcome band limitation of standard RSOA. The frequency response of the RSOA has a smooth roll-off with no relaxation oscillation peak, while its modulation has a good linearity similar to the laser diode [8]. It is perfect for the electronic equalization using the decision feedback equalizer, which can be combined with Forward Error Correction [10]. Set-off optical filtering aided by electronic equalization [11] or detuned ad-hoc delay interferometers [12] are also good perspectives.

On the other hand, introduction of advanced modulation formats with high spectral efficiency in optical communication systems led to significant improvements. Advanced modulation formats have been demonstrated both to effectively increase the transmission capacity of bandwidth-limited transmitters and to implement efficient multiplexing techniques,

Manuscript received 11 November 2014, revised 03 March 2015.

This work was supported by the Hungarian Fund under OTKA No. PD 109288.

E. Udvary is with the Department of Broadband Infocommunications and Electromagnetic Theory, Budapest University of Technology and Economics, Budapest, 1111, Egry 18., Hungary (e-mail: udvary@ mht.bme.hu).

such Orthogonal Frequency Division Multiplexing (OFDM) [4]. However, the aforementioned approaches improve the complexity of the system.

On the other hand, the problem of the limited modulation bandwidth can be overcome with system and device level. In principle, the modulation bandwidth is limited by the speed at which the carrier density can be changed. This is usually determined by the lifetime of the carriers in the RSOA active layer. Carrier lifetime is mainly governed by emission rate. Based on it, the operational and environmental parameters can be optimized from the viewpoint of the modulation speed. For example the length of the RSOA can be enlarged to increase photon density, hence reducing carrier lifetime. However the transmission will be degraded due to the chirp over long distances. So, the optimization of device and environmental parameters is effective method, but it demands circumspect design of the device and system parameters.

Naturally, SOA/RSOA's direct modulation capability is characterized by several parameters, like chirp, extinction ratio, ASE noise, Noise Figure, optical signal-to-noise ratio, Q-factor, linearity, etc. Each of these modulator characteristics are influenced by same device and system parameters. However the main problem is the bandwidth limitation. So the paper focuses the investigation of the modulation bandwidth.

III. Experimental Work

The optimal environmental parameters and operating point of the device must be selected cautiously. It is difficult to give individual optimization and a trade-off is necessary, because different parameters are optimal for amplifier gain, modulation bandwidth, linearity, etc. The modulation depends on system and device parameters. So the presented investigation involves the theoretical and experimental study of transfer function and modulation bandwidth with different environmental constraints.

A. Measurement set-up

The frequency response of the RSOA can be measured by modulating the carrier density. Fig. 1 shows the simplified measurement setup. During the experimental work the RSOA-modulator under test was driven by the sum of bias (dc) current and sinusoidal modulation radio frequency (RF) signal via a bias tee. The RSOA under test was packaged in butterfly

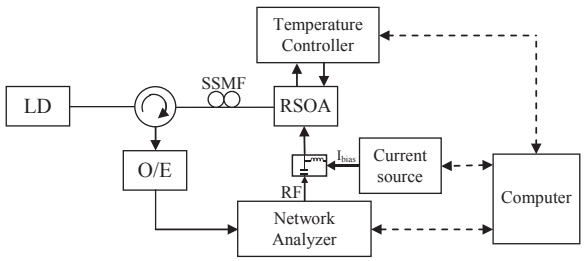


Fig. 1.Measurement setup

package and the impedance mismatching was moderate. Frequency of the RF signal is varied between 100 MHz and 10

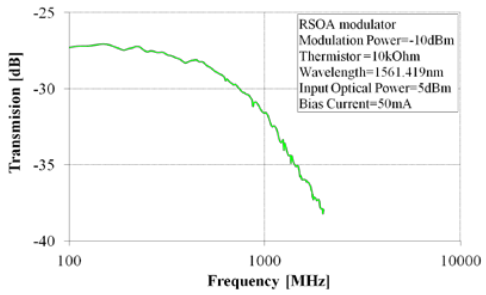


Fig. 2. Typical measured frequency response

GHz and it is generated by network analyzer. The required optical power and wavelength were produced by a tunable laser source. The incoming continuous wave and the reflected, modulated optical signals were separated by optical circulator.

The intensity modulated optical signal was detected by an amplified PIN photodetector. At the electrical output a network analyzer registered the detected electrical signal. The setup was controlled by a computer program, hence the measurement parameters were carefully set by the program and the measurement results were processed and stored.

Fig. 2 describes the typical frequency response of the RSOA. The bias current was 50 mA and the injection power was 5dBm, which is 7 dB higher than the 3 dB gain compression point. The 10 dB modulation bandwidth was measured to be 1.8 GHz. The slope of the curve is -20 dB/decade.

B. Modulation bandwidth versus bias current of the RSOA

The amplitude and shape of the transfer function depend on several parameters. The input optical power and the bias current of the RSOA-modulator are the two most important effects. Fig. 3 depicts the relative modulation bandwidth versus bias current as a function of the input optical power. The improvement of the bandwidth is about linear proportional to the bias current. The slope of the curve is higher in case of higher input power. So the bias current effect is more significant in the saturated regime.

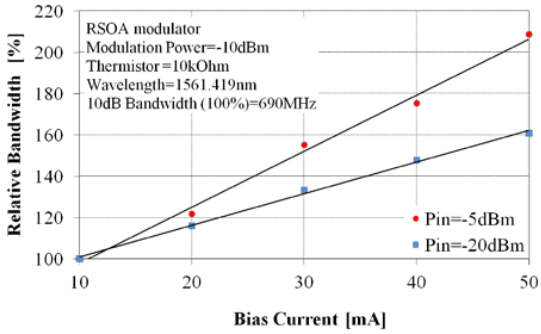


Fig. 3. Measured bandwidth improvement versus bias current in saturated and unsaturated regimes

C. Modulation bandwidth versus incident optical power

In general, the lifetime of the carriers in the presence of a strong, saturating input signal is reduced due to stimulated recombination. Hence the modulation bandwidth can be improved applying higher input optical power. Fig. 4 represents this effect, where the relative bandwidth versus input optical power curve is plotted. The modulation bandwidth is constant in the unsaturated regime. It can be extended by about 70 percents compared with the unsaturated value in the saturated regime. Same time, linearity will be improved [7], but the modulation depth will be decreased. Hence the detected electrical signal at the centre part of the system will be improved, because the higher output power of the RSOA modulator.

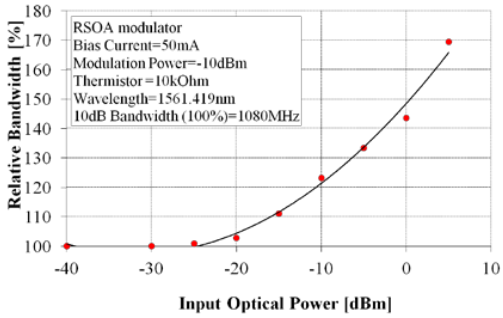


Fig. 4. Measured bandwidth improvement versus input optical power

D. Modulation bandwidth versus modulation index and temperature

The behavior of the RSOA is temperature sensitive. However it does not cause significant change in the modulation bandwidth. In a similar manner, the level of the electrical modulation power has no any remarkable effect for the modulation bandwidth. Naturally, the modulation depth, therefore the detected electrical power is proportional with the electrical modulation power. Hence the higher modulation signal is more effective, but the level of modulation electrical signal is limited by electrical nonlinearity.

To summarize the measured results, the optimization of operational and environmental parameters can double improve the modulation speed.

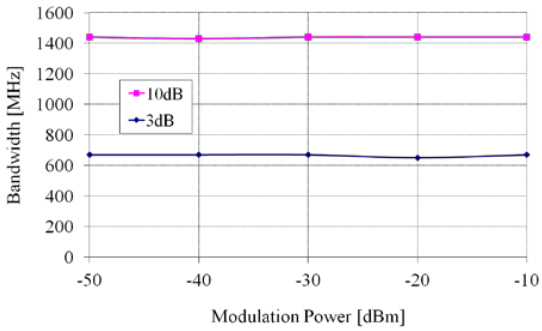


Fig. 5. Measured 3dB and 10dB bandwidths versus modulation power

IV. SIMULATION WORK

The experimental work presented the effect of the environmental parameters (like temperature, input optical power, bias current, etc.). However the investigation of inside device parameters with simulation method is more efficient.

A. Description of the model

An optical field which propagates along a travelling wave semiconductor optical amplifier was considered in the model. The interaction of light with carriers in the SOA is governed by the carrier rate and field propagation wave equations. The amplifier's output power is calculated by solving numerically the coupled rate and wave equations [5]. First the operating point is calculated by a steady state consideration. Next a change in the carrier density around the mean value was considered due to a change in the injected signal and the differential equation was obtained. The associated change in the stimulated and spontaneous emissions can be calculated.

$$\frac{dN}{dt} = \frac{I}{e \cdot A} - R_{sp}(N) - v_g \cdot \Gamma \cdot g_m \cdot S$$
$$\frac{dS}{dz} = (\Gamma \cdot g_m - \alpha) \cdot S \tag{1}$$

Where N is the carrier density, I is the injection current, e is the electron charge, A is the volume of active layer, $R_{sp}(N)$ is the spontaneous recombination rate, v_g is group velocity, Γ is the optical confinement factor, g_m is the material gain, α is the internal loss, S is the photon flux density, t is the local time, z is the spatial coordinate along the amplifier.

In case of modulation, the current is divided into dc and modulation part. So the carrier density and consequently the photon density include modulation part, too.

$$I = I_{DC} + \Delta I \cdot \exp(i \cdot \omega \cdot t)$$
$$S(t,z) = S_{DC}(z) + \Delta S(z) \cdot \exp(i \cdot \omega \cdot t) \tag{2}$$
$$N(t,z) = N_{DC} + \Delta N(z) \cdot \exp(i \cdot \omega \cdot t)$$

Based on this model the modulation amplitude of the photon density can be calculated at the output of the device.

$$\frac{d\Delta S}{dz} = (g_{sat}(z) - \alpha) \cdot \Delta S(z) + \Gamma \cdot a \cdot \Delta N(z) \cdot S_{DC}(z) =$$
$$= (g_{sat}(z) - \alpha) \cdot \Delta S(z) +$$
$$+ \Gamma \cdot a \cdot \frac{\frac{\tau_s}{e \cdot V} \cdot \Delta I(z) - \frac{v_g \cdot \tau_s \cdot g_{sat}(z)}{\Gamma} \cdot \Delta S(z)}{1 + i \cdot \omega \cdot \tau_s + \frac{S_{DC}(z)}{S_{sat}}} \cdot S_{DC}(z) \tag{3}$$

where g_{sat} is the saturated gain, S_{sat} is the saturated photon flux density, τ_s is the carrier lifetime, a is the differential gain.

The equation suggests that an increased bandwidth will be obtained with increased current, input optical power, differrential gain and confinement factor.

The carrier density is non-uniform along the SOA active region. To solve the problem the SOA is divided into many longitudinal sections and in each section my model assumes both uniform carrier and photon densities. The sectioned amplifier cavity is used to account for longitudinal effects; and the spatial variation of the material gain and other parameters of the SOA can be captured. The suitable adjustment of the model parameters enables the simulation of both in-line and reflective devices. The effect of spatial dependence in in-line device is more significant than in reflective device.

B. Simulation results

The slope of the transfer function is determined by different effects. The local modulation is low pass type, but the saturation and propagation effects [13] causes high pass filtering and it can improve the modulation bandwidth. Fig. 8 represents the typical transfer functions in case of different device length and bias current. During the simulation the modulation response was investigated, next 3 dB and 10 dB modulation bandwidths were calculated from the simulated transfer function.

1) Local modulation, operation point

The general transfer function is low-pass type and the bandwidth is usually limited by the carrier lifetime. The carrier lifetime is inversely proportional to the recombination rate. The carrier recombination rate can be described by different parts; there are the nonradiative recombination rate, the radiative recombination coefficient (spontaneous and stimulated) and the defect or Auger recombination coefficient. The effective carrier lifetime depends on the operating conditions. At low input optical power (unsaturated regime), the spontaneous and non-radiative recombination rates are dominant, and the carrier lifetime depends on these recombination terms. So the bandwidth increases by increasing the electrical bias current, because it increases all recombination terms and the carrier lifetime is decreased. The experimental work validated this effect (Fig. 3). On the other hand, it is advantageous to employ long SOA, since it tolerates larger bias currents, and have a larger differential gain for a given chip gain compared to short SOA.
The carrier lifetime in amplifier is larger than in laser because of the smaller photon density. The unsaturated carrier lifetime of a typical device is about 200-300 ps. It determines less than 1 GHz bandwidth.

2) Saturation effect

High input optical power and electrical current induce high photon density inside the active zone increasing the stimulated recombination rate. So, the stimulated recombination rate tends to overcome the spontaneous and non-radiative recombination terms. The stimulated lifetime is reduced and the average carrier lifetime is also smaller. The saturation level depends on the electrical bias current. At large optical input power, the saturation effect is much stronger than with low input injection at low bias current. Similarly, the saturation effect is significant at high bias current, when signal and ASE

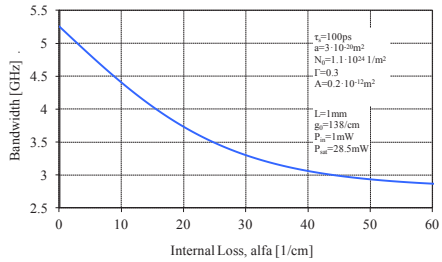


Fig. 6. Simulated bandwidth versus internal loss

are strongly amplified along the device. Consequently, the carrier lifetime decreases (the bandwidth increases) as the optical power is increased. Contrary, the bandwidth decreases as the internal loss increases, because the attenuation of the optical signal higher, namely the optical power is lower (Fig. 6). The experimental work validated this effect (Fig. 3). So, the 10-20 ps carrier lifetime can improve the bandwidth to the 100 GHz regime.

3) Propagation effect

The non-uniformity of the carrier lifetime along the device affects the transfer function. To illustrate the intrinsic propagation effect multisection model was applied. The current amplitude was independent of localization, but the spatial variation of average carrier and photon densities were taken into account.

The simulation results demonstrated the photon density modulation amplitude is high pass filtered as it travels along the SOA, since higher frequencies saturate the amplifier less. This tends to compensate the low pass filter type transfer function and increase the bandwidth of the device. Consequently, the whole transfer function has resonance, but it is also low past type.

So, the modulation bandwidth increases versus device length (Fig. 7). The slope of the curve depends on the optical power propagating over the device. As the optical power increases (the internal loss decreases) the situation goes to saturation and the propagation has more significant effect.

Summarizing, the modulation bandwidth cannot be accounted by the simple low pass transfer function determined by the effective carrier lifetime. It is affected by the evaluation of the signal as it propagates through the amplifier and the saturation. Figure 8 shows modulation responses for different

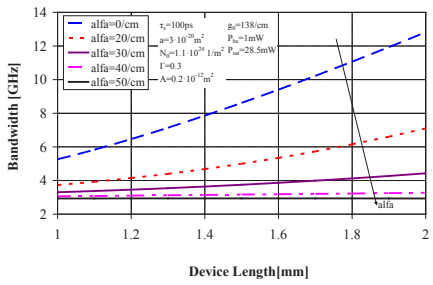


Fig. 7. Simulated bandwidth versus device length

amplifier lengths and different values of the internal loss. It represents the effect of the propagation. As the device length increases the bandwidth is improved and a resonance can be observed. On the other hand lower attenuation, namely higher intensity also improves the bandwidth.

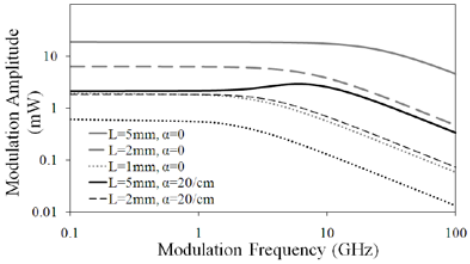


Fig. 8. Simulated modulation response for different amplifier lengths (L) and internal losses (α)

4) Effect of microwave and optical signal mismatch

The modulation signal propagates with a speed different from the optical mode the same or opposite directions and it is attenuated over the device. The model can take into account these effects, if appropriate phase and amplitude of the current density change are applied in the sections. The phase velocity of the microwave is in the range of 8-12% of the velocity of the light in vacuum for the frequencies in the range of 5-40 GHz [12]. So the refractive index of the microwave signal (n_μ=light speed/microwave signal speed) is in the range 14.3-8.3. The mismatch between the microwave and the light leads to a dip in the modulation response (Fig. 9).

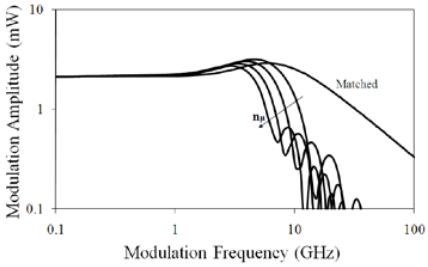


Fig. 9. Simulated modulation response for different mismatches

Fig.10 demonstrates that the modulation bandwidth decreases versus the level of the mismatch. The shape of the curve depends on the length of the device.

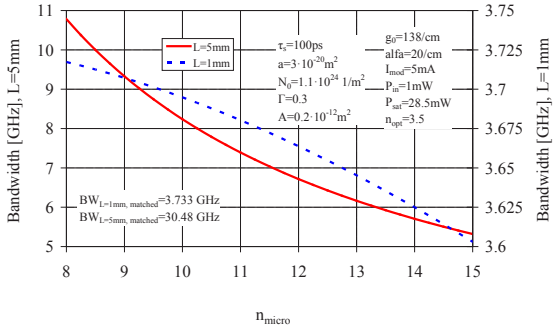


Fig. 10. Simulated bandwidth versus mismatch with different device length

5) Effect of bonding position

The simulation results show that the waveguide (scattering loss) plays a very important role. The transfer function depends on the location of the bonding point. In the previous simulations the modulation signal was coupled at the starting point of the device. So I calculated with copropagation between the microwave electrical and the optical signals. Fig. 11 presents the transfer functions when the bonding position is the end of the device. It causes counter propagation. The transfer function shows the effect when the microwave signal is faster than the optical signal. Fig. 12 shows the situation when the bonding position is the center of the device. It means both co- and counter propagations.

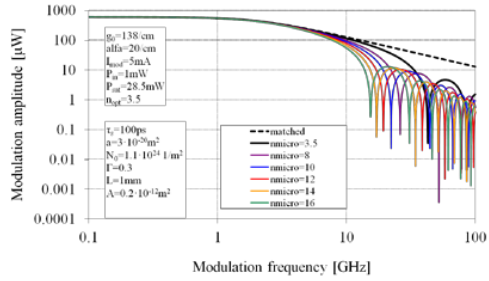


Fig. 11. Effect of bonding position, bonding point: end of the device

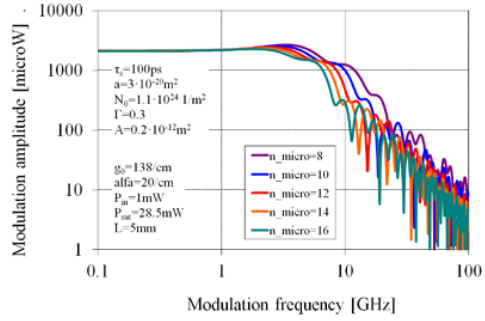


Fig. 12. Effect of bonding position, bonding point: middle of the device

6) Effect of microwave attenuation

In case of a real device the microwave modulation signal is attenuated, too. The level of the attenuation depends on the microwave frequency, it is about 500 dB/cm at 40 GHz and „just" 10 dB/cm at 10 GHz. The attenuation of the modulation signal decreases the intensity modulation amplitude, but same

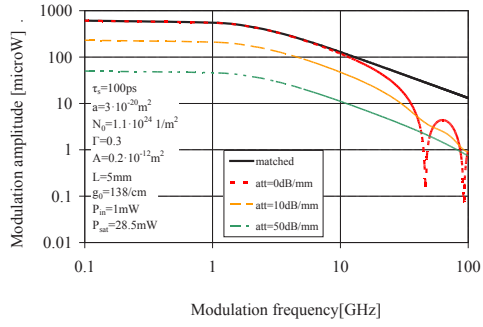


Fig. 13. Effect of electrical attenuation

time it decreases the mismatch effect. Hence reduced dips can be observed (Fig. 13).

V. CONCLUSION

This paper has provided a summary of the work, which proves the importance of electrical modulation bandwidth investigation of in-line and reflective semiconductor optical amplifier based external intensity modulators in next generation optical access networks. It means both baseband modulated WDM-PON and high speed WDM-RoF applying Subcarrier Multiplexing. The paper proposes device and system level solutions for enhancing the data speed at which the device is directly modulated.

The experimental results confirm that a significant improvement in the modulation bandwidth can be achieved by proper adjustment of the operation point and environmental conditions. Mainly the bias current and the level of the incident optical power determine the carrier lifetime. The modulation bandwidth can be doubled by this approach.

Bandwidth enhancement can be also achieved by applying optimal device structure. The comparison of in-line and reflective structure and the effect of device length and electrode structure were investigated by simulations.

The obtained results suggest that the methods enable the use of the RSOA as intensity modulator with improved performance at an extended data rate or subcarrier frequency.

Eszter Udvary (M'97) was born in Budapest, Hungary. She received Ph.D. degree in electrical engineering from Budapest University of Technology and Economics (BME), Hungary, in 2009. Her research interests are in the broad areas of optics, include communication systems, optical and microwave interactions, electro-optical devices and sensors.

She is currently an associate professor at BME, Department of Broadband Infocommunications and Electromagnetic Theory where she leads the Optical and Microwave Telecommunication Lab. She has authored or co-authored more than 80 papers and received more than 60 citations. Dr. Udvary is a member of IEEE and HTE.

REFERENCES

[1] Tong S., Beltrán, M., Rui Z., Anandarajah, P.M., Lorente, R., Barry, L.P., "60 GHz Radio Over Fiber System Based on Gain-Switched Laser," IEEE Journal of Lightwave Technology, vol.32, no.20, pp.3695-3703, Oct.15, 2014

[2] Takushima, Y., Cho, K. Y. and Chung, Y. C., "Design Issues in RSOA-based WDM PON", PhotonicsGlobal, IPGC 2008, Singapore, 2008. IPGC 2008. IEEE, 8-11 Dec. 2008, pp.1-4

[3] Papagiannakis et all,"Investigation of 10-Gb/s RSOA-based upstream transmission in WDM-PONs utilizing optical filtering and electronic equalization," IEEE Photonics Technology Letters, vol. 20, pp.2168–2170 (2008)

[4] Hong M. et all, "10-Gb/s transmission over 20-km single fiber link using 1-GHz RSOA by discrete multitone with multiple access," Optics Express 19, B486–B495 (2011).

[5] Udvary, E. ; Berceli, T., "Optical subcarrier label swapping by semiconductor optical amplifiers", IEEE Journal of Lightwave Technology, 2003, vol.21, issue 12, pp. 3221-3225

[6] Udvary, E., Berceli, T., "Semiconductor Optical Amplifier for Detection Function in Radio over Fiber Systems", Journal of Lightwave Technology, 2008, vol.26, issue 8, pp. 2563-2570.

[7] Udvary, E., Berceli, Ti., "Improvements in the Linearity of Semiconductor Optical Amplifiers as External Modulators", IEEE MTT Transactions & J Lightwave Technology Special Issue on Microwave Photonics, November 2010, Volume: 58 Issue:11, pp.3161 - 3166

[8] Udvary, E., Berceli, T., "Performance Improvements for Optical Links Applying Semiconductor Optical Amplifiers", Mediterranean Microwave Symposium. Morocco, 2009.11.15-2009.11.17., pp. -. Paper MMS2009-8.

[9] Watanabe, T. et al., "Transmission performance of chirp-controlled signal by using semiconductor optical amplifier", IEEE Journal of Lightwave Technology, August 2000, pp. 1069-1077

[10] Schrenk, B., Valicourt, G., Omella, M., Lazaro, J. A. Brenot, R., Prat, J. "Direct 10-Gb/s modulation of a single-section RSOA in PONs with high optical budget", IEEE Photon. Technol. Lett., 22, Mar. 2010.

[11] Kim, H. "10-Gb/s operation of RSOA using a delay interferometer", IEEE Photon. Technol. Lett., 22, Mar. 2010.

[12] Kim, H. "Transmission of 10-Gb/s directly modulated RSOA signals in single-fiber loopback WDM PONs", IEEE Photonics Technology Letters, vol. 22, no. 14, Jul. 15, 2011.

[13] Durhuus, T., Mikkelsen, B., Jørgensen, C., Carsten, S. Elmholdt, K. "All optical wavelength conversion by semiconductor optical amplifiers". IEEE Journal of Lightwave Technology, Vol. 14, 1996, p. 942-954.

[14] Cho, K. Choi, B. Takushima, Y. and Chung, Y., "25.78-Gb/s operation of RSOA for next-generation optical access networks," IEEE Photonics Technology Letters, IEEE23, 495–497 (2011).

E-band terrestrial radio – propagation and availability aspects

László Csurgai-Horváth and István Frigyes

***Abstract*— This paper is focusing on the E-band terrestrial radio channel, especially on availability calculations and on the relationship between fade duration statistics and availability. This frequency band is applied in high speed data transmission links between endpoints at a distance of few kilometres. There is a high demand for such connections with high reliability features; mobile backhaul networks are good example for that. The main propagation impairment causing bursty drop-outs on the transmission links is rain. Based on our long term measurements we will show the availability characteristics of the E-band radio and propose a new method for availability calculations. We investigate the relationship between fade duration statistics and the availability as well.**

***Index Terms*— E-band propagation, rain fading, path attenuation, availability, fade duration**

I. INTRODUCTION

THIS paper derives the availability characteristics of millimetre-band radio links established by E-band devices with high transmission speed, exceeding the Gbit/s rate. The demand for such connections is high and even increasing because the new broadband services like high-speed internet, multimedia services, etc. require higher and higher bandwidth and data transmission speed. However, rain as the main reason of channel degradation severely influences the quality of service and may cause unavailable periods during the transmission. Therefore in our paper we concentrate on the rain attenuation on terrestrial E-band radio connections, furthermore on its first and second order statistics. Based on long-term measurements we derive some of the channel's propagation and availability characteristics.

A comprehensive description of atmospheric effects that are significantly influencing the millimetre-band propagation channel can be found in [1]. Several calculation methods are detailed in this book, while the related ITU-R recommendations and standardized computation procedures are defined in [2]-[5]. In the beginning of our paper we summarize our latest results on terrestrial E-band propagation measurements and compare them with the related ITU-R path attenuation models.

In our paper one of the question that we focused on is the availability, thus the operation of the investigated E-band link at low input signal levels –near to the fade margin- is particularly interesting. The dynamic range of our measurement system is limited by the nature of the applied equipment. Therefore in order to increase the dynamic range and ensure to study the operation even during the deep fade events we apply some signal processing techniques [12] to reduce the noise floor and extend the observed measurement range. After data processing we approximate the upper tail of the measured attenuation distribution with an exponential function that serves to create a final closed form to estimate the required antenna diameter for a specific probability margin of interruption.

As the main scope of our interest is availability calculation, we apply the above mentioned dynamic range extension method to estimate the attenuation statistics in this frequency band. As the data processing technique and the first model constitution was already published in [8], besides the refinement of this model and using a longer measured dataset for higher precision, we are focusing on system availability calculations and prove the correctness of the relationship between interruption probability margin and the antenna diameter.

A second important question is in this paper the connection between fade duration statistics and the availability of the radio link. According to its definition, fade duration is the amount of time that the signal envelope stays below a specific level [5]. The distribution of fade duration at specific levels – especially around the fade margin- serves as an alternative method to calculate the availability of the radio system. It will be shown that the attenuation distribution and the fade duration statistics results similar values for the probability of interruption. Nevertheless, fade duration statistics comprise additional information about the nature of fading events. Therefore it can be used to indicate the number of bits or blocks that are affected by a fade event. This information is well applicable during the design of coding schemes for wireless channels [6].

This paper is organized in five sections. The introduction is followed by the description of the measurement environment, the discussion of the data processing method, comparing with

Submitted January 30, 2015, revised March 14, 2015.

László Csurgai-Horváth is with Budapest University of Technology and Economics (BME), Department of Broadband Infocommunications and Electromagnetic Theory, Budapest, Hungary. He is member of IEEE and HTE. (e-mail: csurgai@mht.bme.hu).

István Frigyes is with Budapest University of Technology and Economics, Department of Broadband Infocommunications and Electromagnetic Theory, Budapest, Hungary. He is senior member of IEEE. (e-mail: frigyes@mht.bme.hu).

This research work was carried out within the frame of cooperation between Telenor Hungary and BME.

the ITU-R recommendations and introducing the improved model for attenuation statistics. Section III derives the relationship between availability and attenuation statistics. In Section IV we study how fade duration statistics can be applied to estimate the availability. The paper ends with some concluding remarks.

II. E-BAND MEASUREMENTS AND STATISTICS

Our measurements were performed on an experimental 72.56 GHz radio link [11], serving as part of the backhaul network of Telenor Hungary connecting a base station with the corresponding base station controller. The measured endpoint of the radio link was located in a dense built-in city area (Budapest) at the top of a building at a height of 102 m. The exact geographical position is N47.48° latitude and E19.06° longitude. Further technical parameters can be found in Table I.

TABLE I.
THE E-BAND LINK PARAMETERS

Frequency	72.56 GHz
Path length/Effective path length (d/d_{eff})	2.3/2.08 km
Transmit power	16 dBm
Antenna gain	44 dBi
Antenna diameter	31cm
Receive sensitivity	-61 dBm
Polarization	horizontal
Receiver noise figure (NF)	7 dB
Bandwidth at 1000Mbps (B)	1400 MHz
Minimum signal-to-noise ratio for BER<10^{-3} (C/N)	12 dB
Campaign period	05.2009-11.2012

During a 43 month measurement campaign we recorded the level of received power with 1 sample/sec rate. The median value of the received power level was -42.66 dBm; we applied this as clear sky level during the relating calculations. In Fig. 1 a weekly time series with typical rain attenuation events can be seen.

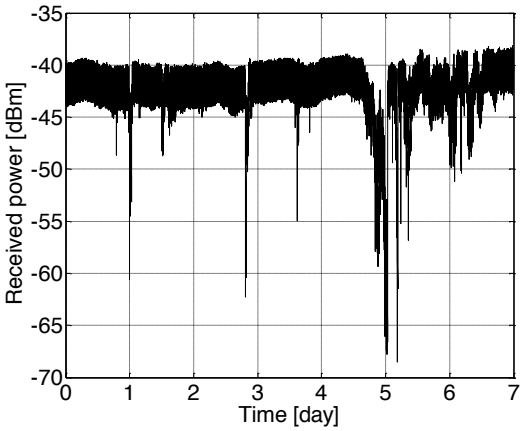


Fig. 1. One week received power time series with rain events

Considering the system maintenance periods, valid data was recorded in 84.2% time of the whole campaign. As the system was dedicated only for this measurement, there was no real data transfer on the radio link. This explains the relatively high outage rate caused by different other tests that we performed.

The clear sky level is the function of the transmitted power, transmitter and receiver antenna parameters and receiver gain. We considered these parameters as constant during the measurement campaign. The clear sky value is characterizing the actual measurement system and serves as a reference level for attenuation calculations.

The atmospheric scintillation and the constant but not negligible noise of some system elements (amplifiers, downconverters, etc.) cause a continuous, high-speed low-level variation of the received power. This variation will be removed by an appropriate filtering as it is detailed in Section IV. The reason of the observable high attenuation peaks is rain; other effects like interference, shadowing, multipath propagation is negligible due to the careful link design and the lack of interfering radio sources in this frequency band.

Generally a terrestrial radio link like the investigated one operates trouble-free as long as the attenuation is less than the fade margin of the system. The quality of the radio connection only decreases when the attenuation is high enough to reduce the received power level near to the receiver sensitivity. Therefore it is especially important to know the high-attenuation statistics of the received power time series. In order to increase the accuracy of the low-level signal measurement, we applied the noise subtraction method to reduce the noise floor and extend the dynamic range of our measurements [12].

The details of this method were presented in [8] therefore only by referring to this paper we apply its results. With (1) we can calculate the actual attenuation A_{actual} after the noise subtraction:

$$A_{actual} = \frac{A_{measured}}{1 - A_{measured}(N / P_{median})}, \quad (1)$$

where $A_{measured}$ denotes the measured attenuation, P_{median} is the clear-sky level, N is the noise power.

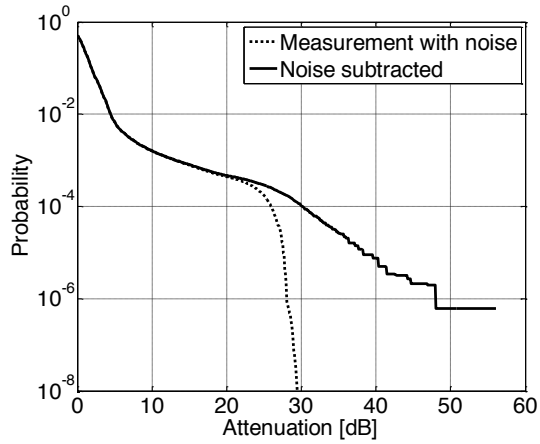


Fig. 2. Measured attenuation CCDF and its value after noise subtraction

Fig. 2 depicts the measured attenuation CCDF and the result of noise subtraction by using (1). The value of N/P_{median} can be

determined from the vertical tangent of the measured data CCDF (Complementary Cumulative Distribution Function) at the highest observable attenuation. The noise subtraction significantly increases the dynamics in the high attenuation range that is very important for the forthcoming availability calculations.

To estimate the long-term statistics of rain attenuation, the recommendation ITU-R-P 530 [2] can be applied for E-band as well. We compared the ITU model-based rain attenuation CCDF with our measurements after noise subtraction (Fig. 3). As in the range of high attenuation, the two curves are rather different from each other, in [8] we proposed the following exponential approximation of the attenuation probability in the *A=33-45dB* range:

$$\begin{aligned} &if\ 33dB < A < 45dB: \\ &p_A = a \cdot \exp(-d_{eff} b \cdot A) \\ &a = 0.2991 \\ &b = 0.1281 \end{aligned} \quad (2)$$

In (2) a and b are empirical parameters, while d_{eff} denotes the effective path length [2]. Effective path length is the corrected value of the real path length and it is the function of frequency and the actual climatic zone. The validity of (2) is fairly high if we consider the long (43 months) measurement period, and accurately estimates the probability of the E-band attenuation in the higher range.

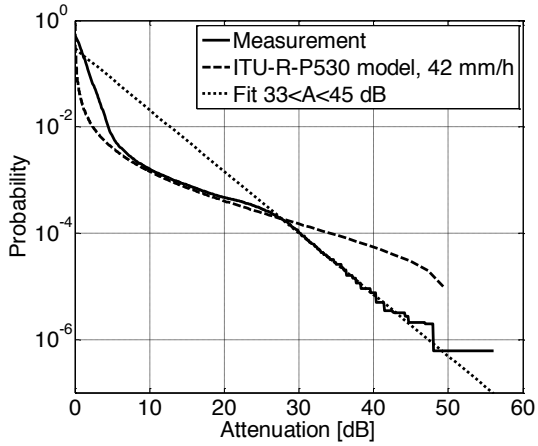


Fig. 3. Measured attenuation CCDF, the ITU-R approximation and the proposed exponential fit

For the ITU-R calculations the rain zone *K* [3] was applied with $R_{0.01}=42mm/h$, as this is recommended for the measurement location (Budapest).

III. ATTENUATION CHARACTERISTICS AND AVAILABILITY

The ITU-R F.1703 recommendation defines the availability objectives for the backhaul network [4]. In [9] the availability ratio is defined as *AR=0.9995 (99.95%)*. This value is significantly lower than the value required by different mobile operators (usually 0.99998). Based on our measurements firstly we determine the availability of the investigated E-band radio link.

The minimum required input signal for nominal operation can be determined from the physical link parameters (see Table I.). The value of the receiver threshold *Th* is the following [7]:

$$\begin{aligned} &Th = kT_0 + NF + 10\log B + C/N = \\ &= -174 + 7 + 91.46 + 12 = -63.54dBm \end{aligned} \quad (3)$$

This value yields the first estimation of the link availability. The difference between the threshold *Th* and the nominal received power (-42.66 dBm) is 20.88 dB. This value can be considered as the upper attenuation limit of the error-free operation. The distribution of attenuation after noise subtraction (see Fig. 2) is $4.06 \cdot 10^{-4}$ at 20.88 dB, resulting 99.9594 % link availability.

A different approach to calculate the availability is using the physical system parameters (path length, carrier frequency and system gain). In the following we will do such calculations and prove the method by comparing the result with our long-term measurement data.

For a given radio link the antenna is the most flexible component, therefore in [8] we constructed a practical equation to express the minimum required antenna gain G_A if we know the path length d, the carrier frequency f_C, and the system gain G_S. The main input parameter of the calculation is the p_m probability margin of interruption. The system gain can be expressed as the difference of the transmit power and the receiver threshold:

$$G_S = P_T - Th = 16dBm + 63.54dBm = 79.54dB \quad (4)$$

Applying (1) the final form to determine the minimal antenna gain for the required p_m interruption probability margin is (for details see [8]):

$$G_A^{dB} \geq \left(\begin{array}{c} 32.44^{\mathrm{dB}} + 20\log d^{km} + 20\log f^{MHz} + \\ + 0.3 d^{km} - \frac{1}{b d_{eff}^{km}} \ln \frac{p_m}{a} - G_S^{dB} \end{array} \right) / 2 \quad (5)$$

By parameterizing the above equation with $p_m=4.06 \cdot 10^{-4}$, for the investigated link we get $G_A=41.4dB$. If we compare the antenna gain given by the manufacturer [11] (*44dB*), the precision of (5) is apparent.

IV. FADE DURATION STATISTICS AND AVAILABILITY

In this section we will derive the relationship between fade duration statistics and link availability. The distribution of fade duration is often called as second-order statistics of the attenuation process. According to its definition, fade duration is always considered at a specific attenuation threshold and it is the time interval between two crossings of the signal level above the threshold [5].

Fade duration statistics is an additional tool of radio link design engineers because it describes a different aspect of the propagation channel. The attenuation statistics informs us about the probability that the fading depth exceeds a specified level, but the length of the individual fade events and thus the possible outage periods cannot be determined with examining only the attenuation distribution. This can be demonstrated with the identical attenuation statistics for two different radio connections: one with several short fade events and another with less but longer fade events. This kind of channels may result different operational characteristics, therefore solely the attenuation distribution is not enough to describe them.

Fade duration can be characterized for example with the statistics of number the fade events longer than a specific duration. This kind of statistics can be seen in Fig. 4 for the investigated E-band link, depicting the most relevant (≥20dB) threshold ranges in point of view the availability. Before calculating the fade duration, the effect of scintillation (fast, low level fluctuation) should be removed from the measured time series, as proposed in [5]. According to the recommendation a low-pass filter can be applied with 3 dB cut-off frequency at 0.02 Hz to eliminate the scintillation and other rapid variations of rain attenuation. If scintillation and rapid variations of the attenuation process are not filtered out, the signal will exhibit stronger fluctuations and the fade duration statistics will be significantly different. During data processing we applied a 20 sec moving average filter that performs the recommended filtering process.

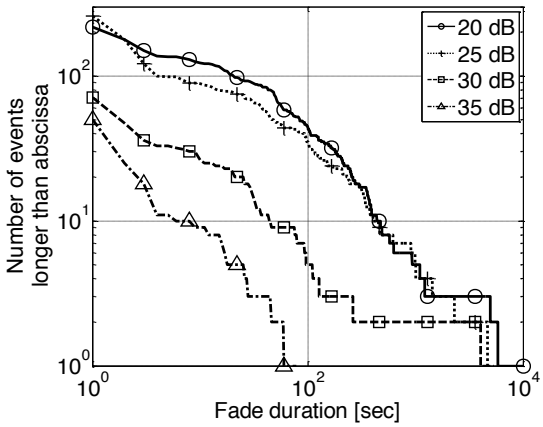


Fig. 4. Fade duration statistics based on 43 month measurements

In the previous sections it was shown how the long term attenuation statistics and the physical system parameters can be applied to estimate the availability for E-band radio.

While the system gain determines the maximal attenuation that can be allowed for a specific error probability, the fade duration statistics helps to determine the length of the unavailability periods and it may lead to different channel characterization aspects.

A. Availability and measured fade duration statistics

Rain with moderate intensity usually does not affect the quality of the radio connection if the attenuation stays below the fade margin. Therefore the length of deep fades becomes particularly interesting if a system operates near to the fade margin. Fade duration statistics is applicable to extract relevant data about the timing characteristics of the fading process and the behaviour of the individual fade events can be even better observed. In the following the statistics of the measured fading length will be shown at this threshold an above it, respectively.

We have seen in Section III that the attenuation threshold of the error-free operation is 20.88 dB, so the closest integer value, 21 dB will be also included in the following calculations. In Table II, the number of the fade events with a specific/or longer duration at different thresholds is given. The total number of recorded samples was $93.84 \cdot 10^6$ for the whole measurement campaign.

TABLE II.
DETAILED FADE DURATION DATA (STATISTICS OF 43 MONTH)

Duration [s]	10	100	1000
Number of events at 20dB	123	44	5
Number of events at 21dB	108	43	5
Number of events at 25dB	88	35	4
Number of events at 30dB	26	5	2
Number of events at 35dB	9	0	0

The total duration of the fade events at specific thresholds are summarized in Table III, considering the whole measurement campaign. In the table one can find also the durations projected to a 1 year period as well.

TABLE III.
TOTAL FADE DURATION

Threshold [dB]	43 month duration [s]	Projected to 1 year [s]
20	43107	12030
21	40410	11277
25	26112	7287
30	10209	2849
35	366	102

To evaluate these results we can find the p_m probability margin by dividing the 1 year fade duration at 21 dB with the total duration of year:

$$p_m = \frac{fd^{[21dB]}}{d^{[year]}} = \frac{11277s}{24 \cdot 3600 \cdot 365s} = 3.58 \cdot 10^{-4}, \qquad (6)$$

that is 99.9642% availability. It is very close to 99.9594% link availability that was calculated from the attenuation statistics. It is not surprising that the fade duration statistics provides the same availability rate as the attenuation statistics.

The main difference and the relevance of fade duration statistics can be derived from Fig. 4. There are numerous long duration fade events even at high attenuation that may cause the radio channel unavailable for a period that is longer that the duration which is tolerable by the system. This is the main

reason why fade duration statistics is important: a deep fading may interrupt the operation of the link for significant period; however its impact to the yearly attenuation statistics could be negligible.

In such cases to improve the availability of the radio link, one of the possible solutions is to increase the transmit power or the antenna size. Increasing the transmit power have several side-effects: besides the increased power consumption it increases the probability of unwanted interference as well. Larger antennas affect the system costs and several mechanical problems may also arise. Besides these obvious but not always feasible solutions, various coding techniques are available. In the following section the effects of block coding and their relationship with the length of fade events will be shortly discussed.

B. Block coding and fade duration

The propagation channel often produces burst errors instead of independent ones due to the behaviour of the fade events. This is one of the main reasons why fade duration statistics may help to select the right modulation scheme, error correction method or interleaving structure. These aspects will be briefly discussed in the following.

Forward Error Correction (FEC) adds redundant information to the source data in order to improve the capacity of the radio channel. One of the major forms of FEC is block coding that is often applied to channels with burst impairments [10].

The RF interface of the investigated E-band device operates with BFSK modulation mode and utilizes an RS(204,188) FEC [11]. Considering 1 Gbps data rate and the redundancy of the FEC, the bit time T_b=*0.92 ns* and the block time is *188 ns*. By comparing this time with the much higher durations of fade events, it is obvious that the difference between them is several orders of magnitude.

At this high data rate even a more complex code like low-density parity-check (LDPC) block code or LDPC convolutional code [13] transfers several blocks during the rain fade events. This means that the error correction capability of these codes cannot completely mitigate the effect of the rain fading. The coding gain of LDPC is around 2-3 dB [14] compared it with convolutional code. On the other hand, attenuation during rain events could be significantly higher than this level. Considering this assumption, LDPC coding enhances the system performance when the radio link operates near to the system threshold. Another benefit is that the error correction expands the connectivity range of the individual links therefore it reduces the installation costs as larger coverage area can be achieved with less equipment.

As a conclusion, block coding at the physical layer may eliminate the effects of fast fading, scintillation or white Gaussian noise but it is not responsible to combat again slow fading. One of the solutions is the introduction of coding at higher level: at the transport, network or application layer where the transmission structure is longer than the duration of fade events (several seconds or minutes), or the protocol ensures retransmission in case of an error. If a feedback channel exists and channel state information (CSI) is available, with automatic repeat request (ARQ) or by using adaptive power control (APC), further improvement can be achieved to eliminate the effect of the slow and deep fading. This technique is the dynamic FEC, nevertheless this is out of the scope of our study.

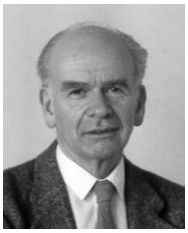
V. CONCLUSIONS

In this paper a 43-month duration E-band terrestrial radio-wave propagation measurement was analysed and investigated from point of view the RF link availability. We have shown the rate of the availability based on the measured attenuation statistics and we applied a new model by using the physical system parameters for the calculations and an approximation of the availability with using the fade duration statistics. It was shown that the attenuation statistics does not completely characterize the channel with rain fading. Therefore one of the known second-order statistics, fade duration distribution was calculated and investigated. This is an important statistics that informs us about the length of the fade events and provides additional information for radio link designers. A further application of second order statistics is to support the implementation of advanced fade mitigation techniques like block coding or adaptive transmission control.

REFERENCES

[1] L. Castanet (ed.), *Influence of the Variability of the Propagation Channel on Mobile, Fixed Multimedia and Optical Satellite Communications*, Shaker, 2008.
[2] ITU-R Rec. P.530-14 *Propagation data and prediction methods required for the design of terrestrial line-of-sight systems,* 2012.
[3] ITU-R Rec. P.837-5, *Characteristics of precipitation for propagation modelling,* 2007.
[4] ITU-R F.1703, *Availability objectives for real digital fixed wireless links used in 27 500 km hypothetical reference paths and connections,* 2005.
[5] ITU-R Rec. P.1623-1, *Prediction method of fade dynamics on Earth-space paths,* 2003-2005.
[6] Andrea Goldsmith, *Wireless Communications*, Cambridge University Press, 2005.
[7] A.F. Molish, *Wireless Communications*, Wiley, 2011.
[8] L. Csurgai-Horvath, I. Frigyes, J. Bito, *Propagation and availability on E-band terrestrial radio*, 6th European Conference on Antennas and Propagation, EUCAP 2012: pp. 73-75, 2012.
[9] Finnish Communications Regulatory Authority, *Guidelines for Implementation; Fixed Wireless Systems,* 2005.
[10] G. Corazza (ed.), *Digital Satellite Communications*, Springer, 2007.
[11] Bridgewave Communications Inc., *Bridgewave AR80 user's manual*, www.bridgewave.com
[12] Rohde&Schwarz, *Improved Dynamic Range with Noise Correction*, AN1EF76, 2010.
[13] R.M. Tanner, D. Sridhara, A. Sridharan, T.E. Fuja, D.J. Costello, *LDPC Block and Convolutional Codes Based on Circulant Matrices*, IEEE Transactions on Information Theory, vol.50, no.12: 2966- 2984, 2004.
[14] Da Xinyu, Wang Yanling, Xie Tiecheng, *Performance of LDPC Codes for Satellite Communication in Ka Band*, 5th International Conference on Wireless Comm., Networking and Mobile Computing, pp. 1-4, 2009.

László Csurgai-Horváth is an associate professor at Budapest University of Technology and Economics, Department of Broadband Infocommunications and Electromagnetic Theory. He received the M.Sc. degree in 1985 and the Ph.D. degree in 2010 in Electrical Engineering from BME. His current research interests are focused on microwave propagation measurement systems, modeling and time series synthesis for terrestrial and satellite radio links. He participates in several national and international research projects and he is the author of more than 60 journal and conference publications and several book chapters.



István Frigyes, PhD. habil, DSc. graduated at BME as Electrical Engineer in 1955. After working with various industrial institutions as member and later head of R&D departments he has been with BME since 1983 as associate professor and full professor since 1995, actually Professor Emeritus. His research interest included microwave antennas and circuits while in recent decades digital wireless communications (radio and optics), system design, propagation effects, modeling and countermeasures. He is senior life member of the IEEE, he founded and chaired the first Hungarian IEEE Chapter (Communication & Microwaves) and was member of various IEEE ComSoc boards for about two-and-half decades.

CALL FOR PAPERS

Special Issue on Smart Cities:

Crowdsourcing and M2M communication for a connected society

The urbanization of cities is increasing, and nowadays about 54 percent of the world's population lives in cities. By the year of 2025, this number will be around 70 percent. In big cities, this will put a lot more pressure on streets and traffic control. There is a growing importance of Information and Communication Technologies in profiling the competitiveness of cities. There is extensive ongoing research in a wide range of enabling information and communication technologies, including cloud and network infrastructure, wireless and sensing technologies, mobile crowdsourcing, social networking, and big data analytics for smart cities. The next step for the smart city is the automated city – one that is predictive and responsive without human intervention. Such a city could avoid traffic congestion before it occurs and distribute resources, such as emergency services and maintenance, without time-consuming human decision-making. In this Special Issue we will catch up with the latest research and product developments, measurement methods, application scenarios and concept studies.

Our journal is calling for original and unpublished contributions to this important area that will be peer-reviewed. Selected papers will appear in a Special Issue to be published in September of 2015. Original and unpublished papers should be submitted by 15th of July, and by 30th of September in the form of pdf files in IEEE format according to the formatting instructions available at

http://www.ieee.org/publications_standards/publications/authors/authors_journals.html#sect2

Contributions are expected from the following areas:

- Mobile crowdsourcing for urban analytics
- Sensing and IoT for smart cities
- ICT in road vehicles: on-board and connected car services
- Safety, security, and privacy for smart cities
- Crisis and disaster management in a smart city
- Human mobility modeling and analytics
- Senseable city networks
- Mobile crowdsourcing applications
- M2M communications architectures and middleware

The paper submission deadline is 31 July, 2015.

Guest Editors:

ISTVÁN GÓDOR is a research fellow at Ericsson Research, Traffic Analysis and Network Performance Laboratory of Ericsson Hungary. He is a member of the IEEE and a member of public body of Hungarian Academy of Sciences. He received both his M.Sc. and Ph.D. degree in Electrical Engineering from Budapest University of Technology and Economics, Budapest, Hungary in 2000 and 2005, respectively. He has been serving a number of Technical Program Committees or as referee for international journals and conferences, such as IEEE Communications Magazine, IEEE ICC, IEEE VTC, IEEE PIMRC, IEEE WCNC and the like. He has been awarded the 2014 IEEE Communications Society Fred W. Ellersick Prize. His research interests include network design, combinatorial optimization, cross-layer optimization, self-organizing networks, energy efficiency, traffic analysis and modeling.

VILMOS SIMON received his PhD from the Budapest University of Technology and Economics (BME) in 2009 and is currently an associate professor at the Department of Networked Systems and Services and Head of the Multimedia Networks and Services Laboratory. His research interests include self-organizing mobile networks, mobile crowdsensing, Internet of Things, spatial computing. He participated in several research projects including the EU ICST-FET FP6 BIONETS where he also acted as a WP leader. He published more than 40 papers in international journals and conferences, and acts as a reviewer or organizer for numerous scientific conferences. He serves as a president of the Telecommunications Section in the Scientific Association for Infocommunications Hungary.

MARIO KUŠEK is an Associate Professor at the University of Zagreb, Faculty of Electrical Engineering and Computing, Croatia. He holds an the Ph.D. degree (2005) in electrical engineering, major in telecommunications and informatics, from the University of Zagreb. His main research interests include distributed systems, software agents in next generation networks, and converged services on mobile terminals. He participated in two scientific projects financed by the Ministry of Science, Education and Sports of the Republic of Croatia, two EU COST actions, one bilateral project with The Telecommunications Research Center Vienna (FTW) and he led research projects funded by companies Ericsson Nikola Tesla, Kate-Kom and Agrokor. He has coauthored over 70 scientific journal and conference papers. Prof. Kušek is a member of IEEE, currently also serving as a the Chair of the IEEE ComSoc Croatia Chapter, the KES International and the European Telecommunications Standards Institute (ETSI). He published more than 70 papers in journals, conference proceedings and books in the area of distributed systems, multi-agent systems, self-organized systems and machine-to-machine (M2M) Communications.

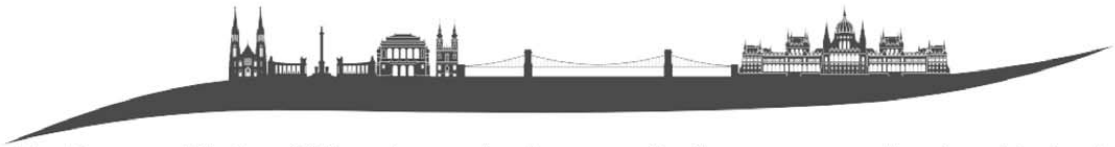
May 20-22, 2015

Call for papers

Budapest, Hungary

EUROPEAN WIRELESS 2015

ew2015.european-wireless.org



The European Wireless (EW) conference is a key venue for European researchers to get in touch with the latest trends in wireless communications and networking. The 21th EW conference will take place in Budapest, the lively capital of Hungary, organized by the Budapest University of Technology and Economics (BME). The main theme of EW 2015 will be "5G and beyond."

Topics of interest include, but are not limited to, the following:

FW Fundamental Wireless

- Modulation and coding for wireless communications
- Signal processing for wireless communications
- Wireless models, synchronization, estimation, equalization
- MIMO systems, space-time coding, diversity
- Fundamental limits, information theory for wireless
- Multiple access schemes, multiuser detection
- Interference mitigation and management
- Distributed coding and cooperative diversity
- Localization and positioning in wireless systems
- Source and joint source/channel coding
- Spectrum sensing and wireless parameter estimation

TW Technology for Wireless

- Migration, integration, and convergence towards 5G
- WiFi, LTE, 3GPP, Heterogeneous Networks
- Wireless LAN/PAN/BAN, Ad Hoc, Mesh networks
- Near-field communications & RFID
- Wired-wireless integration
- Ultra-Wideband (UWB) communications
- Mm-wave communications
- Wireless sensors & actuators networks
- Vehicular and disruption tolerant wireless networks
- Optical wireless and visible light communications

EW Efficient Wireless

- Power: green communications, energy harvesting devices
- Spectrum: cognitive radio, spectrum-aware techniques
- Implementation: low-complexity and scalable systems
- Reliability: robust and dependable wireless systems
- Cost: low-cost radio, sustainable wireless
- Security: privacy and trust in wireless networks

AW Advanced Wireless

- Protocols and architectures for wireless networks
- Channel coding, error protection, network coding
- Cross-layer issues in wireless networks
- Cognitive radio for wireless communications
- QoS and resource allocation in wireless networks
- Mobile/wireless networks modeling and simulation
- Localization and positioning in wireless scenarios
- Optimization and game theory for wireless
- Topology control, self-organizing wireless networks
- Transport layer for wireless communications
- Relays and buffers in wireless networks
- Tools for modeling and analysis of wireless systems

PW Practical Wireless

- Implementation issues in wireless systems
- Testbeds and experimental systems
- Antenna and RF modeling and design
- Mobility management and billing technologies
- Mobile apps and platforms
- Regulation and standardization for wireless
- Context awareness
- Emerging applications in wireless networks

VW Vision for Wireless

- Personal wireless communications beyond 5G
- Software defined wireless networks and re-configurability
- M2M communications and the Internet of Things
- Storage, smart caching, and cloud for wireless
- Wireless social networks, participatory computing
- Molecular and nano-scale wireless communications
- New disruptive concepts for wireless systems

General chair: Hassan Charaf (BME, HU)
General co-chair: Marcos Katz (University of Oulu, FI)
TPC chair: Leonardo Badia (University of Padova, IT) and Mischa Dohler (King's College London, UK)
Steering committee chair: Frank Fitzek (Aalborg University, DK)
Tutorial chairs: Sergio Palazzo (University of Catania, IT) and Morten V. Pedersen (Aalborg University, DK)
Workshops: Christian Weitfield (TU Dortmund, DE), Péter Ekler (BME) and László Lengyel (BME).
Publicity chair: Stefan Valentin (Alcatel Lucent, DE) and Leonardo Militano (Mediterranea University of Reggio Calabria, IT)
Financial chair: Volker Schanz (VDE ITG, DE)
Secretariat/Registration: Christina Gaußmann (VDE ITG, DE)
BME local committee: István Vajk, János Levendovszky, Sándor Imre, Bertalan Forstner, László Lengyel, Péter Ekler, Imre Kelényi, József Bíró, János Tapolcai, Attila Vidács and Rolland Vida

The EW conference is committed to high publication ethics standards through a rigorous single-blind peer-review process. Submitted manuscripts must be original and not published or under consideration elsewhere. They must not infringe any copyright or third party right. Proceedings of EW 2015 will be available on IEEEXplore and Scopus (approval pending). Authors of selected papers will be invited to submit a journal extended version for a special issue of Wiley Transactions on Emerging Telecommunications Technologies.

Important dates
paper submission: February 2, 2015
notification of acceptance: March 22, 2015
camera ready due: March 29, 2015

Guidelines for our Authors

Format of the manuscripts

Original manuscripts and final versions of papers should be submitted in IEEE format according to the formatting instructions available on
http://www.ieee.org/publications_standards/publications/authors/authors_journals.html#sect2,
"Template and Instructions on How to Create Your Paper".

Length of the manuscripts

The length of papers in the aforementioned format should be 6-8 journal pages.
Wherever appropriate, include 1-2 figures or tables per journal page.

Paper structure

Papers should follow the standard structure, consisting of *Introduction* (the part of paper numbered by "1"), and *Conclusion* (the last numbered part) and several *Sections* in between.
The Introduction should introduce the topic, tell why the subject of the paper is important, summarize the state of the art with references to existing works and underline the main innovative results of the paper. The Introduction should conclude with outlining the structure of the paper.

Accompanying parts

Papers should be accompanied by an *Abstract* and a few *index terms (Keywords)*. For the final version of accepted papers, please send the *short cvs* and *photos* of the authors as well.

Authors

In the title of the paper, authors are listed in the order given in the submitted manuscript. Their full affiliations and e-mail addresses will be given in a footnote on the first page as shown in the template. No degrees or other titles of the authors are given. Memberships of IEEE, HTE and other professional societies will be indicated so please supply this information. When submitting the manuscript, one of the authors should be indicated as corresponding author providing his/her postal address, fax number and telephone number for eventual correspondence and communication with the Editorial Board.

References

References should be listed at the end of the paper in the IEEE format, see below:

a) Last name of author or authors and first name or initials, or name of organization
b) Title of article in quotation marks
c) Title of periodical in full and set in italics
d) Volume, number, and, if available, part
e) First and last pages of article
f) Date of issue

[11] Boggs, S.A. and Fujimoto, N., "Techniques and instrumentation for measurement of transients in gas-insulated switchgear," IEEE Transactions on Electrical Installation, vol. ET-19, no. 2, pp.87–92, April 1984.

Format of a book reference:

[26] Peck, R.B., Hanson, W.E., and Thornburn, T.H., Foundation Engineering, 2nd ed. New York: McGraw-Hill, 1972, pp.230–292.

All references should be referred by the corresponding numbers in the text.

Figures

Figures should be black-and-white, clear, and drawn by the authors. Do not use figures or pictures downloaded from the Internet. Figures and pictures should be submitted also as separate files. Captions are obligatory. Within the text, references should be made by figure numbers, e.g. "see Fig. 2."
When using figures from other printed materials, exact references and note on copyright should be included. Obtaining the copyright is the responsibility of authors.

Contact address

Authors are requested to send their manuscripts via electronic mail or on an electronic medium such as a CD by mail to the Editor-in-Chief:

Csaba A. Szabo
Department of Networked Systems and Services
Budapest University of Technology and Economics
2 Magyar Tudosok krt.
Budapest, 1117 Hungary
szabo@hit.bme.hu

Eurocon 15

http://eurocon2015.usal.es

16th International Conference on Computer as a Tool

University of Salamanca (Spain), 8th -11th September of 2015

DATES

Paper Submission Deadline	4th May 2015
Notification of Acceptance	25th May 2015
Camera-Ready Copy Due	15th June 2015
Early Registration	22nd June 2015
Conference Presentation	8th -11th September 2015

COMMITTEE

General Chair

Emilio Corchado – University of Salamanca (Spain)

Honorary Chairs

Alfonso Fernández Mañueco - Mayor of Salamanca (Spain)

Costas M. Stasopoulos - IEEE Region 8 Director-Elect.

Technical Programme Committee

TPC Co-chairs

Carl James Debono - University of Malta (Malta)

Magdalena Salazar - Universidad Carlos III de Madrid (Spain)

Manuel Castro - UNED (Spain)

Manuel Graña - University of País Vasco/EHU (Spain)

Publicity Co-chairs

Matej Zajc - University of Ljubljana (Slovenia)

Marios Antoniou - CYTA (Cyprus)

Ali El-Mousa - University of Jordan (Jordan)

Shaun Kaplan - CapeSoft (South Africa)

Track Chairs

Jan Haase – Track: Smart Cities

Athanasios Kakarountas – Track: Circuits and Systems for Signal Processing

Organizing Committee

Jesús Fraile - Universidad Politécnica de Madrid (Spain)

Ana Collado - CTTC (Spain)

Victorino Franco - University of Seville (Spain)

Alfonso Lago - University of Vigo (Spain)

Mislav Grgic - University of Zagreb (Croatia)

Igor Kuzle - University of Zagreb (Croatia)

Peter Nagy - HTE (Hungary)

Álvaro Herrero - University of Burgos (Spain)

Bruno Baruque - University of Burgos (Spain)

Héctor Quintián - University of Salamanca (Spain)

SUBMISSION

Manuscripts must be written in English. All papers must comply with the format of the IEEE Manuscript Templates for Conference Proceedings and each paper should not exceed 6 pages.

All submissions will be refereed by experts in the field based on originality, significance, quality and clarity. All contributions must be original, must not have been published elsewhere and must not be intended to be published elsewhere (conference or journal) during the review period. Accepted papers will be included in EUROCON 2015 Proceedings, which will be published by IEEE and submitted to IEEE Xplore. All accepted papers will be considered for extension for possible publication in journal special issues dedicated to this conference.

All papers must be initially submitted in electronic PDF format using the online paper submission system: https://easychair.org/conferences/?conf=eurocon2015

SCOPE

The IEEE Region 8 EuroCon 2015 Conference is a premier forum for the exchange of ideas, open and direct discussion on the development of the Circuits and Systems, Multimedia, Information and Communication Technology and Energy and Power Systems. It has achieved a considerable success during the past 15 editions covering majority of the fields in the area of electrical engineering.

EUROCON 2015 is organized by the IEEE Spanish Section and IEEE Region 8.

KEYNOTE SPEAKERS

Prof. Marios M. Polycarpou - University of Cyprus (Cyprus)

Prof. Francisco Herrera - University of Granada (Spain)

CONFERENCE TOPICS

EUROCON is a multidisciplinary conference, focusing on emerging technologies, and covers a variety of topics in the fields of:

- Information and Communications Technologies
- Circuits and Systems for Signal Processing
- Power and Energy.

Further information about topics: http://eurocon2015.usal.es/Scope

CALL FOR SPECIAL SESSIONS

In addition to regular sessions, participants are encouraged to organize special sessions on specialized topics. Each special session should have at least 4 or 5 quality papers.

Special session organizers will solicit submissions; conduct reviews (at least 6 international members of the review team) jointly with the EUROCON 2015 PC and in the same way recommend accept/reject decisions on the submitted papers.

Special Session submission deadline: 30th March, 2015

Submission of Special Sessions are welcome!

Up to now:

- Enhanced Interference Management in Advanced Mobile Communications Networks for Energy Saving
- Software Defined Radio for Education
- Applications of Computational Intelligence in Bioinformatics and Biomedical Engineering
- Soft Computing Models in Industrial and Environmental Applications by SMC

SOCIAL ACTIVITIES

Guided visit to Salamanca and Tapas Night

Guided visit to Ciudad Rodrigo followed by:

- Bullfighting Show (No pain or blood involved for the brave animals)
- Gala Dinner

CONTACT

Emilio Corchado
IEEE Spain Section Chair
Department of Computer Science and Automatic Control
University of Salamanca, Salamanca, Spain
e-mail: eurocon@usal.es

SCIENTIFIC ASSOCIATION FOR INFOCOMMUNICATIONS

Who we are

Founded in 1949, the Scientific Association for Infocommunications (formerly known as Scientific Society for Telecommunications) is a voluntary and autonomous professional society of engineers and economists, researchers and businessmen, managers and educational, regulatory and other professionals working in the fields of telecommunications, broadcasting, electronics, information and media technologies in Hungary.

Besides its 1000 individual members, the Scientific Association for Infocommunications (in Hungarian: Híradástechnikai Tudományos Egyesület, HTE) has more than 60 corporate members as well. Among them there are large companies and small-and-medium enterprises with industrial, trade, service-providing, research and development activities, as well as educational institutions and research centers.

HTE is a Sister Society of the Institute of Electrical and Electronics Engineers, Inc. (IEEE) and the IEEE Communications Society.

What we do

HTE has a broad range of activities that aim to promote the convergence of information and communication technologies and the deployment of synergic applications and services, to broaden the knowledge and skills of our members, to facilitate the exchange of ideas and experiences, as well as to integrate and harmonize the professional opinions and standpoints derived from various group interests and market dynamics.

To achieve these goals, we…

- contribute to the analysis of technical, economic, and social questions related to our field of competence, and forward the synthesized opinion of our experts to scientific, legislative, industrial and educational organizations and institutions;
- follow the national and international trends and results related to our field of competence, foster the professional and business relations between foreign and Hungarian companies and institutes;
- organize an extensive range of lectures, seminars, debates, conferences, exhibitions, company presentations, and club events in order to transfer and deploy scientific, technical and economic knowledge and skills;
- promote professional secondary and higher education and take active part in the development of professional education, teaching and training;
- establish and maintain relations with other domestic and foreign fellow associations, IEEE sister societies;
- award prizes for outstanding scientific, educational, managerial, commercial and/or societal activities and achievements in the fields of infocommunication.

Contact information

President: **DR. GÁBOR MAGYAR** • *elnok@hte.hu*

Secretary-General: **DR. ISTVÁN BARTOLITS** • *bartolits@nmhh.hu*

Operations Director: **PÉTER NAGY** • *nagy.peter@hte.hu*

International Affairs: **ROLLAND VIDA, PhD** • *vida@tmit.bme.hu*

Address: H-1051 Budapest, Bajcsy-Zsilinszky str. 12, HUNGARY, Room: 502

Phone: +36 1 353 1027, Fax: +36 1 353 0451

E-mail: *info@hte.hu*, Web: *www.hte.hu*