Infocommunications Journal

A PUBLICATION OF THE SCIENTIFIC ASSOCIATION FOR INFOCOMMUNICATIONS (HTE)

June 2015 Volume VII Number 2 ISSN 2061-2079

PAPERS FROM OPEN CALL

Attribute-Based Encryption Optimized for Cloud Computing ..... *M. Horváth* 1

A Novel Framework for Semantic Discovery of Web Services Using Integrated Semantic Model ..... *S. Sharma, J. Singh Lather, and M. Dave* 10

Fairness in Kademlia with Random Node Joins ..... *Z. Novák, Z. Pap* 19

The Use of Software-Defined Radio Systems in Multilateral Navigation Radio Systems ..... *G. M. Mashkov, E.G. Borisov, A.G. Vladyko, and A.I. Gomonova* 26

FROM IEEE COMMUNICATIONS MAGAZINE

Opportunities in Mobile Crowd Sensing ..... *H. Ma, D. Zhao, and P. Yuan* 32

CALL FOR PAPERS

Special Issue on Advanced Wireless and Mobile Technologies and Services 9

Special Issue on Smart Cities: Crowdsourcing and M2M Communication for a Connected Society 39

ADDITIONAL

Guidelines for our Authors 40

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
Subscription rates for foreign subscribers:

Articles can be sent also to the following address:
Budapest University of T echnology 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
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

Attribute-Based Encryption Optimized for Cloud Computing

Máté Horváth

***Abstract*—In this work, we aim to make attribute-based encryption (ABE) more suitable for access control to data stored in the cloud. For this purpose, we concentrate on giving to the encryptor full control over the access rights, providing feasible key management even in case of multiple independent authorities, and enabling viable user revocation, which is essential in practice. Our main result is an extension of the decentralized CP-ABE scheme of Lewko and Waters [8] with identity-based user revocation. Our revocation system is made feasible by removing the computational burden of a revocation event from the cloud service provider, at the expense of some permanent, yet acceptable overhead of the encryption and decryption algorithms run by the users. Thus, the computation overhead is distributed over a potentially large number of users, instead of putting it on a single party (e.g., a proxy server), which would easily lead to a performance bottleneck. The formal security proof of our scheme is given in the generic bilinear group and random oracle models.**

***Index Terms*—storage in clouds, access control, attribute-based encryption, user revocation, multi-authority.**

I. INTRODUCTION

Recent trends show a shift from using companies' own data centres to outsourcing data storage to cloud service providers. Besides cost savings, flexibility is the main driving force for outsourcing data storage, although in the other hand it raises the issue of security, which leads us to the necessity of encryption. Traditional cryptosystems were designed to confidentially encode data to a target recipient (e.g. from Alice to Bob) and this seems to restrict the range of opportunities and flexibility offered by the cloud environment. Imagine the following scenario: some companies are cooperating on a cryptography project and from each, employees are working together on some tasks. Suppose that Alice wants to share some data of a subtask with those who are working on it, and with the managers of the project from the different companies. We see that encrypting this data with traditional techniques, causes that recipients must be determined formerly, moreover either they have to share the same private key or several encrypted versions (with different keys) must be stored. These undermine the possible security, efficiency and the flexibility which the cloud should provide.

Attribute-based encryption (ABE) proposed by Sahai and Waters [16] is intended for one-to-many encryption in which ciphertexts are encrypted for those who are able to fulfil certain requirements. The most suitable variant for fine-grained access control in the cloud is called ciphertext-policy (CP) ABE, in which ciphertexts are associated with access policies, determined by the encryptor and attributes describe the user, accordingly attributes are embedded in the users' secret keys. A ciphertext can be decrypted by someone if and only if, his attributes satisfy the access structure given in the ciphertext, thus data sharing is possible without prior knowledge of who will be the receiver preserving the flexibility of the cloud even after encryption.

Returning to the previous example, using CP-ABE Alice can encrypt with an access policy expressed by the following Boolean formula: "CRYPTOPROJECT" AND ("SUBTASK Y" OR "MANAGER"). Uploading the ciphertext to the cloud, it can be easily accessed by the employees of each company, but the data can be recovered only by those who own a set of attributes in their secret keys which satisfies the access policy (e.g. "CRYPTOPROJECT", "SUBTASK Y").

In spite of the promising properties, the adoption of CP-ABE requires further refinement. A crucial property of ABE systems is that they resist collusion attacks. In most cases (e.g. [2], [19]) it is achieved by binding together the attribute secret keys of a specific user with a random number so that only those attributes can be used for decryption which contains the same random value as the others. As a result private keys must be issued by one central authority (CA) that would need to be in a position to verify all the attributes or credentials it issued for each user in the system. However even our example shows that attributes or credentials issued across different trust domains are essential and these have to be verified inside the different organisations (e.g. "MANAGER" attribute). To overcome this problem, we are going to make use of the results of Lewko and Waters [8] about decentralising CP-ABE.

The other relevant issue is user revocation. In everyday use, a tool for changing a user's rights is essential as unexpected events may occur and affect these. An occasion when someone has to be revoked can be dismissal or the revealing of malicious activity. Revocation is especially hard problem in ABE, since different users may hold the same functional secret keys related with the same attribute set (aside from randomization). We emphasise that user revocation is applied in *exceptional cases* like the above-mentioned, as all other cases can be handled simpler, with the proper use of attributes (e.g. an attribute can include its planned validity like "CRYPTOPROJECT2015").

Simultaneous solutions for these two problems could enhance flexible access control in cloud-based secure data storage. Such "optimized" CP-ABE could hide symmetric keys, which are used to efficiently encode large amounts of data, and reveal them only for authorized users, who can be identified through expressive access policies (for details see Figure 1).

M. Horváth works in the Laboratory of Cryptography and System Security (CrySyS Lab) at Budapest University of Technology and Economics, Department of Networked Systems and Services, Magyar tudósok krt. 2, 1117 Budapest, Hungary. E-mail: mhorvath@crysys.hu

Manuscript received February 12, 2015; revised May 25, 2015

Related Work.: The concept of ABE was first proposed by Sahai and Waters [16] as a generalization of identity-based encryption. Bethencourt et al. [2] worked out the first ciphertext-policy ABE scheme in which the encryptor must decide who should or should not have access to the data that she encrypts (ciphertexts are associated with policies, and users' keys are associated with sets of descriptive attributes). This concept was further improved by Waters in [19].

The problem of building ABE systems with multiple authorities was first considered by Chase [5] with a solution that introduced the concept of using a global identifier (GID) for tying users' keys together. Her system relied on a central authority and was limited to expressing a strict AND policy over a pre-determined set of authorities. Decentralized ABE of Lewko and Waters [8] does not require any central authority and any party can become an authority while there is no requirement for any global coordination (different authorities need not even be aware of each other) other than the creation of an initial set of common reference parameters. With this it avoids placing absolute trust in a single designated entity, which must remain active and uncorrupted throughout the lifetime of the system. Several other multi-authority schemes (e.g. [14], [18]) were shaped to the needs of cloud computing, although these lack for efficient user revocation.

Attribute revocation with the help of expiring attributes was proposed by Bethencourt et al. [2]. For single authority schemes Sahai et al. [15] introduced methods for secure delegation of tasks to third parties and user revocation through piecewise key generation. Ruj et al. [14], Wang et al. [18] and Yang et al. [20] show traditional attribute revocation (in multi-authority setting) causing serious computational overhead, because of the need for key re-generation and ciphertext re-encryption. A different approach is identity-based revocation, two types of which were applied to the scheme of Waters [19]. Liang et al. [11] gives the right of controlling the revoked set to a "system manager" while Li et al. [10], follow [7], from the field of broadcast encryption systems and give the revocation right directly to the encryptor. This later was further developed by Li et al. [9] achieving full security with the help of dual system encryption. For this approach, but in key-policy ABE, Qian and Dong [13] showed fully secure solution.

To the best of our knowledge no multi-authority system is integrated with identity-based user revocation and our work is the first in this direction.

Contribution.: Based on [8] and [7] we propose a scheme that adds identity-based user revocation feature to distributed CP-ABE. With this extension, we achieve a scheme with multiple, independent attribute authorities, in which revocation of specific users (e.g. with ID_i) from the system with all of their attributes is possible without updates of attribute public and secret keys (neither periodically, nor after revocation event). We avoid re-encryption of all ciphertexts the access structures of which contain a subset of attributes of the revoked user. The revocation right can be given directly to the encryptor, just like the right to define the access structure which fits to the cloud computing scenario.

A preliminary version of this work appeared in [6]. In this paper, we make substantial extensions to the contributions presented in [6], including a new, detailed security analysis of our proposed scheme, with a rigorous proof in the generic bilinear group and random oracle models, as well as proposal for an application approach in the cloud storage scenario and detailed explanations and reflections on related works.

Organization.: In Section II we introduce the later used theoretical background. In Section III the details of our scheme can be found together with efficiency and security analysis. Directions for further research are proposed in the last section.

II. BACKGROUND

We first briefly introduce bilinear maps, and provide the relevant background on access structures and secret sharing schemes. Then we give the algorithms of Ciphertext Policy Attribute-Based Encryption with identity-based user revocation.

A. Bilinear maps

We present the most important facts related to groups with efficiently computable bilinear maps.

Let $\mathbb{G}_0$ and $\mathbb{G}_1$ be two multiplicative cyclic groups of prime order p. Let g be a generator of $\mathbb{G}_0$ and e be a bilinear map (pairing), $e : \mathbb{G}_0 \times \mathbb{G}_0 \to \mathbb{G}_1$, with the following properties:

1) Bilinearity: $\forall u, v \in \mathbb{G}_1$ and $a, b \in \mathbb{Z}_p$, we have $e(u^a, v^b) = e(u, v)^{ab}$
2) Non-degeneracy: $e(g, g) \neq 1$.

We say that $\mathbb{G}_0$ is a bilinear group if the group operation in $\mathbb{G}_0$ and the bilinear map $e : \mathbb{G}_0 \times \mathbb{G}_0 \to \mathbb{G}_1$ are both efficiently computable. Notice that the map e is symmetric since $e(g^a, g^b) = e(g, g)^{ab} = e(g^b, g^a)$.

B. Access Structures and Secret Sharing

The requirements of decryption in an ABE scheme can be expressed using access structures (for formal definition see [1]), which determines all the authorised sets of attributes that allow decryption. Most ABE schemes (like ours) are restricted to *monotone access structures*, meaning that any superset of an authorized set is authorized as well. We note that (inefficiently) general access structures also can be realized using our techniques by having the not of each attribute as separate attribute.

To enforce the access structure, determined by the encryptor, we are going to make essential use of Linear Secret Sharing Schemes (LSSS). Here we adopt the definitions from those given in [1].

Definition 1 (Linear Secret Sharing Scheme [1]): A secret-sharing scheme Π over a set of attributes U is called linear (over $\mathbb{Z}_p$) if

1) the shares for each attribute form a vector over $\mathbb{Z}_p$,
2) there exists a matrix A with ℓ rows and n columns called the share-generating matrix for Π. For all $x = 1, \ldots, \ell$, the x^{th} row of A is labelled by an attribute $\rho(x)$, where ρ is a function from $\{1, \ldots, \ell\}$ to U. When we consider the column vector $v = (s; r_2, \ldots, r_n)$, where $s \in \mathbb{Z}_p$ is the secret to be shared, and $r_2, \ldots, r_n \in \mathbb{Z}_p$ are randomly chosen, then $Av = \lambda$ is the vector of ℓ shares of the

secret s according to Π. The share $(Av)_x = \lambda_x$ belongs to attribute $\rho(x)$.

In [1] it is shown that every linear secret sharing-scheme according to the above definition also enjoys the *linear reconstruction property*, defined as follows. Suppose that Π is an LSSS for the access structure $\mathbb{A}$. Let $S \in \mathbb{A}$ be any authorized set, and let $I \subset \{1, 2, \ldots, \ell\}$ be defined as $I = \{i | \rho(i) \in S\}$. Then, there exist constants $\{\omega_i \in \mathbb{Z}_p\}_{i \in I}$ such that, if $\{\lambda_i\}$ are valid shares of any secret s according to Π, then $\sum_{i \in I} \omega_i \lambda_i = s$. Furthermore, it is also shown in [1] that these constants $\{\omega_i\}$ can be found in time polynomial in the size of the share-generating matrix A and for unauthorized sets, no such $\{\omega_i\}$ constants exist.

We use the convention that $(1, 0, 0, \ldots, 0)$ is the "target" vector for any linear secret sharing scheme. For any satisfying set of rows I in A, we will have that the target vector is in the span of I, but for any unauthorized set, it is not.

Using standard techniques (see [8] - Appendix G) one can convert any monotonic boolean formula into an LSSS representation. An access tree of ℓ nodes will result in an LSSS matrix of ℓ rows.

C. Revocation Scheme for Multi-Authority CP-ABE

A multi-authority Ciphertext-Policy Attribute-Based Encryption system with identity-based user revocation is comprised of the following algorithms:

Global Setup$(\lambda) \rightarrow GP$
The global setup algorithm takes in the security parameter λ and outputs global parameters GP for the system.

Central Authority Setup$(GP) \rightarrow (SK^*, PK^*)$
The central authority (CA) runs this algorithm with GP as input to produce its own secret key and public key pair, SK^*, PK^*.

Identity KeyGen$(GP, RL, GID, SK^*) \rightarrow K^*_{GID}$
The central authority runs this algorithm upon a user request for identity secret key. It checks whether the request is valid and if yes (i.e. the user's global identifier, denoted by GID, is not part of the RL revocation list: $GID \notin RL$), generates K^*_{GID} using the global parameters and the secret key of the CA.

Authority Setup$(GP) \rightarrow (PK, SK)$
Each attribute authority runs the authority setup algorithm with GP as input to produce its own secret key and public key pair, SK, PK.

KeyGen$(GP, SK, GID, i) \rightarrow K_{i,GID}$
The attribute key generation algorithm takes in an identity GID, the global parameters, an attribute i belonging to some authority, and the secret key SK for this authority. It produces a key $K_{i,GID}$ for this attribute-identity pair.

Encrypt$(GP, \mathcal{M}, (A, \rho), \{PK\}, PK^*, RL) \rightarrow CT$
The encryption algorithm takes in a message $\mathcal{M}$, an access matrix (A, ρ), the set of public keys for relevant authorities, the public key of the central authority, the revoked user list and the global parameters. It outputs a ciphertext CT.

Decrypt$(GP, CT, (A, \rho), \{K_{i,GID}\}, K^*_{GID}, RL) \rightarrow \mathcal{M}$
The decryption algorithm takes in the global parameters, the revoked user list, the ciphertext, identity key and a collection of keys corresponding to attribute, identity pairs all with the same fixed identity GID. It outputs either the message $\mathcal{M}$ when the collection of attributes i satisfies the access matrix corresponding to the ciphertext. Otherwise, decryption fails.

III. OUR RESULTS

To build our model we will use the prime order group construction of Lewko and Waters [8], because of its favourable property of having independent attribute authorities. In order to achieve identity-based revocation we supplement the distributed system with a Central Authority. However it seems to contradict with the original aim of distributing the key generation right, this additional authority would generate only secret keys for global identifiers ($GID \in \mathbb{Z}_p$) of users and the attribute key generation remains distributed. Our Central Authority does not possess any information that alone would give advantage during decryption, in contrast to single authority schemes, where the authority is able to decrypt all ciphertexts. Regarding this, we can say that our system remains distributed, in spite of launching a Central Authority.

Approach to the Cloud Storage Scenario: We give a high-level description about a possible application of the algorithms that we proposed in Subsection II-C (for graphical depiction see Figure 1). Because of efficiency reasons it is practical to encrypt data using a symmetric cipher, always with fresh random number as key. Access control is achieved by encrypting the symmetric key using CP-ABE and attaching the encrypted key to the ciphertext that is stored by the cloud service provider (CSP). Decryption is possible for users, who can obtain the symmetric key, or with other words those, who possess the necessary attributes and were not revoked. Attribute Authorities are run locally on trusted servers of organisations, that are using the system, while the Central Authority is run by the CSP, which also maintains (archives, publishes) the RL revocation list, based on the revocation requests from authorised parties of the organisations. The ABE encryption always uses the fresh RL and ABE decryption is run with the RL at the encryption time of the ciphertext, which are obtained from the CSP. This approach automatically leads to lazy re-encryption of ciphertext, as fresh symmetric key and RL are used whenever data is edited.

a) Our Technique.: We face with the challenges of identity-based revocation. To realize the targeted features, we use some ideas from public key broadcast encryption systems [7]. A recent[1] work of Cao and Liu [4] points out an inherent drawback of the [7] scheme, namely that for malicious users it is worth to exchange their decryption keys in order to maximize their interests. However we utilize similar techniques as [7], our system is not vulnerable to this kind of misuse, because unlike in broadcast encryption, where having a non-revoked secret key is the only requirement for decryption, in ABE, users are also required to fulfil requirements related to their attributes. Thus such collusion could have

[1] [4] appeared on ePrint some months later than our work.

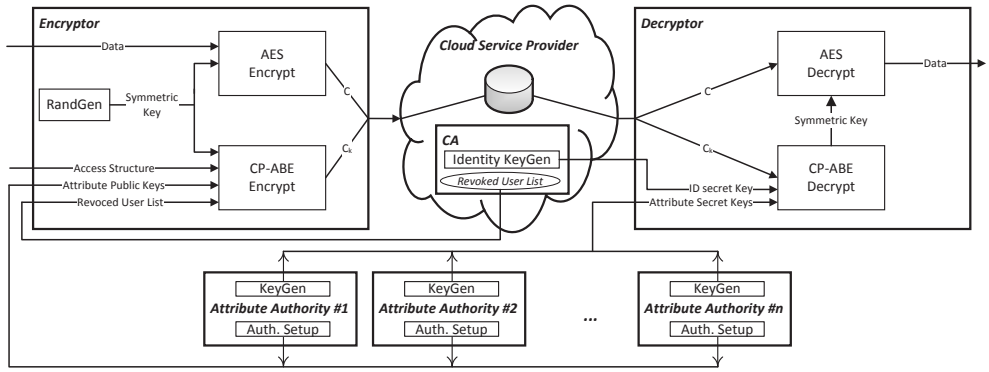


Figure 1. A possible usage of the proposed multi-authority CP-ABE scheme for access control in a cloud storage scenario.

only a restricted benefit[2] as the set of ciphertexts that can be decrypted is also restricted by the used attribute secret keys (which cannot be mixed between different users).[3]

We use secret sharing in the exponent. Suppose an encryption algorithm needs to create an encryption with a revocation set $RL = GID_1^*, \ldots, GID_r^*$ of r identities. The algorithm will create an exponent $s^* \in \mathbb{Z}_p$ and split it into r random shares $s_1, \ldots, s_r$ such that $\sum_{k=1}^{r} s_k = s^*$. It will then create a ciphertext such that any revoked user with GID_k^* will not be able to incorporate the k^{th} share and thus not decrypt the message.

This approach presents the following challenges. First, we need to make crucial that the decryptor needs to do the GID comparisons even if his attributes satisfy the access structure of the ciphertext. Second we need to make sure that a user with revoked identity GID_k^* cannot do anything useful with share k. Third, we need to worry about collusion attacks between multiple revoked users.

To address the first one we are going to take advantage of the technique of [8] that is used to prevent collusion attacks. Here the secret s, used for the encryption, is divided into shares, which are further blinded with shares of zero. This structure allows for the decryption algorithm to both reconstruct the main secret and to "unblind" it in parallel. If a user with a particular identifier GID satisfies the access structure, he can reconstruct s in the exponent by raising the group elements to the proper exponents. This operation will simultaneously reconstruct the share of 0 and thus the $e(H(GID), g)$ blinding terms will cancel out. When we would like to make this algorithm necessary, but not enough for decryption it is straightforward to spoil the "unblinding" of the secret by changing the shares of zero in the exponent to shares of an other random number, $s^* \in \mathbb{Z}_p$. Thus we can require an other computation, namely the comparison of the decryptor's and the revoked users' GIDs. If correspondence is found, the algorithm stops, otherwise reveals the blinding, enabling decryption.

The second challenge is addressed by the following method. A user with $GID \neq GID_k^*$ can obtain two linearly independent equations (in the exponent) involving the share s_k, which he will use to solve for the share s_k. However, if $GID = GID_k^*$, the obtained equations are going to be linearly dependent and the user will not be able to solve the system.

In the third case, the attack we need to worry about is where a user with GID_k^* processes ciphertext share l, while another user with GID_l^* processes share k, and then they combine their results. To prevent collusion, we use $H(GID)$ as the base of the identity secret key, such that in decryption each user recovers shares $s_k \cdot \log_g H(GID)$ in the exponent, disallowing the combination of shares from different users.

A. Our Construction

To make the following notions more understandable, in Table I we summarize the new keys and variables (compared to [8]) which we introduce in our construction. Based on the above principles, the proposed algorithms are the following:

Table I
THE SUMMARY OF OUR NEW NOTATIONS

Notation	Meaning	Role
PK^*	$\{g^a, g^{1/b}\}$	public key of the Central Authority
SK^*	$\{a, b\}$	secret key of the Central Authority
K^*_{GID}	$H(GID)^{(GID+a)b}$	global identity secret key of a user
$C^*_{1,k}$	$\left(g^a g^{GID_k^*}\right)^{-s_k}$	revoked user identification in CT
$C^*_{2,k}$	$g^{s_k/b}$	k^{th} secret share in the CT
RL	$\{GID_1^*, \ldots, GID_r^*\}$	list of r revoked users

Global Setup$(\lambda) \rightarrow GP$

In the global setup, a bilinear group $\mathbb{G}_0$ of prime order p is chosen. The global public parameters, GP, are p and a generator g of $\mathbb{G}_0$, and a function H mapping global identities $GID \in \mathbb{Z}_p$ to elements of $\mathbb{G}_0$ (this is modelled as a random oracle in the security proof).

[2]Of course, when users reveal their secret keys, we cannot hope for security in any encryption method, but assuming honest users, it is their interest to keep the secrets. As long as the attributes of (still non-revoked) colluding users do not cover all the access policies, our scheme will not reveal all ciphertexts for the malicious group.

[3]We also note that the flaw of [7]'s security proof, mentioned by [4] does not affect our results, as we use different proof technique.

Central Authority Setup$(GP) \rightarrow (SK^*, PK^*)$
The algorithm chooses random exponents $a, b \in \mathbb{Z}_p$, keeps them as secret key $SK^* = \{a, b\}$ and publishes $PK^* = \{g^a, g^{1/b}\}$.

Identity KeyGen$(GP, RL, GID, SK^*) \rightarrow K^*_{GID}$
Upon the request of a user it first checks whether the user is on the list of revoked users (RL) or it has been queried before, if yes refuses the request, otherwise computes $H(GID)$ and generates the global identity secret key:

$$K^*_{GID} = H(GID)^{(GID+a)b}.$$

Authority Setup$(GP) \rightarrow (PK, SK)$
For each attribute i belonging to the authority (these indices i are not reused between authorities), the authority chooses two random exponents $\alpha_i, y_i \in \mathbb{Z}_p$ and publishes $PK = \{e(g,g)^{\alpha_i}, g^{y_i}\ \forall i\}$ as its public key. It keeps $SK = \{\alpha_i, y_i\ \forall i\}$ as its secret key.

KeyGen$(GP, SK, GID, i) \rightarrow K_{i,GID}$
To create a key for a GID, for attribute i belonging to an authority, the authority computes:

$$K_{i,GID} = g^{\alpha_i} H(GID)^{y_i}$$

Encrypt$(GP, \mathcal{M}, (A, \rho), \{PK\}, PK^*, RL) \rightarrow CT$
The encryption algorithm takes in a message $\mathcal{M}$, an $n \times \ell$ access matrix A with ρ mapping its rows to attributes, the global parameters, the public keys of the relevant authorities, the user identity public key and the most recent list of revoked users.
It chooses random $s, s^* \in \mathbb{Z}_p$ and a random vector $v \in \mathbb{Z}_p^\ell$ with s as its first entry. Let λ_x denote $A_x \cdot v$, where A_x is row x of A. It also chooses a random vector $w \in \mathbb{Z}_p^\ell$ with s^* as its first entry. Let ω_x denote $A_x \cdot w$.
For each row A_x of A, it chooses a random $r_x \in \mathbb{Z}_p$ and supposed that the number of revoked users is $|RL| = r$ it chooses s_k such that $s^* = \sum_{k=1}^{r} s_k$. The CT ciphertext is computed as

$$C_0 = \mathcal{M} \cdot e(g,g)^s,$$
$$C_{1,x} = e(g,g)^{\lambda_x} e(g,g)^{\alpha_{\rho(x)} r_x},$$
$$C_{2,x} = g^{r_x}, \quad C_{3,x} = g^{y_{\rho(x)} r_x} g^{\omega_x},$$
$$C^*_{1,k} = \left(g^a g^{GID^*_k}\right)^{-s_k}, \quad C^*_{2,k} = g^{s_k/b}$$

for all $x = 1, \ldots, n$ and $k = 1, \ldots, r$.

Decrypt$(GP, CT, (A, \rho), \{K_{i,GID}\}, K^*_{GID}, RL) \rightarrow \mathcal{M}$
We assume the ciphertext is encrypted under an access matrix (A, ρ). If the decryptor is not on the list of revoked users (RL) and has the secret keys K^*_{GID} for his GID and $\{K_{i,GID}\}$ for a subset of rows A_x of A, such that $(1, 0, \ldots, 0)$ is in the span of these rows, then the decryptor proceeds as follows. First chooses constants $c_x \in \mathbb{Z}_p$ such that $\sum_x c_x A_x = (1, 0, \ldots, 0)$ and denoting $r = |RL|$ computes:

$$\frac{\mathscr{A}}{\mathscr{B}} = \frac{\prod_x \left(\frac{C_{1,x} \cdot e(H(GID), C_{3,x})}{e(K_{\rho(x),GID}, C_{2,x})}\right)^{c_x}}{\prod_{k=1}^{r} \left(e(K^*_{GID}, C^*_{2,k}) e(C^*_{1,k}, H(GID))\right)^{1/(GID - GID^*_k)}}$$

which equals to $e(g,g)^s$, so the message can be obtained as $\mathcal{M} = C_0 / e(g,g)^s$.

To see the soundness of the Decryption algorithm observe that after substituting the corresponding values we get the following:

$$\begin{aligned}\mathscr{A} &= \prod_x \left(e(g,g)^{\lambda_x + \omega_x \log_g H(GID)}\right)^{c_x} \\ &= e(g,g)^{\sum_x \lambda_x c_x} \cdot e(H(GID), g)^{\sum_x \omega_x c_x} \\ &= e(g,g)^{s + s^* \log_g H(GID)} \\ \mathscr{B} &= \prod_{k=1}^{r} \left(e(g,g)^{(GID - GID^*_k) s_k \log_g H(GID)}\right)^{1/(GID - GID^*_k)} \\ &= e(g,g)^{-\sum_{k=1}^{r} s_k \log_g H(GID)} = e(g,g)^{s^* \log_g H(GID)}\end{aligned}$$

Remark 1. We note that an almost equivalent result can be achieved, with some different modifications on the decentralized scheme (splitting $C_{1,x}$ into two parts, using $e(g,g)^{\beta s}$ for encryption, where β is the secret of the CA, and publishing g^s) and fitting these to the method of [10]. However in this way additional modifications are still needed to prevent the CA from being able to decrypt any ciphertext by computing $e(g^\beta, g^s)$.

Remark 2. Supposing that we have a honest but curious CSP, which does not collude with the users, it is also possible to achieve indirect revocation (similarly to [11], [15]), with simple modifications on our scheme. With other words, the CSP could fully supervise user revocation based on the revocation requests from parties, authorised for this. We only need to modify the Encrypt algorithm to compute $C, C_0, C_{1,x}, C_{2,x}$ as originally and $C'_{3,x} = g^{y_{\rho(x)} r_x} \quad \forall x = 1, \ldots, n$. These values would form CT' that is sent to the CSP, where the collusion resistant CT with the revocation information is computed and published. CT has the same form as earlier, the only difference is that the blinding vector w is chosen by the CSP, so $\omega_x, C^*_{1,k}, C^*_{2,k}$ (as previously) and $C_{3,x} = C'_{3,x} \cdot g^{\omega_x}$ are computed also by the CSP. The main advantage of this approach is that immediate and efficient (partial) re-encryption can be achieved as only $w, s_k, \omega_x, C^*_{1,k}, C^*_{2,k}$ and $C_{3,x}$ need to be recomputed after a revocation event.

Remark 3. Alternatively, it is also possible to give revocation right directly to the encryptor by simply publishing a user list instead of RL. In this case RL would be defined by the user, separately for each ciphertext, and attached to CT.

B. Efficiency

Traditional, attribute-based user revocation (e.g. [14], [18], [20]) affects attributes, thus the revocation of a user may cause the update of all the users' attribute secret keys who had common attribute with the revoked user (a general attribute can affect big proportion of the users) and the re-encryption of all ciphertext the access structure of which contain any of the revoked user's attributes (most of these could not be decrypted by the revoked user).

In our scheme, a revocation event does not have any effect on the attributes as it is based on identity. Although it is a trade-off and in the other hand there is some computational

overhead on the encryption and decryption algorithms. In this way the necessary extra computation of authorities is reduced and distributed between the largest set of parties, the users, preventing a possible performance bottleneck of the system. At the same time the extra communication is also reduced to the publication of the revoked user list. Our revocation scheme has the following costs.

The ciphertext has $2r$ additional elements, if the number of revoked users is r. For the computation of these values $3r$ exponentiations and r multiplications are needed in $\mathbb{G}_0$. Alternatively, the revoked user list may contain $g^a g^{GID_i^*}$ instead of the global identifiers. In this case the encryptor need to do only $2r$ additional exponentiations in $\mathbb{G}_0$, compared with the scheme of [8], to compute the ciphertext. The overhead of the decryption algorithm is $2r$ pairing operations, r multiplications and exponentiations in group $\mathbb{G}_1$.

Note that, as in all model that uses LSSS to express the access structure, the access matrix and the mapping ρ must be part of the ciphertext, increasing its length. However, it is possible to reduce this length by attaching only a formatted Boolean formula instead and compute the necessary components of LSSS more efficiently, using the algorithm of Liu and Cao in [12].

C. Security

Before giving the formal proof, we point out that from the point of view of a user, whose attributes have never satisfied the access structure defined in the ciphertext, our construction is at least as secure as the one by [8], because the computation of $\mathscr{A}$ is equivalent to the decryption computation given there. However in our case, it is not enough to obtain the message. Changing the first entry of the blinding vector w from zero to a random number (as we did), causes that the blinding will not cancel out from $\mathscr{A}$, but we need to compute $\mathscr{B}$ which can divide it out. $\mathscr{B}$ can be computed with any GID different from any GID_k^* of the revocation list and we ensure that the decryptor must use the same GID both in $\mathscr{A}$ and $\mathscr{B}$ by using $H(GID)$ in both the identity and attribute secret keys.

1) Security Model: We now define (chosen plaintext) security of multi-authority CP-ABE system with identity-based revocation. Security is defined through a *security game* between an attacker algorithm $\mathcal{A}$ and a challenger. We assume that adversaries can corrupt authorities only statically, but key queries are made adaptively. The definition reflects the scenario where all users in the revoked set RL get together and collude (this is because the adversary can get all of the private keys for the revoked set). Informally, $\mathcal{A}$ can determine a set of corrupted attribute authorities, ask for any identity and attribute keys and specify messages, on which it will be challenged using the revocation list and access matrix of its choice. The only (natural) restriction in the above choices is that $\mathcal{A}$ cannot ask for a set of keys that allow decryption, in combination with any keys that can be obtained from corrupt authorities in case of a non revoked GID_k. In case of revoked identities we can be less restrictive: corrupted attributes alone cannot satisfy the access policy, but it it might be satisfied together with attributes from honest authorities. $\mathcal{A}$ wins the game if it respects the rules and can decide which of its challenge messages were encrypted by the challenger. The formal security game consists of the following rounds:

Setup. The challenger runs the Global Setup algorithm to obtain the global public parameters GP. $\mathcal{A}$ specifies a set $AA' \subseteq AA$ of corrupt attribute authorities and uses the Authority Setup to obtain public and private keys. For honest authorities in $AA \setminus AA'$ and for the Central Authority, the challenger obtains the corresponding keys by running the Authority Setup and Central Authority Setup algorithms, and gives the public keys to the attacker.

Key Query Phase. $\mathcal{A}$ adaptively issues private key queries for identities GID_k (which denotes the k^{th} GID query). The challenger gives $\mathcal{A}$ the corresponding identity keys $K^*_{GID_k}$ by running the Identity KeyGen algorithm. Let UL denote the set of all queried GID_k. $\mathcal{A}$ also makes attribute key queries by submitting pairs of (i, GID_k) to the challenger, where i is an attribute belonging to a good authority. The challenger responds by giving the attacker the corresponding key, K_{i,GID_k}.

Challenge. The attacker gives the challenger two messages M_0, M_1, a set $RL \subseteq UL$ of revoked identities and an access matrix (A, ρ).
RL and A must satisfy the following constraints. Let V denote the subset of rows of A labelled by attributes controlled by corrupt authorities. For each identity $GID_k \in UL$, let V_{GID_k} denote the subset of rows of A labelled by attributes i for which the attacker has queried (i, GID_k). For each $GID_k \in UL \setminus RL$, we require that the subspace spanned by $V \cup V_{GID_k}$ must not include $(1, 0, \ldots, 0)$ while for $GID_k \in RL$, it is allowed and we only require that the subspace spanned by V must not include $(1, 0, \ldots, 0)$.
The attacker must also give the challenger the public keys for any corrupt authorities whose attributes appear in the labelling ρ.
The challenger flips a random coin $\beta \in (0, 1)$ and sends the attacker an encryption of M_β under access matrix (A, ρ) with the revoked set RL.

Key Query Phase 2. The attacker may submit additional attribute key queries (i, GID_k), as long as they do not violate the constraint on the challenge revocation list RL and matrix (A, ρ).

Guess. $\mathcal{A}$ must submit a guess β' for β. The attacker wins if $\beta' = \beta$. The attacker's advantage in this game is defined to be $\mathbb{P}(\beta' = \beta) - \frac{1}{2}$.

Definition 2: We say that a multi-authority CP-ABE system with identity-based revocation is (chosen-plaintext) secure (against static corruption of attribute authorities) if, for all revocations sets RL of size polynomial in the security parameter, all polynomial time adversary has at most a negligible advantage in the above defined security game.

2) Security Analysis: We are going to prove the security of our construction in the generic bilinear group model previously used in [2], [3], [8], modelling H as a random oracle. Security in this model assures us that an adversary cannot break the scheme with only black-box access to the group operations and H. Intuitively, this means that if there are any vulnerabilities in

our construction, then these must exploit specific mathematical properties of elliptic curve groups or cryptographic hash functions used when instantiating the scheme.

Theorem 1: For any adversary $\mathcal{A}$, let q be a bound on the total number of group elements it receives from queries it makes to the group oracles and from its interaction with the security game, described in III-C1. The above described construction is secure according to Definition 2 in the generic bilinear group and random oracle models. The advantage of $\mathcal{A}$ is $\mathcal{O}(q^2/p)$.

In our proof we are going to use the following strategy. First we identify events that occur only with negligible probability, namely that the attacker is able to guess certain values successfully and that the oracle returns the same value for different queries. Assuming that these do not happen, we examine the (exponent) values which the attacker can obtain during the game. We show that $\mathcal{A}$ can recognise the challenge ciphertext only if is has used $GID_K \notin RL$ with a satisfying attribute set or has broken the rules of the game.

Proof:

We describe the generic bilinear model as in [3]. We let ψ_0 and ψ_1 be two random encodings of the additive group $\mathbb{Z}_p$. More specifically, each of ψ_0, ψ_1 is an injective map from $\mathbb{Z}_p$ to $\{0,1\}^m$, for $m > 3\log(p)$. We define the groups $\mathbb{G}_0 = \{\psi_0(x) : x \in \mathbb{Z}_p\}$ and $\mathbb{G}_1 = \{\psi_1(x) : x \in \mathbb{Z}_p\}$. We assume to have access to oracles which compute the induced group operations in $\mathbb{G}_0$ and $\mathbb{G}_1$ and an oracle which computes a non-degenerate bilinear map $e : \mathbb{G}_0 \times \mathbb{G}_0 \to \mathbb{G}_1$. We refer to $\mathbb{G}_0$ as a generic bilinear group. To simplify our notations let g denote $\psi_0(1)$, g^x denote $\psi_0(x)$, $e(g,g)$ denote $\psi_1(1)$, and $e(g,g)^y$ denote $\psi_1(y)$.

The challenger and the attacker play the security game (described in III-C1) and compute each value with respect to the generic bilinear group and random oracle models (i.e. send queries to the group oracle that responds with randomly assigned values). When $\mathcal{A}$ requests e.g. $H(GID_k)$ for some GID_k for the first time, the challenger chooses a random value $h_{GID_k} \in \mathbb{Z}_p$, queries the group oracle for $g^{h_{GID_k}}$, and gives this value to the attacker as $H(GID_k)$. It stores this value so that it can reply consistently to any subsequent requests for $H(GID_k)$.

We are going to show that in order to determine $\beta \in \{0,1\}$, $\mathcal{A}$ has to be able to compute $e(g,g)^{s^* h_{GID_k}}$ for any $k = 1,\ldots,r$, which is possible only with negligible probability without breaking the rules of the game.

We can assume that each of the attacker's queries to the group oracles either have input values that were given to $\mathcal{A}$ during the security game or were received from the oracles in response to previous queries. This is because of the fact that both ψ_0 and ψ_1 are random injective maps from $\mathbb{Z}_p$ into a set of at least p^3 elements, so the probability of the attacker being able to guess an element in the image of ψ_0, ψ_1 which it has not previously obtained is negligible.

Under this condition, we can think of each of the attacker's queries as a multi-variate expressions[4] in the variables $y_i, \alpha_i, \lambda_x, r_x, \omega_x, h_{GID_k}, a, b, s_k$, where i ranges over the attributes controlled by uncorrupted authorities, x ranges over the rows of the challenge access matrix, k ranges over the revoked identities. (We can also think of λ, ω_x as linear combinations of the variables $s, v_2, \ldots, v_\ell$ and $s^*, w_2, \ldots, w_\ell$.)

Furthermore we also assume that for each pair of different queries (corresponding to different polynomials), $\mathcal{A}$ receives different answers from the oracle. Since the maximum degree of polynomials is 8 (see the possible polynomials later), using the Schwartz-Zippel lemma [17] we get that the probability of a collusion is $\mathcal{O}(1/p)$ and a union bound shows that the probability of that any such collusion happens during the game is $\mathcal{O}(q^2/p)$, which is negligible. Now suppose that it does not happen.

In order to determine β, the attacker clearly needs to recover s. [8] showed that without a satisfying set of attributes an attacker cannot make a query of the form $c(s + 0 \cdot h_{GID_k})$ (where c is a constant) thus has only negligible advantage in distinguishing an encoded message from a random group element (when using their original scheme). This result implies that in our modified construction, the attacker cannot make a query of the form $c(s + s^* h_{GID_k})$ without a satisfying set of attributes (as the first element of the blinding vector w is changed to s^* from 0) which also shows - following their reasoning - that an expression in the form cs cannot be formed either. In our case, however, the possession of the necessary attributes are not enough to make a cs query, but $-c(s^* h_{GID_k})$ is also indispensable for this.

It can be seen that the case when $GID_k \in UL \setminus RL$ is equivalent to the original scheme of [8]. Consequently, from now on we can assume that all $GID_k \in RL$ and the challenge access policy is satisfied, thus simulating that all revoked users are colluding and prior to their revocation they were all able to decrypt. We will show that $\mathcal{A}$ cannot make a query of the form $-c(s^* h_{GID_k})$ and so not cs.

Based on the above assumptions the attacker can form queries which are linear combinations of

$$1, h_{GID_k}, y_i, \alpha_i + h_{GID_k} y_i, \lambda_x + \alpha_{\rho(x)} r_x, r_x, y_{\rho(x)} r_x + \omega_x, a, 1/b, b h_{GID_k}(GID_k^* + a), s_k(a + GID_k^*), s_k/b,$$

the product of any two of these and α_i. (Note that GID_k^* for all $k = 1,\ldots,r$ and α_i, y_i for attributes i controlled by corrupted authorities are constants, known by the attacker.) In these queries shares of s^* can appear in two different forms: as ω_x and s_k, so we investigate whether $\mathcal{A}$ can achieve the desired value from these or not.

1) In order to gain $s^* h_{GID_k}$ by utilizing ω_x, $\mathcal{A}$ must use the product $h_{GID_k} y_{\rho(x)} r_x + h_{GID_k} \omega_x$ for all rows of A, as these are the only terms which contain $h_{GID_k}\omega_x$ and thus which can lead to $s^* h_{GID_k}$. To cancel out $h_{GID_k} y_{\rho(x)} r_x$ the attacker should form this product, which is possible only if $y_{\rho(x)}$ or r_x are known constants, because these elements appear alone in the above list and besides those, $\mathcal{A}$ can only form the product of any two but not three. However if $y_{\rho(x)}$ or r_x are constants for all x, that contradicts with the rules of the security game, because in that case corrupted attributes alone would satisfy the access structure.

[4]These expressions can appear in the exponent of $e(g,g)$.

Table II
POSSIBLE RELEVANT QUERY TERMS

s_k/b
$s_k s_l/b^2$
$s_k a/b$
s_k/b^2
$s_k a h_{GID_l} + GID_l^* s_k h_{GID_l}$
$s_k h_{GID_l}/b$
$s_k s_l a/b + GID_k^* s_k s_l/b$
$s_k a + GID_k^* s_k$
$s_k s_l a^2 + GID_k^* GID_l^* s_k s_l + (GID_k^* + GID_l^*) s_k s_l a$
$s_k a^2 + GID_k^* s_k a$
$s_k a/b + GID_k^* s_k/b$
$s_k b h_{GID_l}(a^2 + (GID_k^* + GID_l^*)a + GID_k^* GID_l^*)$
$s_k a h_{GID_l} + GID_k^* s_k h_{GID_l}$

2) When trying to obtain $s^* h_{GID_k}$ using s_k, we can observe that in each possible query term, s_k appears as multiplier either in all monads or in none of them. Evidently, terms without s_k are useless (see Table II for the relevant terms) for the attackers purposes and terms containing the $s_k h_{GID_l}$ monad can be useful. As it can be seen in Table II, there are two types of terms which contain the necessary monad:

$$s_k a h_{GID_l} + GID_k^* s_k h_{GID_l}$$

and

$$s_k a h_{GID_l} + GID_l^* s_k h_{GID_l}.$$

Multiplying their subtraction by $c/(GID_k^* - GID_l^*)$ it is possible to gain $c \cdot s_k h_{GID_l}$, if $k \neq l$. In case of $k = l$ the two terms are equal, and $s_k a h_{GID_l}$ cannot be cancelled out, as no other terms contain this product. Nevertheless, according to our assumption that $GID_l^* \in RL$ for all $l = 1, \ldots, r$ there must be a $k = l$ as k runs over $1, \ldots, r$. We conclude that it is possible to gain $s_k h_{GID_l}$ for all k for any fixed l, if the attacker has used some $GID_l \notin RL$, which is again contradiction.

Hence, we have shown that under conditions that hold with all but $\mathcal{O}(q^2/p)$ probability, $\mathcal{A}$ cannot query $c(s^* h_{GID_k})$ (neither using ω_x nor s_k) therefore cannot get s without breaking the rules of the security game. It follows than, that the advantage of $\mathcal{A}$ is at most $\mathcal{O}(q^2/p)$. ■

IV. CONCLUSION

We proposed a scheme for efficient identity-based user revocation in multi-authority CP-ABE with several advantageous feature compared with attribute-based revocation. Our results fulfil specific needs of the cloud environment, thus optimizes ABE for real world usage. In the future, our work can be continued in several directions.

First and foremost, extensive comparisons are needed between the different revocation schemes proposed for CP-ABE to understand better their performance between different circumstances.

Securely forwarding the revocation related computations to the CSP (or even to the user), as we mentioned in Remark 2, could allow immediate banning of a user, disallowing the decryption of all previously (and later) encrypted ciphertexts. Steps in this direction, without assuming trusted CSP, would be useful.

The method of identity-based user revocation can be the foundation of a future method that allows non monotonic access structures in multi-authority setting. However our scheme cannot be applied directly for this purpose, it may be used to develop ideas in this field.

The security of our construction is proved in the generic bilinear group model, although we believe it would be possible to achieve full security by adapting the dual system encryption methodology, which was also used by Lewko and Waters [8] in their composite order group construction. This type of work would be interesting even if it resulted in a moderate loss of efficiency from our existing system.

ACKNOWLEDGMENTS

This work was started as a master thesis at Eötvös Loránd University, in the Security&Privacy program of EIT ICT Labs Masterschool. The author would like to thank all the help and valuable advice of Levente Buttyán from CrySyS Lab. He is also grateful to Viktória Villányi and Péter Ligeti for the useful discussions and to the anonymous reviewers of SOFSEM'15 and Infocommunications Journal for the valuable remarks.

REFERENCES

[1] Amos Beimel. *Secure schemes for secret sharing and key distribution*. PhD thesis, Israel Institute of Technology, Technion, Haifa, Israel, 1996.
[2] John Bethencourt, Amit Sahai, and Brent Waters. Ciphertext-policy attribute-based encryption. In *IEEE Symposium on Security and Privacy*, pages 321–334, 2007.
[3] Dan Boneh, Xavier Boyen, and Eu-Jin Goh. Hierarchical identity based encryption with constant size ciphertext. In *Advances in Cryptology–EUROCRYPT 2005*, pages 440–456. Springer, 2005.
[4] Zhengjun Cao and Lihua Liu. Analysis of Lewko-Sahai-Waters Revocation System. Cryptology ePrint Archive, Report 2014/937, 2014. http://eprint.iacr.org/.
[5] Melissa Chase. Multi-authority Attribute Based Encryption. In *Theory of Cryptography*, volume 4392 of *LNCS*, pages 515–534. Springer Berlin Heidelberg, 2007.
[6] Máté Horváth. Attribute-Based Encryption Optimized for Cloud Computing. In G.F. Italiano et al., editor, *SOFSEM 2015: Theory and Practice of Computer Science*, number 8939 in LNCS, pages 566–577. Springer, 2015.
[7] Allison Lewko, Amit Sahai, and Brent Waters. Revocation systems with very small private keys. In *IEEE Symposium on Security and Privacy*, pages 273–285, 2010.
[8] Allison Lewko and Brent Waters. Decentralizing attribute-based encryption. In *Advances in Cryptology–EUROCRYPT 2011*, pages 568–588. Springer, 2011.
[9] Qinyi Li, Hu Xiong, and Fengli Zhang. Broadcast revocation scheme in composite-order bilinear group and its application to attribute-based encryption. *International Journal of Security and Networks*, 8(1):1–12, 2013.
[10] Yang Li, Jianming Zhu, Xiuli Wang, Yanmei Chai, and Shuai Shao. Optimized Ciphertext-Policy Attribute-Based Encryption with Efficient Revocation. *International Journal of Security & Its Applications*, 7(6), 2013.
[11] Xiaohui Liang, Rongxing Lu, Xiaodong Lin, and Xuemin Sherman Shen. Ciphertext policy attribute based encryption with efficient revocation. *TechnicalReport, University of Waterloo*, 2010.
[12] Zhen Liu and Zhenfu Cao. On Efficiently Transferring the Linear Secret-Sharing Scheme Matrix in Ciphertext-Policy Attribute-Based Encryption. *IACR Cryptology ePrint Archive*, 2010:374, 2010.
[13] Jun-lei Qian and Xiao-lei Dong. Fully secure revocable attribute-based encryption. *Journal of Shanghai Jiaotong University (Science)*, 16:490–496, 2011.

[14] Sushmita Ruj, Amiya Nayak, and Ivan Stojmenovic. Dacc: Distributed access control in clouds. In *IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications*, pages 91–98, 2011.
[15] Amit Sahai, Hakan Seyalioglu, and Brent Waters. Dynamic credentials and ciphertext delegation for attribute-based encryption. In *Advances in Cryptology–CRYPTO 2012*, pages 199–217. Springer, 2012.
[16] Amit Sahai and Brent Waters. Fuzzy identity-based encryption. In *Advances in Cryptology–EUROCRYPT 2005*, pages 457–473. Springer, 2005.
[17] J. T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *J. ACM*, 27(4):701–717, October 1980.
[18] Guojun Wang, Qin Liu, Jie Wu, and Minyi Guo. Hierarchical attribute-based encryption and scalable user revocation for sharing data in cloud servers. *Computers & Security*, 30(5):320–331, 2011.
[19] Brent Waters. Ciphertext-policy attribute-based encryption: An expressive, efficient, and provably secure realization. In *Public Key Cryptography–PKC 2011*, pages 53–70. Springer, 2011.
[20] Kan Yang, Xiaohua Jia, Kui Ren, and Bo Zhang. DAC-MACS: Effective data access control for multi-authority cloud storage systems. In *INFOCOM, 2013 Proceedings IEEE*, pages 2895–2903, 2013.

Máté Horváth obtained his MSc diploma in computer science in the Security and Privacy program of EIT ICT Labs at the University of Trento (Italy) and Eötvös Loránd University (Hungary). His bachelor degree is in mathematics from the Technical University of Budapest. He has been doing research in the CrySyS Lab under the guidance of prof. Levente Buttyán since 2014.

A Novel Framework for Semantic Discovery of Web Services using Integrated Semantic Model

Shailja Sharma, Jagdeep Singh Lather, Mayank Dave

Abstract — **Semantic web technology plays a very critical role in the automatic web service discovery by assigning formal semantics to the service descriptions. Practically, it is not feasible to explicitly annotate the formal semantics to millions of existing services. Further, in user context, the request formation for services in semantic web is a complex process as it requires the user to be technically aware of the underlying technologies of the web services, discovery frameworks, description languages and implementation details. In this paper, we propose a semantic framework that enables Web service discovery based on the combination of semantic and syntax information contained in the service profiles. This novel approach for automatic discovery of Web services employs measures of semantic relatedness, Natural Language Processing techniques and information retrieval based statistical models to match a user request. Additionally, we present an efficient semantic matching technique to compute the intra service semantic similarity scores which further facilitates semantic ranking of services. The efficiency of the proposed approach has been demonstrated through experimental evaluations which clearly show that high degree of automation can be achieved with high precision.The results have been further authenticated by providing comparisons with other Information Retrieval based methods.**

Keywords: Semantic Web Service Discovery, Measures of Semantic Relatedness, Machine learning, Text Mining, OWL-S.

I. INTRODUCTION

Web services provide standard mechanisms for integration of distributed application over heterogeneous platforms [1]. Major challenge in the current service oriented architecture is to have automatic discovery process for the desirable services.

Manuscript received March 26, 2015, revised June 10, 2015.
Shailja Sharma is a Ph.D student in the Department of Computer Applications at National Institute of Technology, Kurukshetra, India (email: shailjakaushik@gmail.com)
Jagdeep Singh Lather is working as a Professor at Department of Electrical Engineering at National Institute of Technology, Kurukshetra (email: jslather@gmail.com)
Mayank Dave is working as Professor at Department of Computer Engineering at National Institute of Technology, Kurukshetra, India (email: mdave67@gmail.com)

The existing service standards like Extensible Markup Language (XML) [2], Simple Object Access Protocol (SOAP) [3], Web Service Description Language (WSDL) [4], Universal discovery and description Integration (UDDI) [5] are confined to syntax based matching scheme, where matching of service profiles is purely syntactic. This type of matching considers only those services whose syntactic descriptions exactly match the query keywords irrespective of semantic relatedness between the terms. Therefore, the outcome is either no results or a list of irrelevant services. Due to usage of different keywords, several services having semantically similar terminology are excluded from the result set, although they are potentially good candidates for the user's request. Therefore, human intervention is required in the service discovery process to filter the relevant services out of the resultant list. The numerous approaches for web service discovery range from Information Retrieval based similarity measures to semantic logic-based inference rules [6].

The current semantic approaches for web services discovery presume the services to be described in semantic description languages like OWL-S [7], WSMO [8], WSDL-S [9], SAWSDL [10], *etc*. Although, the semantic web technology is a promising way towards the realization of automatic discovery of web services but practical limitations of the semantic based approaches is that it is not possible to expect all service requestors and service providers to have same understanding of context and to use same ontological concepts.

Frameworks like OWL-S [7], WSMO [8] and WSDL-S [9] assume request as a web service and ask the user to express the request in a formal specified language which requires that web service users be friendly with these technologies and frameworks. Existing Web Services, whose descriptions are written in syntax based languages like WSDL, do not have explicitly associated semantics in their descriptions. Further, it is not feasible to have semantic tagged descriptions for all the new services. Thus, it can be concluded that annotating semantics to the existing services is time consuming, cumbersome and nearly impractical.

In order to address some of the cited limitations of existing approaches, this paper proposes an easy to use semantic framework for the discovery of services described in Web Service Description Language. The proposed approach generates concrete similarity score which facilitates the ranking of existing services. Our framework supports formulation of easy user request, keeping the user technical expertise independent of the technical knowhow of the services. Service provider's comments written in natural language as documentation have been utilized to match the user requirements in a better way. The proposed solution for semantic web service discovery integrates the syntactic as well as semantic information of a web service and tries to utilize the hidden semantics of the existing services. The rest of the paper is divided into following sections: section 2 provides the literature review and section 3 illustrates the proposed framework for describing new semantic matching algorithm; evaluations and results have been presented in section 4 and finally conclusion and future scope are covered in section 5.

II. RELATED WORK

It has been analyzed that most of the existing service discovery approaches adhered strictly to single service description languages and standards like DAML-S, OWL-S, WSMO, WSDL, SAWSDL, *etc*. These approaches vary on I/O matching vs. IOPE matching. DAML-S based approach [17] and OWL-S based approaches [12, 18, 19, 20, 21 and 22] and SAWSDL based approach [14] perform Input-Output (IO) matching on service profiles whereas approaches in [13, 23] perform IOPE matching. Research has been focused on addition of the semantics into the frameworks like WSDL using SAWSDL and WSDL-S. In [24], a WSDL-S based discovery technique over federated registries using the METEOR-S infrastructure has been proposed.

Some hybrid approaches [12, 13, 15 and 14] have also been proposed which considers semantic as well as syntactic description of the services. OWLS-MX [12] is an OWL-S based hybrid matchmaker that gives semantic as well as hybrid degree of matching.

In comparison to input and output (I/O) parameter matching by Klusch *et al*.[12], ITL based LARKS [15] perform IOPE matching. LARKS do not supports logical subsumes and hybrid nearest neighbor and has never been evaluated experimentally. Similar to OWL-MX, WSMO-MX [13] is also a hybrid matchmaker but it matches IOPE's of profiles instead of (I/O) and is based on WSMOframework. SAWSDL-iMatcher [14] annotates semantics to the existing profiles and supports user customizable matching strategies according to different application requirements. One of the major shortcomings of OWL-S based approaches is that it does not support mapping. Compared to OWL-S, WSMO is capable of modeling mediation for handling ontology heterogeneities in ontologies. Further, it has been observed that hybrid approaches based on logical reasoning are computationally expensive.

Among the various data mining based approaches, majority of them use classification and clustering to find semantic similarity between the services [25, 26, and 27]. SWSC [25] method uses Jaccard coefficient and hierarchical agglomerative clustering whereas in comparison Wen *et. al* [27] has modified the K-Mediod clustering mechanism to sort out the problem of web service discovery.

Batra and Bawa [26] propose a classification based approach that uses the Normalized semantic score MSR for semantic web service discovery.

In [28], García *et al.* use SPARQL-based repository filter for improving the semantic web service discovery.

III. SEMANTIC MATCHING APPROACH FOR WEB SERVICE DESCRIPTIONS

The proposed discovery framework aims to provide efficient discovery of services by combining the semantic and syntactic information from the service profiles. The complete discovery framework and algorithm is depicted in Figure 1 and Figure 2. Initially, the services are parsed using the text miner. This involves removal of markup(s), translation of upper case characters into lower case, punctuation and white space removal, followed by the stop word removal, *etc*. The textual documentation, arguments, operations and service name from the service profiles are extracted out. After preprocessing, the statistical weights are assigned to the terms using four different widely used weighing schemes of information retrieval i.e. TFIDF scheme, Binary scheme, Term Occurrence and Term Frequency. Further, a semantic relatedness matrix is generated using WordNet based measure of semantic relatedness. In next step, the syntactic vectors are transformed into the semantically enriched service vectors through semantic integration engine. The query is also transformed to the semantic query vector using the semantic integration engine.

The similarity match engine calculates the $semantic\_similarity$ between the semantic query vector $Query$ with the semantic description vectors of services present in the four kernels. Based on user specified threshold value $User\_Threshold$, a ranked list of relevant services having $semantic\_similarity$ greater than $User\_Threshold$ are returned to the user.

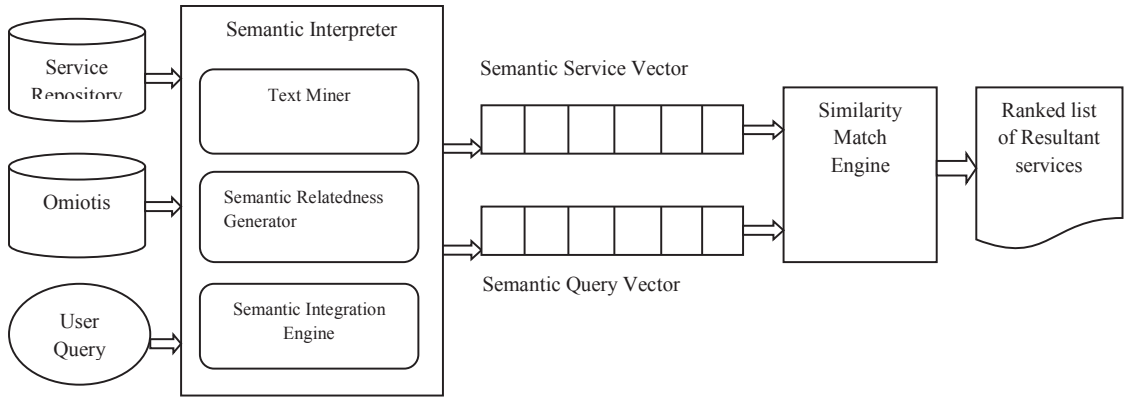


FIGURE 1: THE PROPOSED ARCHITECTURE

Input: Service set, $Query$: user query and
$User\_Threshold$:Threshold value;
Output: Resultant service set $results$;

1. Extractthe terms from the textual documentation, arguments, operations and service name from the service profiles of the m WSDL profiles through Text mining.
2. Assign weight to all the terms in the $Terms(X)$ and generate

a) $TF-IDF(m,X)$;
b) $Binary(m,X)$;
c) $Term\_Occurance(m,X)$;
d) $Term\_Frequency(m,X)$;

3. Calculate semantic relatedness matrix $Omiotis(X,n)$ for all the terms in the $Terms(X)$ vector.
4. Generate semantic kernel for all four weight representations by merging semantic information in them:

a) $Omiotis\_TF-IDF(m,n)$;
b) $Omiotis\_Binary(m,n)$;
c) $Omiotis\_Term\_Occurance(m,n)$;
d) $Omiotis\_Term\_Frequency(m,n)$;

5. Generate user query semantic vector $Query$;
6. For all the semantic description vectors within a kernel

i) Calculate cosine angle between the $Query$ vector and semantic description vectors of semantic kernel.
ii) If $(semantic\_similarity \geq User\_Threshold)$ Then append service no. and $semantic\_similarity$ to the $results$ vector
else drop the service.

7. Sort the $results$ and return ranked list of services $results$;

FIGURE 2: PROPOSED ALGORITHM

The proposed approach is divided into three phases:

i. *Parsing and weight generation of WSDL service profiles*
ii. *Calculation of integrated Semantic vectors for WSDL profiles*
iii. *Semantic matchmaking module for services*

i. *PARSING AND WEIGHT GENERATION OF WSDL SERVICE PROFILES*

In the proposed framework, in first stage service profiles are pre-processed to generate a set of tokens where relevant information under <element name> and <documentation> tags of the WSDL service profiles is extracted. All these extracted terms are stored in $Terms(X)$ vector. These terms are assigned weight using the four different schemes used in Information Retrieval [44]:

1. Term Frequency–Inverse Document frequency weighing ($TF-IDF$) scheme: The $TF-IDF$ is used for knowing the rare or important features in the corpus:

$$TF-IDF_{ij} = TF_{ij} * (\log \frac{N}{\{T_j \in D_n\}}) \; ; \qquad (1)$$

2. Binary Weightage: This kind of weightage scheme counts only the presence of a term within a WSDL profile. It does not consider the frequency of the term. It is calculated as:

$$Binary_{ij} = 1, \; if \; TF_{ij} > 0;$$
$$Binary_{ij} = 0, if \; TF_{ij} = 0; \qquad (2)$$

Here, TF_{ij} is the number of times that term j appears in WSDL profile for service i .

3. Term Occurrence wise weightage: This scheme counts the number of times a term has appeared in the WSDL profile.
$$Term\_Occurance_{ij} = TF_{ij}; \quad (3)$$
4. Term Frequency wise weightage [43]: In addition to the term frequencies counted in Term Occurrence weighing scheme; this approach also normalizes the normal term frequencies by the square root of the sum of squares of all frequencies of the terms present in the WSDL profiles.
$$Term\_Frequency_{ij} = \frac{TF_{ij}}{\sqrt{\sum_{k=1\,to X}(TF_{ik})^2}}; \quad (4)$$
The numerator is the frequency of the word being considered and the denominator is the square root of the sum of the squares of the frequency of each unique word. The term weighing (1-4) of all the services have been calculated using Rapidminer, a free available statistical tool.

After the four weighting matrices have been generated, these are forwarded to next phase of the framework to merge them with the semantic information of the services.

ii. *CALCULATION OF INTEGRATED SEMANTIC VECTORS FOR WSDL*

The semantic relatedness values of all the terms of the WSDL profiles are calculated with the dimension vector to generate a semantic matrix i.e. $Omiotis(X,n)$ for X terms present in the WSDL corpus. Here, dimension vector $Dim(n) = \{Dim_1, Dim_2, Dim_3, \cdots Dim_n\}$ contains different domains for the service set, as many times, it is not possible to allocate a single category to any service as services may potentially belong to more than one category. For calculating semantic relationship value between the terms and the dimensions, WordNet based Omiotis measure of semantic relatedness has been used [15]. Omiotis is the first measure of semantic relatedness between texts that considers all three factors for measuring the pair-wise word-to-word semantic relatedness scores. For weighting the semantic path Omiotis considers three key factors: (a) the semantic path length, (b) the intermediate nodes specificity denoted by the node depth in the thesaurus' hierarchy, and (c) the types of the semantic edges that compose the path. Omiotis is based on a sense relatedness measure, called SR. Semantic relatedness for a pair of terms $T(t_1,t_2)$ is calculated as follows [15]:

Definition 1 Given a word thesaurus O, let $T(t_1,t_2)$ be a pair of terms for which entries exist in O, let X_1 be the set of senses of t_1 and X_2 be the set of senses of t_2 in O. Let $S_1, S_2, \cdots, S_{|X_1||X_2|}$ be the set of pairs of senses, $S_k = (s_i, s_j)$, , with $s_i \in X_1$ and $s_j \in X_2$. The semantic relatedness of $T(SR(T,S,O)$ is defined as
$$max_{S_k}\{max_P\{SCM(S_k,O,P)\cdot SPE(S_k,O,P)\}\}$$
$$= max_{S_k}\{SR(S_k,O)\} \text{ for all } k = 1..|X_1|\cdot|X_2| \quad (5)$$
Semantic relatedness between two terms t_1, t_2 where $t_1 \equiv t_2 \equiv t$ and $t \notin O$ is defined as 1. Semantic relatedness between t_1, t_2 when $t_1 \in O$ and $t_2 \notin O$, or vice versa, is considered 0.

The experimental evaluation have proved that Omiotis measure of semantic relatedness approximates human understanding of semantic relatedness between words better than previous related measures [15].The main benefit of integrating Omiotis with $TF-IDF$ is that it improvesthe discovery accuracy as semantic information contained in the Omiotis is merged with the statistical information present in the $TF-IDF$. Details about the Omiotis and various terms used here are availablein [15].

The term dimension semantic vectors for all the terms of the WSDL profiles are calculated. Further, the semantic syntactic integration is applied to generate four semantic kernels using different weighing schemes. The semantic relatedness values of different terms of the services contained in $Omiotis(X,n)$ is merged with the different weighing schemes based matrices viz. $TF-IDF$, $Binary$, $Term\_Occurance$ and $Term\_Frequency$ thus generating semantic integration kernels of web service description vectors.

Four kernels are:

1. Omiotis-TFIDF Integration Kernel: The $TF-IDF(m,X)$ matrix is integrated with the Omiotis based Semantic Relatedness Matrix i.e. $Omiotis(X,n)$ to generate a Omiotis-TFIDF Integrated Kernel:
$$Omiotis\_TF\text{-}IDF = TF\text{-}IDF \times Omiotis \quad (6)$$
2. Omiotis-Binary Integration Kernel: The $Binary(m,X)$ matrix is integrated with the Omiotis based Semantic Relatedness Matrix i.e. $Omiotis(X,n)$ to generate a Omiotis-Binary Integrated Kernel:
$$Omiotis\_Binary = Binary \times Omiotis \quad (7)$$
3. Omiotis-Term Occurrence Integration Kernel: The $Term\_Occurance(m,X)$ matrix is integrated with the Omiotis based Semantic Relatedness Matrix $Omiotis(X,n)$ to generate a Omiotis- Term Occurance Integration Kernel:

$$Omiotis\_Term\_Occurance = Term\_Occurance \times Omiotis \quad (8)$$

4. Omiotis-Term Frequency Integration Kernel: The $Term\_Frequency(m, X)$ matrix is integrated with the Omiotis based Semantic Relatedness Matrix $Omiotis(X, n)$ to generate a Omiotis- Term Frequency Integration Kernel:

$$Omiotis\_Term\_Frequency = Term\_Frequency \times Omiotis \quad (9)$$

Finally, each row of these semantic kernels represents the semantic description vector for each service.

iii. *SEMANTIC MATCHMAKING MODULE FOR SERVICES*

In this phase, the semantic query vector for the users' query is calculated using Omiotis measure of semantic relatedness and the query is matched with the semantic web service description vectors of the services using cosine similarity measure. Thus, a ranked list of semantically similar services is generated as the output of the algorithm. The services with values less than the user specified threshold values are eliminated from the output list and remaining services are returned to the service consumer.

IV. IMPLEMENTATION OF THE APPROACH FOR WSDL SERVICE DESCRIPTIONS

The proposed approach was implemented on 98 services described in WSDL language; these service descriptions were downloaded from the Internet [45, 46]. These services were chosen from various domains like weather, stock, vehicle, food etc. The proposed algorithm has been evaluated using Precision, Recall and Fscore. Accordingly, set of relevant services for each query were defined and the framework was tested using ten queries with different user specified threshold values. The textual documentation, arguments, operations and service name from the service profiles were extracted out. After pre-processing, 562 terms were extracted from the WSDL profiles and stored in $Terms(562)$ vector. Further, for all the 562 terms in the $Terms(562)$ vector, the $TF-IDF(98,562)$ matrix, $Binary(98,562)$, $Term\_Occurance(98,562)$ and $Term\_Frequency(98,562)$ matrices were generated. Next, the semantic relatedness value between all the extracted terms of the services and all dimensions set $Dim(10)$ was calculated for in the dataset. The dimension vector constitutes the following 10 domains: Dim = {"Automobile", Book", "Film", "Weather", "Food", "Hospital", "Hotel", "SMS", "Stock", "Missile"}

TABLE 1.
COMPARISON OF PROPOSED APPROACH ON WSDL DATA SET WITH LATENT SEMANTIC ANALYSIS AT DIMENSION 10

Weightage scheme	Threshold Value	Proposed Approach			LSA Approach		
		Precision	Recall	Fscore	Precision	Recall	Fscore
TFIDF	0.5	84.45	92.75	87.93	75.45	92.32	78.36
	0.6	87.33	91.63	89.09	78.00	91.61	79.16
	0.7	92.09	89.06	90.27	79.00	88.27	77.01
	0.8	96.39	87.39	90.95	79.00	85.73	75.02
	0.9	97.50	85.91	90.71	80.16	84.62	75.15
Binary	0.5	66.09	92.75	73.78	40.64	71.98	45.88
	0.6	76.55	92.03	81.39	40.97	66.14	43.74
	0.7	86.48	88.24	86.94	39.72	57.59	40.57
	0.8	93.75	83.29	87.78	40.54	54.60	40.66
	0.9	98.75	75.91	84.15	54.67	53.83	49.81
Term Occurrence	0.5	85.16	92.75	88.31	68.62	92.71	75.42
	0.6	86.66	91.84	88.90	69.96	91.81	76.14
	0.7	91.55	90.73	90.88	74.17	90.18	78.46
	0.8	96.39	87.39	90.95	81.76	77.30	77.17
	0.9	97.50	85.85	90.64	89.00	65.89	71.69
Term Frequency	0.5	85.16	92.75	88.31	68.62	92.71	75.42
	0.6	86.66	91.84	88.90	69.96	91.81	76.14
	0.7	91.55	90.73	90.88	74.17	90.18	78.46
	0.8	96.39	87.39	90.95	81.76	77.30	77.17
	0.9	97.50	85.85	90.64	89.00	65.89	71.69

The semantic relatedness score between the terms within each WSDL profile and all the dimensions of the dimension vector was calculated using the Omiotis measure of semantic relatedness and stored in *Omiotis*(562,10) matrix. Further, the semantic information contained in the *Omiotis*(562,10) matrix was merged with the statistical information of the $TF-IDF$(98,562) matrix, *Binary*(98,562) , *Term _ Occurance*(98,562) and *Term _ Frequency*(98,562) matrices. Thus, four semantic kernels viz. $Omiotis\_TF-IDF$, $Omiotis\_Binary$, $Omiotis\_Term\_Occurance$ and $Omiotis\_Term\_Frequency$ based on different weighing schemes were calculated. Each row of these semantic kernels represents the semantic description vector for each service.

A total of ten queries were run to test the proposed approach on service profiles written in WSDL language. The semantic description vector of each query was matched against all the semantic description vectors for services in each of the four kernels and $semantic\_similarity$ was calculated between the queries and the service vectors. The proposed approach allows user to specify a degree of similarity matching threshold, i.e. $User\_Threshold$. Therefore, the services having $semantic\_similarity$ higher than the user specified $User\_Threshold$ will be returned to the user. The proposed approach was run for different threshold values of 0.5, 0.6, 0.8 and 0.9 for filtering the semantic value of match. The result of the proposed approach on WSDL profiles for ten queries in terms of Macro average precision, Macro average recall and Macro average F-Score were calculated.

The results of our approach were compared with information retrieval based Latent Semantic Analysis (LSA), $TF-IDF$ cosine similarity and Jaccard similarity. Latent Semantic Analysis, $TF-IDF$ cosine similarity and Jaccard similarity are all standard prevalent approaches which are normally considered for comparisons in IR research. The detailed results in terms of precision, recall and Fscore of top two performing approaches for ten user queries are provided in Table 1. The results were compared with the LSA (dimensions varying from 10 to 60).By looking at the results in Table 1, it can be clearly observed that the Macro average precision and Macro average Recall of proposed approach are considerably high than the precision of LSA10 approach.From the results, it can be analysed that our approach resulted in more relevant results in the final resultant set as the Macro average precision and macro average F-Measure of the proposed approach are much better than the results obtained from LSA (dimensions varying from 10 to 50).

The cosine similarities of the query vectors with the semantic description vectors of services for all the four kernels based on four weighing schemes were calculated. The comparative output of proposed approach with other approaches for all the four weighing schemes is presented in Figure 3 and Figure 4.

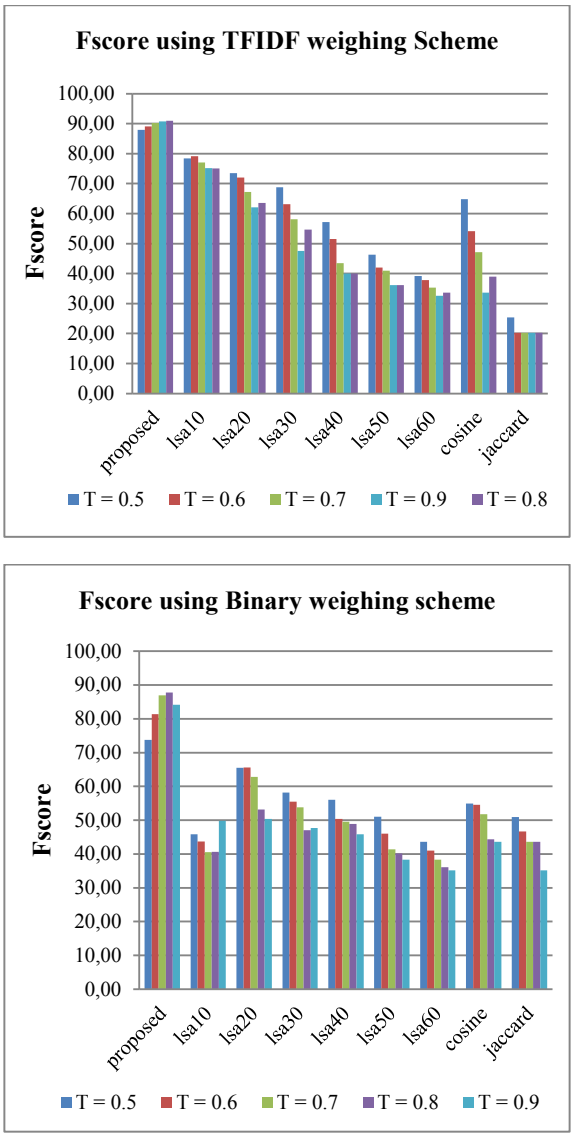
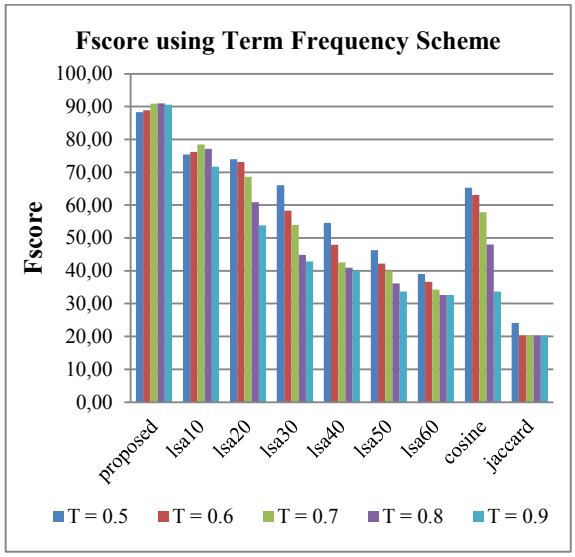


FIGURE 3: FSCORE COMPARISON FOR TFIDF AND BINARY WEIGHING SCHEME

Fscore comparison using TFIDF scheme: The proposed approach performed better than all other methods with Fscore value 87.9% at threshold 0.5. The accuracy improved on increasing the threshold value and attained its peak value of 90.9% at threshold value 0.8.Latent Semantic Analysis approach for dimension 10 performed next to the proposed approach

at different thresholds. The Fscore of cosine similarity was found to be 64.7% at threshold 0.5 and the results of jaccard similarity were found to be the lowest.

Fscore comparison using Binary scheme: Again, the system performed best for the proposed approach at all threshold values. For binary scheme, LSA approach at dimension 20 performed next to the proposed approach. All other approaches (except theproposed and LSA 10) gave better results at threshold value 0.5 whereas the proposed attained peak results at threshold 0.8 with Fscore 87.7% and outperformed the other approaches.



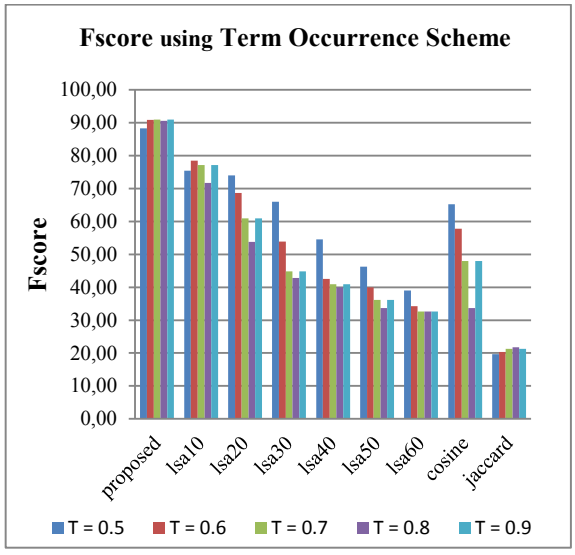


FIGURE 4: FSCORE COMPARISON FOR TERM OCCURRENCE AND TERM FREQUENCY WEIGHING SCHEME

Fscore comparison for Term Occurrence weighing scheme: The results of the proposed approach were quite promising for all threshold values in this scheme. The highest Fscore of 90.9% was achieved at threshold 0.8. The LSA approach performed better at threshold 0.7 at dimension 10 and for rest of the dimensions and approaches, better values were attained at threshold 0.5.

Fscore comparison for Term Frequency weighing scheme: In this scheme too, the results of the proposed approach were again found to be of highest values. Fscore values for this kernel were similar to the Term Occurrence weighing scheme. No major variations were seen in this weighing scheme.

During implementation, it was found that the Fscore of the proposed approach outperforms all the other methods for all the four weighing schemes. Based on the comparative analysis of all the results, it was found that the proposed framework is efficient and easy to use discovery approach, suitable for all kind of users. A novice user, who is not familiar with concerned ontologies, technology and implementation details can easily discover the existing services over the internet without any technical overhead. Ambiguity, polysemy and synonymy issues of the terms used in the service profiles are dealt through Omiotis measure of semantic relatedness. This approach matches the user query to the semantically similar service vectors of the semantic kernels according to the hybrid similarity score and finally returns a ranked list of semantically similar Web services along with their corresponding similarity scores.

V. CONCLUSION AND FUTURE SCOPE

Keeping in view the fact that the users have very less knowledge regarding the underlying technologies of the web services, discovery frameworks, description languages and implementation details, we have proposed a semantic framework that enables the web service discovery based on the combination of semantic and syntactic information of the service profiles. A novel approach has been presented which takes advantages from measures of semantic relatedness and statistical models for providing efficient means to automate the discovery process of Web services. The proposed approach is implemented on WSDL services and the experimental evaluations have shown that the performance of the discovery process can be significantly improved by combining the information retrieval techniques with the measures of semantic relatedness. The proposed approach proved to be effective in retrieving more relevant results for the user requirements.

REFERENCES

[1] S. Robak, B. Franczyk, Modeling web services variability with Feature diagrams, In: Proceedings of web services and database systems, 2002, pp. 20-128.

[2] E. Dashofy, M. van der Hoek, A. Taylor, A highly-extensible, XML-based architecture description language Software Architecture, In: Proceedings of Working IEEE/IFIP Conference, Amsterdam, Netherlands, 2001.

[3] F. Curbera, M. Duftler, R. Khalaf, W. Nagy, N. Mukhi, S. Weerawarana, Unraveling the Web Services, Web: An introduction to SOAP, WSDL, and UDDI, IEEE Internet Computing, 6(2), April 2002.

[4] World Wide Web Consortium (W3C), Web Services Description Language (WSDL) 1.1., 2001. http://www.w3.org/TR/wsdl/.

[5] UDDI Technical White Paper, 2000. http://www.uddi.org/pubs/Iru-UDDI-Technical-White-Paper.pdf

[6] N. Shadbolt, T. B. Lee, W. Hall, The Semantic web revisited, IEEE Intelligent Systems, 21(3), 2006, pp. 96-101.

[7] D.Martin, M.Burstein, J.Hobbs, O.Lassila, K.Sycara, OWL-S: Semantic markup for web services, 2004. http://www.w3.org/submission/OWL-S/.

[8] D. Fensel, H. Lausen , A.Pollers, J.D. Bruijn, M. Stollberg, D. Roman, J. Domingue, Enabling Semantic Web Services: The web service modeling Ontology, Springer Verlag, ISBN 978-3-540-34520-6, 2007, pp. 38-61.

[9] R. Akkiraju, J. Farrell, J. Miller, M. Nagarajan, T. Schmidt, A. Sheth, K. Verma, Web service semantics-WSDL-S,2005. http://www.w3.org/submission/WSDL-S/.

[10] J. Farrell, H. Lausen, Semantic Annotations for WSDL and XML Schema, 2007. http:/www.w3.org/TR/SAWSDL/.

[11] C. Kiefer, A. Bernstein, The creation and evaluation of ISPARQL strategies for matchmaking, In: Proceeding of Fifth European Semantic Web Conference (ESWC), volume-5021, LNCS, Springer Verlag Berlin Heidelberg, 2008, pp. 463-477.

[12] M. Klusch, B. Fries, K. P. Sycara, OWL-MX: A Hybrid semantic Web Service matchmaker for OWL-S services, Web Semantics 7(2), 2009, pp.121-133.

[13] M. Klusch, F. Kaufer, Wsmo-mx: A hybrid semantic web service matchmaker, Web Intelligence and Agent Systems, 7(1) 2009, pp. 23-42.

[14] D. Wei, T. Wang, J. wang, A. Bernstein SAWSDL-iMatcher: A customizable and effective Semantic Web Service Matchmaker, Web Semantics: Science, Services and Agents on the World Wide Web 9(4), 2011, pp. 402-417.

[15] G. Tsatsaronis, I. Varlamis, M. Vazirgiannis, Text relatedness based on a word thesaurus, Artificial Intelligence Research ,Volume-37, 2010, pp. 1-39.

[16] J.A. Nasir, A. Karim, G. Tsatsaronis and I. Varlamis, A Knowledge-based Semantic Kernel for Text Classification, In: Proceedings of the 18th international conference on String processing and information retrieval, 2011, pp. 261-266.

[17] M. Paolucci, T. Kawamura, T.R Payne, K.P Sycara, and Semantic matching of Web Services capabilities In: Horrocks, I., Hendler, J. (eds.) ISWC, LNCS, vol. 2342, Springer, Heidelberg, 2002, pp. 333-347.

[18] D. Bruijn, R. Lara, A. Polleres, D. Fensel, and OWL- DL vs. OWL light: conceptual modeling and reasoning for the semantic Web, In: Proceeding of WWW conference, Chiba, Japan, 2005.

[19] T. Kawamura, J.A.D Blasio, T.Hasegawa, M.Paolucci, K.Sycara, Public deployment of semantic Service Matchmaker with UDDI business Registry, LNCS, Volume 3298, Springer, Heildelberg, 2004, pp.752-766.

[20] R. Gronmo, M.C. Jaeger, Model-Driven Semantic Web Service Composition, In: APSEC December 2005, pp. 79–86.

[21] L. Li, I. Horrocks, A software framework for matchmaking based on Semantic Web technology, In: Proc. of WWW, Budapest, May 2003, pp. 331–339.

[22] H. Wang, Z. Li, A Semantic Matchmaking Method of Web Services Based On SHOIN(D)*, In: Proc. of IEEE APSCC, Guangzhou, China, December 2006, pp. 26–33.

[23] K. Sycara, S.Widoff, M.Clutch, J.Lu, LARKS: Dynamic Matchmaking among Heterogenous Software Agents in Cybersace,Autonomous Agents and Multi agents Systems, Volume 5, 2002, pp.173-203.

[24] K. Verma, K. Sivashanmugam, A. Sheth, A. Patil, S. Oundhakar, J.Miller, METEOR-S WSDI: A Scalable Infrastructure of Registries for Semantic Publication and Discovery of Web Services, Information Technology and Management, 6(1),2005, pp. 17-39.

[25] K. Sycara, S.Widoff, M.Clutch, J.Lu, LARKS: Dynamic Matchmaking among Heterogenous Software Agents in Cybersace,Autonomous Agents and Multi agents Systems, Volume 5, 2002, pp.173-203.

[26] S. Batra, S. Bawa, Semantic discovery of Web Services using principal component analysis, Physical Sciences, 6(18), 2011, pp. 4466-4472.

[27] T. Wen, G. Sheng, Y. Li, Q. Guo, Research on Web Service Discovery with Semantics and Clustering, In: Proceedings of IEEE sixth joint international Conference on Information technology and Artificial Intelligence (ITIAC), 2011, pp. 62-67.

[28] J. M. García, D. Ruiz, A. Cortes, Improving semantic Web Services discovery using SPARQL-based repository filtering, Web Semantics: Science, Services and Agents on the World Wide Web, Elsevier, 2012, pp.12-24.

[29] R. Thiagarajan, W. Mayer, and M. Stumptner, Semantic Service Discovery by Consistency-Based Matchmaking, Advances in Data and web Management, Springer,2009, pp. 492-505.

[30] M. Senvar, A. Bener, Matchmaking of Semantic Web Services Using Semantic-Distance Information, *ADVIS, volume 4243, LNCS, Springer, 2006, pp. 177-186.*

[31] L. Chen, Z. Song, Y. Zhang, and Z. Miao, Wordnet Enhanced Dynamic Semantic Web Services Discovery, Emerging Research in Artificial Intelligence and Computational Intelligence,CCIS volume-237, 2011, pp. 529-536.

[32] A. Grintsvayg, V. D. Veksler, R. Lindsay, W. D. Gray, Vector generation of explicitly defined multidimensional semantic space, In: Proceedings of ICCM Eight international conference on Cognitive Modeling, Oxford, UK, 2007, pp. 231-232.

[33] I. Kaur, A. J. Hornof, A Comparison of LSA wordnet and PMI for predicting User Click Behaviour, In: Proceedings of the conference on human factors in Computing,CHI, pp. 51-60.

[34] Normalized (Point wise) Mutual Information in Collocation Extraction http://www.ling.uni-potsdam.de/~gerlof/docs/npmi-pfd.pdf

[35] R. L. Cilibrasi, P.M.B. Vitanyi, The Google Similarity Distance, ArXiv.org or Clustering by Compression, IEEE Journal of Transactions in Information Theory, 51(4), 2004, pp. 1523 - 1545.

[36] E. Gabrilovich, S. Markovitch , Computing Semantic Relatedness using Wikipedia-based Explicit Semantic Analysis, In: Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI), Hyderabad, India, 2007.

[37] P. Resnik, Using information content to evaluate semantic similarity in a taxonomy, In: Proceedings of the 14th International joint conference on Artificial intelligence,Volume-1, Morgan Kaufmann Publishers Inc., San Francisco CA,USA,1995, pp. 448-453.

[38] D. Lin, An Information-Theoretic Definition of Similarity, In: Proceedings of the Fifteenth International Conference on Machine Learning (ICML '98), Jude W. Shavlik (Ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1998, pp. 296-304.
[39] J. Jiang, W. Conrath, Semantic similarity based on corpus statistics and lexical Taxonomy, In: Proceedings of International conference on Research on Computational Linguistics, 1997, pp. 90-98.
[40] M. Crasso, A. Zunino, M. Campo, AWSC: An approach to Web Services classification based on machine learning techniques, CONICET, Inteligencia Artificial, RevistaIberoamericana de Inteligencia Artificial. 12(37), 2008, pp. 25-36.
[41] http://projects.semwebcentral.org/projects/owls-tc/.
[42] http://Omiotis.hua.gr/WebSite/wsinfo.html.
[43] http://rapid-i.com/rapidforum/index.php?action=printpage;topic=3728.0
[44] R. Cummins, O. Colm, Evolving general term-weighting schemes for information retrieval: Tests on larger collections, Artificial Intelligence Review, 24.3-4, 2005, 277-299.
[45] www.service-repository.com
[46] www.xmethod.com

Shailja Sharma is working as System Analyst in Kurukshetra University, Kurukshetra. She did B.Tech from Kurukshetra University, Kurukshetra and M.E in Computer Sc. and Engineering from Thapar University, Patiala. She is a research scholar in the area of Semantic Web Services from the Department of Computer Applications, National Institute of Technology, Kurukshetra. Her research interests are focused on Semantic Web, Web Services, Data Mining and Machine Learning.

Jagdeep Singh Lather is working as a Professor at Department of Electrical Engineering and Computer Applications, National Institute of Technology, Kurukshetra. He holds a B.Tech, M.Tech and doctorate degree in Electrical Engineering. He has over 15 years of experience in teaching and research. His areas of interests are Robust Control, Flexible AC Transmission systems, Mining Techniques and Decision Trees, ANN and Fuzzy Logic. He is life time member of Scientific and Professional Societies, ISTE, India.

Mayank Dave is working as Professor in the Department of Computer Engineering at National Institute of Technology, Kurukshetra, since 1991. He holds a PhD in Computer Science and Technology from Indian Institute of Technology, Roorkee, India. He has twenty four years of experience in research in the area of Computer Networks. His areas of interests are Computer Networks, Database Systems, QoS in Mobile Adhoc and Sensor Networks, Software Engineering, Web Technologies. He is a member of the IEEE, IEEE Computer Society, IETE, Institution of Engineers (I), CSI and ISTE.

Fairness in Kademlia with Random Node Joins

Zoltán Novák, Zoltán Pap

***Abstract*—Kademlia is among the most prevalent Distributed Hash Table (DHT) protocols in practice. To understand load-balancing and fairness properties of any DHT system one of the key requirements is to study and understand the zone size distribution of the network. Already existing and well known analytical results in this field are not applicable to Kademlia directly, due to its unique addressing mechanism. We show that a direct connection exists between the size of the zones of a given Kademlia network and the shape parameters of the data structure called PATRICIA trie filled with the overlay addresses of the same network. Then analytical description of the asymptotic properties of the Kademlia zone size distribution is provided based on the existing literature on random binary tries. We compare Kademlia to the Chord DHT, and show that Kademlia provides a fairer zone size distribution. These results can be used to achieve better load balancing in DHT systems.**

***Index Terms*—Consistent hashing, load balancing, asymptotic bounds, peer-to-peer networks.**

I. INTRODUCTION

We examine load distribution in the Kademlia [1] distributed hash table (DHT) system which is one of the most widely used peer to peer (P2P) overlay in practice in these days[1].

Distributed hash tables [1]–[3] - as their name suggest - are for storing and retrieving arbitrary data in P2P networks using hash keys. Data is distributed among all participant peers in the network. Each node is responsible for a given part of the hash space called zone. A node stores a given value if the hash key of the value falls into its zone. The zone of the node is usually determined by a predefined relation between the overlay address of the node and the addresses of the other online nodes. The goal is to minimize the number of zones that have to be modified (increased or decreased) when additional nodes join or leave the system. A somewhat contradicting goal is to keep the zone sizes balanced. These goals are usually achieved by a technique called consistent hashing [4].

The relative size of the zone of a given node determines the expected relative number of data items it has to store. It also influences the number of routes directed through the node. Thus having a larger zone size means a larger expected average load for the node in the system.

Using the assumption that both the node addresses and the hash keys are uniformly distributed[2], the zone size distribution of a DHT protocol can be calculated by using probabilistic models.

Manuscript received April 9, 2015, revised June 12, 2015

Zoltán Novák and Zoltán Pap are with Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics, 1117, Budapest Magyar tudósok krt 2, Hungary. E-mail: {novak,pap@tmit.bme.hu}

[1]See http://en.wikipedia.org/wiki/Kademlia#Implementations,

[2]The first is true by definition, because in DHTs each joining node choose a random address. The second part can be considered true due to the properties of the commonly used hash functions.

The exact distribution can be used to describe, compare or evaluate the performance characteristics of different DHT systems, or to devise efficient uniform random node selection algorithms [5], [6]. These algorithms could be used for statistical estimations in large networks or as an algorithmic building block in randomized network algorithms. Uniform random node selection can also be used directly for load balancing purposes. A specific example of application was presented by Scott Lewis et al. [7]. Their scalable Byzantine agreement algorithm is based on the availability of uniform random node selection in a network.

Load balancing is also one of the areas that could benefit much from the exactly known apriori distribution of the load [8].

II. RELATED WORK

A. Consistent hashing

Karger et al. [4] have been introduced consistent hashing to minimize the number of values that have to be moved upon a hash table resize. They have provided the following algorithm:

The storage nodes (buckets) are randomly placed (hashed) onto a unit circle, and each bucket stores the data with hash keys between its hash and the hash of the previous bucket. The handling of the hash space in the Chord [2] DHT system is based upon the same concept.

B. DHT zone distribution

The probabilistic properties of the zone sizes have been investigated in the original article that has introduced cosnsitent hashing [4]. The load distribution of Chord DHT have been first investigated by using simulation in [2]. The asymptotic distribution of the minimal zone size in Chord have been described in [5]. Cuevas et al. [9] have examined routing fairness in the Chord system, based on the Chord zone size distribution (as nodes with larger zone tend to appear in other nodes routing table more frequently).

Finally, Wang et al. [10] have provided several limits – that are true with high probability – for Chord's zone size distribution, such as the distribution of minimal, maximal zones and have also examined different joining strategies like half splitting[3], and multipoint sampling[4]. It may be interesting to mention, that for the half-splitting case, the authors of [10] have relied on results coming from the research of regular tries, but have not recognised the connection between the Kademlia address space and PATRICIA tries. This article also contains a good collection of the prior results of the field.

[3]When a joining node always choose an address that splits the original zone into two equal parts.

[4]When a joining node choose its address by splitting the largest zone it has found after sampling some random locations.

C. Kademlia

For Kademlia the analytical literature is sparse. Recently Cai and Devroye [11] have provided analytical results about the search times in Kademlia. Their method is based on regular tries: they've started with the initial assumption that the address trie is balanced (in this case regular and PATRICIA tries are similar), and then have refined the result by relaxing on this balancedness assumption. They did not recognise the direct connection between these exact (unbalanced) description of the address space and PATRICIA tries.

D. Occupancy problem, poissanization

As we will show in section IV, the zone size of a Kademlia node depends on whether particular address sets – these sets are depend on the address of the given node – contain at least one online node. The general version of this problem is called occupancy problem:

Given n balls and c cells we assign the balls into the cells randomly. What will be the probability that exactly k cells remain empty?

In the most general case, cells can have different selection probabilities, or the number of cells can be infinite – the only restriction then is that the sum of the selection probabilities have to be 1. This field of probability theory [12]–[14] gives a general theoretical framework for our problem.

It is also worth to mention a general technique to solve similar problems: analytic depoissonization [15]. In many cases these kind of balls-in-urns problems can be modelled easier with poissonization. Poissonization means that the exact Bernoulli models are replaced with approximative Poisson models (e.g. imagine balls arriving into the urns according to Poisson processes). With depoissonization it is possible to translate back the results of the Poisson model to the original Bernoulli model.

Reference [13] has also presented results about the equivalence of the moments of the original and the poissonizated occupancy distributions.

E. PATRICIA tries

PATRICIA trie [16] is the compact version of the regular trie (also called prefix tree). These structures are commonly used to efficiently store strings together with their prefixes. By providing efficient prefix search, they are particularly suitable for storing dictionaries or routing tables.

In the generic trie each node of the tree represents a character of the stored string, and a path to an internal node in the tree represents a prefix string. Below that node one can find all the strings sharing that same prefix. A path form the root to a leaf node gives a stored string, where in each step we get the next character of the string.

PATRICIA trie (also called radix tree) is a space optimized version of the regular trie, where each node with only one child is merged with its parent. In this case a node can contain larger fragments of the prefix not just one character. (Figure 1 shows a PATRICIA trie storing five binary strings.)

As we show later, the shape parameters of the binary PATRICIA trie are directly related to the zone size distribution of the Kademlia DHT system.

Unfortunately it has been proved to be notoriously hard to describe the exact shape parameters of random PATRICIA tries, and despite it has been introduced for more than forty years ago, the properties of the PATRICIA trie are still actively researched. Luckily there exists many asymptotic results in this field, that can be applied directly to our problems.

References [17]–[19] provide asymptotic and limiting distributions of various shape parameters of random PATRICIA tries. A recent result about the expected value of the number of tree nodes at a certain level of the trie have been presented in [20]. An interesting result is that the variance of the insertion cost of random strings into PATRICIA tries – which is related to the path length distribution – is constant: $1 + \mathrm{O}(1)$ $(= 1.000000000001237\ldots)$ [21]. Finally there are also results about the asymmetrical case where the input alphabet is not uniformly distributed [22], [23].

III. KADEMLIA

This section is a short introduction to the Kademlia [1] DHT.

Kademlia is based on a 160 bit address space, to which both nodes and keys are mapped. Each key-value pair is stored on the node having the closest overlay address in the system to the given key. Distance is calculated using the result of bitwise binary XOR operator ($\oplus$) interpreted as a natural number:

$$\mathrm{d}(nodeaddress, key) = nodeaddress \oplus key$$

Each node maintains 160 tables to store routing information. In Kademlia terminology these tables are called k-buckets. The i-th k-bucket contains at most K nodes whose distance from the current node is between 2^{160-i} and $2^{160-i+1}$, where K is a pre-chosen system parameter.

K-buckets are ordered lists of nodes, with the most recently seen node at the beginning of the list. If a node A receives a message from another node B, than A tries to insert B into the appropriate k-bucket, if there's still room. If the given k-bucket is full, A sends PING to the node from the end of the list; if it replies, A moves it to the head of the list; if it does not, A deletes it from the list, and replaces it with B. With adequate network traffic, k-buckets remain consistent thanks to the procedures above.

A. Searching

The Kademlia protocol defines four remote procedure calls (RPC). Each participating node have to implement these:

- PING, check if a node is still connected;
- STORE, stores a key and corresponding data;
- FIND_NODE with an address as its parameter, returns the K closest values to the given address from the node's routing tables;
- FIND_VALUE with a key as its parameter, if a node stores data corresponding to the key, it returns the result data; otherwise it behaves identically to FIND_NODE.

Using these RPCs a node can find the closest peer to a given address. For example lets assume that a node (X) wants to

look up the closest node to a key (y). The search goes through the following steps:

1) Node X creates a list L containing the K closest addresses to y. Initially it fills this list from its own k-buckets.
2) X selects α unmarked nodes from the list, and runs the FIND_NODE RPC on them (α is a predefined system-wide parameter).
3) X updates the list L by merging the return values of the FIND_NODE RPCs. Then it keeps only the K closest addresses to y. It also marks every node in the list on which the FIND_NODE RPC has been already run.
4) If the list still contains unmarked nodes, return to step 2.
5) The result node is the one with the closest address to y in L.

When a node leaves the network, it simply copies its data to the nearest node, and disconnects.

B. Defining zone distribution

Definition 1: Let $A = \{0, 1, \ldots, 2^{160} - 1\}$ be the set of all addresses in the system. Let N be the set of occupied addresses (the set of online nodes). We denote the number of online nodes $|N|$ with n.

Joining nodes choose a uniformly distributed random address independently from each other. We assume that $n > 0$, i.e., every system has at least one node online.

Definition 2: Let $X \in N$ be the address of a node in the system, then define $\mathrm{Close}(X) \subseteq A$ as the set of addresses where:

$$\mathrm{Close}(X) = \{Z \mid \forall Y \in N, Y \neq X, X \oplus Z < Y \oplus Z\}$$

where $\oplus$ is the bitwise XOR operation, and its result is interpreted as a natural number.

$\mathrm{Close}(X)$ can be imagined as a kind of Voronoi cell of X: the set of all addresses that are closer to X – according to the XOR distance – than to any other online node.

Definition 3:

Let the zone size $\mathrm{T}(X)$ of a node $X \in N$ be:

$$\mathrm{T}(X) = \frac{|\mathrm{Close}(X)|}{|A|},\ (0 < \mathrm{T}(X) \leq 1).$$

The zone size of X represents the portion of addresses (A) that are closer to X than to any other online node. For example if $\mathrm{T}(X) = 0.5$, then if a uniformly random address R from the address set A is chosen – this is the case in practice when a hash key is calculated to store a value – the closest online node to R will be X with a probability of 0.5.

This way the distribution of the zone sizes in the system corresponds to the load distribution – assuming that the hash keys are uniformly distributed.

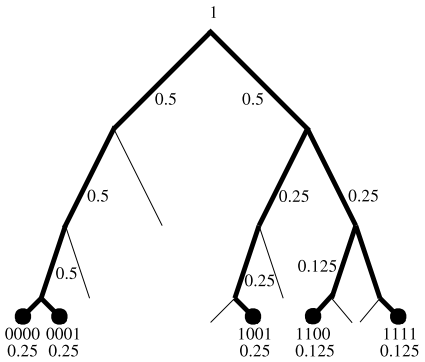
IV. KADEMLIA AND PATRICIA TRIES

In this section we present a connection between the Kademlia zone sizes and the shape of the PATRICIA trie of the Kademlia addresses. Then the cumulative distribution function of Kademlia zone sizes is provided. Utilizing existing results about random binary PATRICIA tries we describe the first two moments of this zone size distribution, and an estimation of the minimal zone size in the system is also provided. Finally we provide an asymptotic estimation of the Jain's fairness index of the zone distribution, as a measure of load fairness.

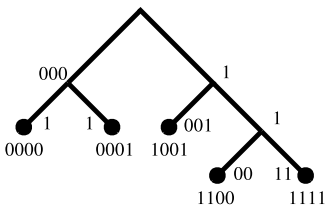
A. Visualizing Kademlia zone sizes

First let us try to visualise the zone size distribution in XOR distance as defined in section III-B.

1) The sum of zone sizes for all online nodes in the system is 1.
2) The sum of zone sizes for online nodes whose addresses start with 0 or 1 are 0.5 and 0.5 respectively, if there is at least one online node in both the 0xxx... and the 1xxx... address space.
3) Groups of nodes with prefix: 00, 01, 10, 11 share 0.25, 0.25, 0.25, 0.25 parts of the whole territory if all address prefixes contain at least one online node.
4) and so on...



(a) Division of zone sizes in XOR distance



(b) Kademlia addresses in PATRICIA trie

Figure 1: Visualizing Kademlia zone sizes

What happens if there isn't any node with prefix 10? Then nodes with prefix 11 will share on a 0.5 territory, as they are the only nodes with address prefix 1, and nodes with prefix 1 share 0.5 territory according to bullet 2 above. In Figure 1a the division of the zones is depicted for addresses: 0000, 0001, 1001, 1100, 1111 in the four bit address space.

B. Kademlia zone sizes and the PATRICIA trie of addresses

We have seen that according to the XOR distance division of the zone happens only if there is a branching in the regular

trie of the addresses. Because in PATRICIA trie the internal nodes with only one child (non-branching nodes) are merged with their parents, the path length of the trie to a given address leaf equals the number of branching on the path in the original trie.

Using this insight the division of the zones can be described with a corresponding PATRICIA prefix trie. The leaves of the trie are the occupied addresses. At the root, we begin with a zone size of 1, and at every lower level the territory is divided by two.

Figure 1 shows a division of zones between nodes: 0000, 0001, 1001, 1100, 1111 in the four bit address space, and the corresponding PATRICIA trie (in Figure 1b) storing the same addresses.

The size of the zone of a node is 2^{-l} where l is the path length of the address in the binary PATRICIA trie.

The sum of the zone sizes is always one – in accordance with the well known Kraft's inequality, that states that for every binary tree:

$$\sum_{\ell \in \text{leaves}} 2^{-\text{depth}(\ell)} \leq 1$$

where equality holds if every internal node has two children, which is true by definition in PATRICIA tries.

C. Distribution of zone sizes in Kademlia

We have two conflicting assumptions:

1) Joining nodes choose their addresses from a finite address space independently and randomly with uniform distribution, meaning that address collision is possible
2) Every node has a unique address

If there would be any measurable chance of address collision, it could be handled by increasing the address space. Simply we assume that this method would be used instead of other methods such as reconnection with a different address. It is worth considering that although the analytical description would be somewhat different for the aforementioned two cases, the practical numerical values would only be different by the order of magnitude of the hash collision probability - which is – by design – negligible in practice. Therefore in the rest of the paper we simply consider the size of the address space (the height of the PATRICIA trie of addresses) unbounded. Similar simplifications (for example modelling the Chord ring with a continuous unit circle) are prevalent in the literature.

Definition 4: Let $\mathrm{P}_n(\mathrm{T} \leq x) = \mathrm{F}_n^T(x)$ be the cumulative distribution function (CDF) of T (the zone sizes) in a system with n independently and randomly chosen node addresses. Then, in a system with one node ($n = 1$):

$$\mathrm{F}_1^T(x) = \begin{cases} 0 & \text{if } x \leq 1, \\ 1 & \text{if } x > 1. \end{cases}$$

We can write a recursive definition of $\mathrm{F}_n^T(x)$ using the law of total probability:

1) Let us visualise the n node addresses at the root of the corresponding regular trie (Fig. 1a). Every address begins with 0 or 1 with a probability of 0.5 respectively. The root node divides the n nodes into two sets.
2) The cardinality of these two sets has a binomial distribution, and they sum up to n:

$$P(\text{no address begins with } 0) = \binom{n}{0} 2^{-n},$$
$$P(1 \text{ address begins with } 0) = \binom{n}{1} 2^{-n},$$
$$\vdots$$
$$P(n \text{ addresses begin with } 0) = \binom{n}{n} 2^{-n}.$$

3) Let us assume that 5 addresses begin with 0 and $n-5$ with 1. If $\mathrm{F}_5^T(x)$ and $\mathrm{F}_{n-5}^T(x)$ is known, the CDF of their combination can be written. As each of the two branches shares upon only 0.5 zone, we have to use $\mathrm{F}_5^T(2x)$ and $\mathrm{F}_{n-5}^T(2x)$. Note that this would not be the case if the division was $(n;0)$ or $(0;n)$, because than the zone is not halved (no new level added in the corresponding PATRICIA trie)!
4) Using the law of total probability we can write the following recursive definition:

$$\mathrm{F}_n^T(x) = \frac{1}{2^n} \left(\binom{n}{0} F_n^T(x) + \binom{n}{n} F_n^T(x) + \right.$$
$$\left. + \sum_{k=1}^{n-1} \binom{n}{k} \left(\frac{k}{n} \mathrm{F}_k^T(2x) + \frac{n-k}{n} \mathrm{F}_{n-k}^T(2x) \right) \right)$$

5) Finally – after rearranging occurrences of $\mathrm{F}_n^T(x)$ to the left – we reach the following recursive formula as the CDF of the exact zone distribution of the Kademlia DHTs:

$$\mathrm{F}_1^T(x) = \begin{cases} 0 & \text{if } x \leq 1, \\ 1 & \text{if } x > 1. \end{cases}$$

$$\mathrm{F}_n^T(x) = \frac{1}{2^n - 2} \sum_{k=1}^{n-1} \left(\binom{n-1}{k-1} \mathrm{F}_k^T(2x) + \right.$$
$$\left. + \binom{n-1}{k} \mathrm{F}_{n-k}^T(2x) \right)$$

D. Moments of the zone sizes in Kademlia

Starting from the known recursive generating function of the PATRICIA trie's path lengths:

$$g_x(1) = 1$$

$$g_x(n) = \frac{2x}{2^n - 2} \sum_{i=1}^{n-1} \binom{n-1}{i-1} g_x(i)$$

We can recognise that by substituting $x = \frac{1}{2}$ to this generating function of the trie path length distribution, we get the formula for the expected value of the Kademlia zone sizes.

Here the result can be inferred much easier considering the fact that nodes share the entire hash space, so on average they get nth part of it:

$$E(T) = \frac{1}{n} \ (= g_{\frac{1}{2}}(n) = \frac{1}{2^n - 2} \sum_{i=1}^{n-1} \binom{n-1}{i-1} g_{\frac{1}{2}}(i))$$

As a by-product this short-cut provides an interesting identity for the recursive formula of the generating function.

Using the similar insight about the generating function, the variance of the zones can be defined as:

$$V_n(T) = E_n(T^2) - E_n^2(T) = g_{\frac{1}{4}}(n) - \frac{1}{n^2}$$

We can analyse the asymptotic behaviour of $g_{\frac{1}{4}}(n)$ by using the generating function of the poissonizated limiting distribution:

$$\mathrm{G}_x(n) = \prod_{i=1}^{\infty} \left[e^{\frac{-n}{2^i}} + (1 - e^{\frac{-n}{2^i}})x \right]$$

This distribution is derived according to section II-D. We simply assume that branching happens at a given node in PATRICIA trie according to Poisson distribution instead of using the proper Binomial distribution.

According to [14] the variance of the original and the poissonizated distribution is asymptotically close. For certain cases the difference is $\mathrm{o}(1)$ (small ordo), meaning that they are asymptotically equivalent[5]. Therefore we calculate $G_{\frac{1}{4}}(n)$ first.

From the generating function we have:

$$G_{\frac{1}{4}}(n) = \prod_{i=1}^{\infty} \left[e^{\frac{-n}{2^i}} + (1 - e^{\frac{-n}{2^i}}) \frac{1}{4} \right] = \prod_{i=1}^{\infty} \left[\frac{1}{4} + \frac{3}{4} e^{\frac{-n}{2^i}} \right]$$

As a simplification let's assume that n is a power of two, then the limit of $G_{\frac{1}{4}}(n)$ as n goes to 2^∞ is:

$$\lim_{n \to 2^\infty} G_{\frac{1}{4}}(n) = \lim_{n \to 2^\infty} \prod_{i=1}^{\infty} \left[\tfrac{1}{4} + \tfrac{3}{4} e^{\frac{-n}{2^i}} \right] =$$

$$= \lim_{n \to 2^\infty} \frac{1}{4^{\log_2 n}} \prod_{i=1}^{\log_2 n} \left[1 + 3e^{\frac{-n}{2^i}} \right] \prod_{i=\log_2 n+1}^{\infty} \left[\frac{1}{4} + \frac{3}{4} e^{\frac{-n}{2^i}} \right] =$$

$$= \lim_{n \to 2^j, j \to \infty} \frac{1}{n^2} \prod_{i=1}^{j} \left[1 + 3e^{-2^{j-i}} \right] \prod_{i=j+1}^{\infty} \left[\frac{1}{4} + \frac{3}{4} e^{-2^{j-i}} \right] =$$

$$= \lim_{n \to 2^j, j \to \infty} \frac{1}{n^2} \prod_{i=0}^{j-1} \left[1 + 3e^{-2^i} \right] \prod_{i=1}^{\infty} \left[\frac{1}{4} + \frac{3}{4} e^{-2^{-i}} \right] =$$

$$= \lim_{n \to 2^\infty} \frac{1.5254695585786\ldots}{n^2}$$

The constant of the last line is the result of calculating the (existing) limits of the products numerically. By relaxing the assumption that n is a power of two, the results may be different due to the nonzero fractional part of $\log_2 n$.

Instead of deriving this more generic solution, we have simply used this specific asymptotic behaviour of the limiting distribution at large powers of 2 as a clue to search for $g_{\frac{1}{4}}(n)$ in the form of: $g_{\frac{1}{4}}(n) \simeq c/n^2$.

By numerical calculations we have found that c is oscillating around 1.525 with a decreasing amplitude as n increases[6]:

$$g_{\frac{1}{4}}(n) \simeq \frac{1.525}{n^2}$$

This gives:

$$V_n(T) = g_{\frac{1}{4}}(n) - \frac{1}{n^2} \simeq \frac{0.525}{n^2}$$

An alternative characterization of the zone size distribution can be given by calculating Jain's fairness index. This index can be used to describe fairness with a constant value, when the participants share on some finite resource (such as the hash space in our case). It has been defined as:

$$\mathcal{J}(x_1, x_2, \ldots, x_n) = \frac{\left(\sum_{i=1}^{n} x_i \right)^2}{n \sum_{i=1}^{n} x_i^2}$$

The result ranges from $1/n$ when one node gets all the resources (worst case) to 1 when nodes have equal shares of the resources (best case). The index is k/n if k nodes equally share the resource, while the other $n - k$ node receive zero amount.

From the variance calculation it follows that in the case of Kademlia:

$$\mathcal{J}_T \approx 1/1.525 \approx 0.655$$

E. Distribution of the size of the minimal zone in Kademlia

The distribution of the minimal zone can be used to achieve efficient uniform random node selection in a Kademlia system. The distribution of the minimal zones can be derived with the method presented in section IV-C.

For the details of the derivation of the exact cumulative density function (CDF) of the minimal sized zone in Kademlia, please refer to the previous work of the authors [6]. Here we present the result without further explanation:

$$\mathrm{F}_1^{T_{min}}(x) = \begin{cases} 0 & \text{if } x \leq 1, \\ 1 & \text{if } x > 1. \end{cases}$$

$$\mathrm{F}_n^{T_{min}}(x) = \frac{1}{2^n - 2} \sum_{k=1}^{n-1} \binom{n}{k} \Big(\mathrm{F}_k^{T_{min}}(2x) + \mathrm{F}_{n-k}^{T_{min}}(2x) - \\ - \mathrm{F}_k^{T_{min}}(2x) \mathrm{F}_{n-k}^{T_{min}}(2x) \Big)$$

The minimal zone size in Kademlia is in direct relationship with the height of the PARTICIA trie of addresses. Having a trie with height h, the corresponding Kademlia network will have a minimal zone size of: 2^{-h}.

Key results about the heights of PATRICIA tries have been presented in [18]. For the random binary case the height of the

[5]It may be interesting to mention here that $G_{1/2}(n)$ – the expected value of the poissonizated distribution – can be given exactly in closed form, it equals: $\frac{1-e^{-n}}{n}$

[6]After $n > 1000$ these four decimal digits of 1.525 gets stabilized.

trie is – depending on n – oscillating around the most probable value of:

$$h_1 = \lfloor \log_2 n + \sqrt{2 \log_2 n} - \frac{3}{2} \rfloor + 1$$

The height of the trie is concentrated on h_1 for most n, and for some n it is either concentrated on h_1 and $h_1 + 1$ or on h_1 and $h_1 - 1$.

From this the most probable value of the minimal zone is simply:

$$T_{min} = 2^{-h_1}$$

The random node selection algorithm of the authors [6] relies on the estimation of minimal zone size in the system. In that particular case using an estimate that is lower than the actual minimal zone size results in a perfectly uniform random node selection. Contrarily underestimating the minimal zone size by a large margin results in a large increase in the run time of the random node selection algorithm. In this special case the probable underestimation of the actual minimal zone size:

$$2^{-h_1 - 1} = 2^{-\lfloor \log_2 n + \sqrt{2 \log_2 n} - \frac{3}{2} \rfloor} \gtrapprox T_{min}$$

could provide a viable trade-off.

Alternatively by using the asymptotics of [20][7], any zone size can be characterized with the expected number of nodes having smaller zone – this can be useful to estimate a (minimal) zone together with the known expected number of outliers.

V. COMPARISON TO CHORD

Some of the basic properties of Chord zone distribution – derived from the model of random points on unit circle – are summarized in Table I for comparison purposes.

Zone size	Chord	Kademlia
Average	$1/n$	$1/n$
Variance	$(n-1)/(n^2+n^3)$	$\approx 0.525/n^2$
Minimal	$1/n^2 (= 2^{-2 \log_2 n})$	$\geq 2^{-\lfloor \log_2 n + \sqrt{2 \log_2 n} - \frac{3}{2} \rfloor}$
Jain's fairness	$0.5 + 1/n$	0.655

Table I: Main parameters of the zone size distributons

The full derivation of these results are available in the literature (section II-B). Only a small summary is presented here for comparison purposes.

Cumulative Density Function (CDF) of zone sizes in Chord:

$$\mathrm{P_n}(T \leq x) = \mathrm{F_n}(x) = 1 - (1-x)^{n-1} \ (0 \leq x \leq 1)$$

Probability density function (PDF) of zone sizes in Chord:

$$\mathrm{f_n}(x) = \mathrm{F_n}'(x) = (n-1)(1-x)^{n-2} \ (0 \leq x \leq 1)$$

Expected value (average zone size), as nodes share the whole hash space this result is the same as for Kademlia:

$$\mathrm{E}_n(x) = 1/n \ (= \int_0^1 x(n-1)(1-x)^{n-2} \, dx)$$

[7]The expected number of nodes at a given level of the PATRICIA trie

Variance of zone sizes:

$$\mathrm{V}_n(x) = \int_0^1 \left(x - \frac{1}{n}\right)^2 (n-1)(1-x)^{n-2} \, dx =$$

$$= \left[\frac{(n-1)(1-x)^n \left(1 - 2x + x^2 n^2\right)}{n^2(n+1)(x-1)} \right]_0^1 = \frac{n-1}{n^2+n^3}$$

This is approaching $1/n^2$ for large n, so Kademlia have a constant factor advantage here.

It follows that the Jain's fairness index of the zone sizes in Chord is:

$$\mathcal{J}_T = \frac{1}{n^2 \left(\frac{n-1}{n^2+n^3} + \frac{1}{n^2} \right)} = \frac{1}{2} + \frac{1}{2n}$$

This is 0.5 asymptotically, which is less (worse) than the result of Kademlia (0.655).

Cumulative Density Function (CDF) for the minimal zone in Chord:

$$\mathrm{P}_n^{\min}(T_{\min} \leq x) = \mathrm{F}_n^{\min}(x) = 1 - (1-nx)^{n-1} \ (0 \leq x \leq \frac{1}{n})$$

Probability density function (PDF) for the minimal zone size in Chord:

$$\mathrm{f}_\mathrm{n}^{\min}(x) = \mathrm{F}_\mathrm{n}^{\min'}(x) = (n^2-n)(1-nx)^{n-2} \ (0 \leq x \leq \frac{1}{n})$$

Expected value of the minimal zone size in Chord:

$$\mathrm{E}_n^{\min}(x) = 1/n^2 \, (= \int_0^{\frac{1}{n}} x(n^2-n)(1-nx)^{n-2} \, dx)$$

Variance of the minimal zone size in Chord:

$$\mathrm{V}_n^{\min}(x) = \int_0^{\frac{1}{n}} \left(x - \frac{1}{n^2}\right)^2 (n^2-n)(1-nx)^{n-2} \, dx =$$

$$= \frac{n-1}{n^4+n^5}$$

For the maximal zone size distribution refer to the asymptotic results of Darling [24]. We have not provided results for the maximal zone size distribution of Kademlia.

VI. CONCLUSIONS

A direct connection between Kademlia zone size distribution and the shape parameters of a PATRICIA trie built from the overlay addresses of the DHT have been presented. By relying on existing literature on random symmetric binary PATRICIA tries, we have derived some of the key parameters of the zone size distribution of Kademlia DHT.

The exact distribution of zone sizes and minimal zone sizes of Kademlia have been provided. We have also provided approximative characterisation of the moments of the zone size distribution based on the literature on PATRICIA tries.

By comparing the zone size distribution of Kademlia and Chord, we have concluded that Kademlia zones are distributed more uniformly. Zone sizes have a smaller deviation by a constant factor, and the minimal zone size is larger compared to

Chord. Kademlia zone size distribution is also fairer compared to Chord's by the measure of Jain's fairness index.

The consequence is that in general cases Kademlia achieves a more fair distribution of data than Chord, and this suggests that it may show a more uniform distribution of routing load too. Finally the results about the size of the minimal zone in Kademlia open the possibility to improve upon the random node sampling algorithm of the authors [6].

Zoltán Novák received an M.Sc. degree in software engineering at Budapest University of Technology and Economics (BME) in 2006. Between 2006 and 2010 he was a Ph.D. student at BME department of Telecommunications and Media Informatics. His main research topics are peer to peer systems and overlay networking. Since 2011 he has been working at Ericsson Hungary. Currently he is pursuing Ph.D. degree under the supervision of Zoltán Pap.

Zoltán Pap received an M.Sc. degree in Electrical Engineering at Budapest University of Technology and Economics (BME) in 2000, and an M.Sc. degree in Business Administration at Corvinus University of Budapest in 2002. From 2000 to 2006 he worked on various research fields including telecommunication networks and protocols, grid computing, peer to peer networks, model-based software development and testing at BME department of Telecommunications and Media Informatics. He received a Ph.D degree at BME in 2006. Since 2007 he has been working at Ericsson initially as systems engineer later as product manager and functional manager on several products such as the Telecom Server Platform (TSP), Ericsson's Operations Support System Radio and Core (OSS-RC) and the Smart Services Router (SSR).

REFERENCES

[1] P. Maymounkov and D. Mazières, "Kademlia: A peer-to-peer information system based on the xor metric," in *IPTPS '01: Revised Papers from the First International Workshop on Peer-to-Peer Systems*, (London, UK), pp. 53–65, Springer-Verlag, 2002.

[2] R. Morris, D. Karger, F. Kaashoek, and H. Balakrishnan, "Chord: A Scalable Peer-to-Peer Lookup Service for Internet Applications," in *ACM SIGCOMM 2001*, (San Diego, CA), September 2001.

[3] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Schenker, "A scalable content-addressable network," in *Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications (SIGCOMM '01)*, vol. 31, pp. 161–172, ACM Press, October 2001.

[4] D. Karger, E. Lehman, T. Leighton, R. Panigrahy, M. Levine, and D. Lewin, "Consistent hashing and random trees: Distributed caching protocols for relieving hot spots on the world wide web," in *Proceedings of the Twenty-ninth Annual ACM Symposium on Theory of Computing*, STOC '97, (New York, NY, USA), pp. 654–663, ACM, 1997.

[5] V. King, S. Lewis, J. Saia, and M. Young, "Choosing a random peer in chord," *Algorithmica*, vol. 49, no. 2, pp. 147–169, 2007.

[6] Z. Novák and Z. Pap, "Random node sampling in kademlia," in *6th International ICST Conference on Broadband Communications, Networks, and Systems*, IEEE, 11 2009.

[7] C. Scott and L. J. Saia, "Scalable byzantine agreement," tech. rep., 2004.

[8] M. D. Mitzenmacher, *The power of two choices in randomized load balancing*. PhD thesis, 1996. Chair-Alistair Sinclair.

[9] R. C. Rumín, M. Urueña, and A. Banchs, "Routing fairness in chord: Analysis and enhancement.," in *INFOCOM*, pp. 1449–1457, IEEE, 2009.

[10] X. Wang and D. Loguinov, "Load-balancing performance of consistent hashing: asymptotic analysis of random node join," *Networking, IEEE/ACM Transactions on*, vol. 15, no. 4, pp. 892–905, 2007.

[11] X. S. Cai and L. Devroye, "A probabilistic analysis of kademlia networks," in *Algorithms and Computation*, pp. 711–721, Springer, 2013.

[12] A. Gnedin, B. Hansen, J. Pitman, *et al.*, "Notes on the occupancy problem with infinitely many boxes: general asymptotics and power laws," *Probability surveys*, vol. 4, pp. 146–171, 2007.

[13] T. Emigh, "On the number of observed classes from a multinomial distribution," *Biometrics*, pp. 485–491, 1983.

[14] H.-K. Hwang and S. Janson, "Local limit theorems for finite and infinite urn models," *The Annals of Probability*, pp. 992–1022, 2008.

[15] P. Jacquet and W. Szpankowski, "Analytical depoissonization and its applications," *Theoretical Computer Science*, vol. 201, no. 1-2, pp. 1–62, 1998.

[16] D. R. Morrison, "Patricia - practical algorithm to retrieve information coded in alphanumeric.," *J. ACM*, vol. 15, no. 4, pp. 514–534, 1968.

[17] B. Rais, P. Jacquet, and W. Szpankowski, "A limiting distribution for the depth in patricia tries," 1990.

[18] C. Knessl and W. Szpankowski, "Limit laws for the height in patricia tries," *Journal of Algorithms*, vol. 44, no. 1, pp. 63–97, 2002.

[19] L. Devroye, "Laws of large numbers and tail inequalities for random tries and patricia trees," *Journal of Computational and Applied Mathematics*, vol. 142, no. 1, pp. 27–37, 2002.

[20] A. Magner, C. Knessl, and W. Szpankowski, "Expected external profile of PATRICIA tries," in *2014 Proceedings of the Eleventh Workshop on Analytic Algorithmics and Combinatorics, ANALCO 2014, Portland, Oregon, USA, January 6, 2014*, pp. 16–24, 2014.

[21] H. Prodinger, "Compositions and patricia tries: no fluctuations in the variance," SODA, 2004.

[22] J. Bourdon, "Size and path length of patricia tries: dynamical sources context," *Random Structures & Algorithms*, vol. 19, no. 3-4, pp. 289–315, 2001.

[23] L. Devroye, "A study of trie-like structures under the density model," *The Annals of Applied Probability*, pp. 402–434, 1992.

[24] D. A. Darling, "On a class of problems related to the random division of an interval," *Ann. Math. Statist.*, vol. 24, pp. 239–253, 06 1953.

The Use of Software-Defined Radio Systems in Multilateral Navigation Radio Systems

G. M. Mashkov, E. G. Borisov, A. G. Vladyko, and A. I. Gomonova

Abstract — The paper describes an optional application of software-defined radio systems technology in multilateral range finding systems for solving tasks of determining the flying objects coordinates. The design feature of the described system is the use of cooperative processing of range measurements aggregate to improve accuracy of object positioning.

***Keywords* — multilateral, range measurements, cooperative processing, mean square error (MSE), and software-defined radio systems (SDR).**

I. INTRODUCTION AND PROBLEM DEFINITION

The last decade has shown a significant growth in passenger and cargo air traffic, as well as a considerable increase of flights of privately owned aircraft. This leads to growing of air traffic density, overloading of airfield areas and flight routes. Besides, constantly tightened security requirements place higher demand on the accuracy of flying objects positioning in the shortest possible period. That is the reason why the developing of multi-position navigation radio stations is being carried out worldwide nowadays (for range finding only as well as for range finding and multilateration – MLAT - together).

Multilateral radio systems represent an independent cooperative navigation system of a new type, combining in single unit measurement subsystems, means of communication, data transfer, and computing devices. An example of such system is the MLAT system developed by Czech company Era [1]. Successful achievements in this field have been achieved by the French multinational group "Thales" [2]. The multilateral surveillance system "Mera", developed by NIIRA JSC [3], is actively used in this country. Multilateral navigation system, developed by Australian company "Locata", incorporates pinpoint accuracy characteristic, working in the 2.4 GHz frequency range [4]. General requirements for multilateral systems are given in [5].

The need for the emergence of such systems arrived due to the fact that the existing satellite navigation systems (GPS, GLONASS, and prospective European GALILEO system), have the following main disadvantages: low resistance when exposed to electronic interference, low signal value, complexity of working indoors, as well as in areas of dense urban development, in the mountain gorges, etc. Furthermore, the aforementioned systems lack positioning accuracy in urban areas and to the North of the 60° parallel. The positioning error of GPS/GLONASS can reach over 30 m and more. The main advantages of MLAT systems in comparison with single position systems are the following: the possibility to form spatial view areas of complex configuration with a given overlap ratio, the ability to control and redistribute energy within the system, precise accuracy of flying objects positioning, the ability to measure objects complete velocity vector, etc [6, 7].

Manuscript received May 12, 2015, revised June 12, 2015. The authors are with the Bonch-Bruevich Saint Petersburg State University of Telecommunications, 22 Prospekt Bolshevikov, Saint Petersburg, Russia (e-mail: begspb1967@mail.ru).

II. BASIC PRINCIPLES OF MLAT SYSTEM DESIGN. NI USRP SOFTWARE-DEFINED RADIO SYSTEMS TECHNOLOGY

For mass adoption, MLAT systems need to have a low cost of installation with minimum operating costs, small size combined with low power consumption by using different power supply, easy to build up, update and reconfigurable hardware platform.

It is necessary to allow the operation of navigation equipment developed in conjunction with the systems used in the management of air traffic, such as ADS-B (Automatic dependent surveillance-broadcast) and their modifications, enabling the pilots in the cockpit and air traffic controllers on the ground point to observe the traffic of aircraft movements with greater accuracy and receive aeronautical information.

The instruments of software-defined radio systems could be used as prospective MLAT system transceiver modules. It would allow carrying on the tasks of generating signals of any modulation type, range finding, communication with an object being located, an exchange of information between modules, synchronization of functioning modes of the modules, and optimization of frequencies allocation inside the system etc. Software-defined radio system is a radio communication system in which the functions of the main instruments are implemented by software solutions. These instruments can include filters, amplifiers, modulators or/and demodulators. As soon as these instruments are configurable by software only, there is a possibility of modifying such a system without any significant changes in the hardware configuration. When using SDR, almost the entire volume of work on signal processing is shifted to the software that can run on digital signal processors or special DSP-purpose high-speed PLD. The main reason for such an approach is to create a system that can receive and transmit radio signals in a given frequency range and easily select the desired modulation law [10].

National Instruments Corporation proposes its own solution - NI USRP (NI Universal Software Radio Peripheral). It is a

HF transceiver controlled by a computer. The company offers the functional abilities of a graphical development environment NI LabVIEW for configuring SDR platform. Due to the programmability of the measuring equipment, this is a unique possibility to generate periodic test signals, depending on navigation receiver trajectory, i.e. there is no need for costly tests with participation of air traffic control. The NI USRP software contains a set of functions implemented in the form of virtual instruments (VIs for LabVIEW) to control one or more platforms for USRP. At the higher level, NI USRP driver proposes to use Vis for session opening, configuring of hardware, performance of read/write operations, and session closing. The basic principle of programming is the creation of virtual objects: a satellite, positions of a receiver and an HF generator. Each object is being operated by a link specially designated for it in the software. All of the properties, status and control are exercised by using functional tools incorporated into the set of built-in libraries for visualization [11].

Figure 1 shows a scheme of the MLAT navigation system, consisting of N ground-based transceiver points (GTP); each point emitting a broadband signal on a lettered frequency. Onboard equipment contains a multi-channel transceiver, receiving signals from GTP, and then relaying them. Each of the spatially separated GTPs receives relayed signals due to the request from each position, thus forming range-finding and summarized range-finding measurements. One of the GTPs is nominated as a primary point and takes all the measurements from the others (thus implementing cooperative processing), calculates estimates of rectangular coordinates and speed of their change.

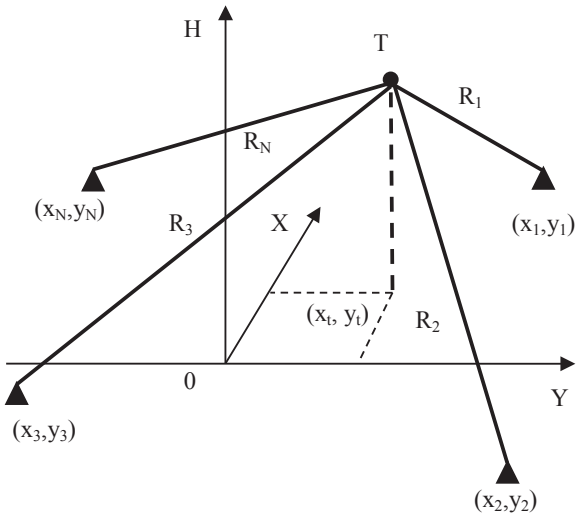


Fig.1 Layout of flying object positioning

For this case let us write a system of linear algebraic equations (SLAE), which takes into account the number of direct x - N and summarized N(N-1) measurements - , forming N^2 measurements, i.e.:

$$\begin{cases} \begin{cases} \hat{R}_1 = 1\cdot R_1 + 0\cdot R_2 + 0\cdot R_3 + \ldots + 0\cdot R_N \\ \hat{R}_2 = 0\cdot R_1 + 1\cdot R_2 + 0\cdot R_3 + \ldots + 0\cdot R_N \\ \vdots \\ \hat{R}_N = 0\cdot R_1 + 0\cdot R_2 + 0\cdot R_3 + \ldots + 1\cdot R_N \end{cases} N\ meas. \\ \begin{cases} \hat{R}_{\Sigma 12} = 1\cdot R_1 + 1\cdot R_2 + 0\cdot R_3 + \ldots + 0\cdot R_N \\ \hat{R}_{\Sigma 21} = 1\cdot R_1 + 1\cdot R_2 + 0\cdot R_3 + \ldots + 0\cdot R_N \\ \hat{R}_{\Sigma 13} = 1\cdot R_1 + 0\cdot R_2 + 1\cdot R_3 + \ldots + 0\cdot R_N \\ \hat{R}_{\Sigma 31} = 1\cdot R_1 + 0\cdot R_2 + 1\cdot R_3 + \ldots + 0\cdot R_N \\ \vdots \\ \hat{R}_{\Sigma N(N-1)} = 0\cdot R_1 + 0\cdot R_2 + 0\cdot R_3 + \ldots + 1\cdot R_{N(N-1)} \end{cases} N(N-1)\,meas. \end{cases} \quad (1)$$

$\hat{R}_1, \hat{R}_2 \ldots \hat{R}_N$ - primary measuring the slant range;

$\hat{R}_{12}, \hat{R}_{21} \ldots \hat{R}_{N(N-1)}$ - the primary measure summary ranges.

Equations (1) contain measurements relative to N^2 estimated parameters that allow us to implement their solution using the least-squares method [8,9]:

$$\tilde{X} = \left[\left(A^T \Lambda W^{-1} A\right)^{-1} A^T \Lambda W^{-1}\right] H, \quad (2)$$

where A is the matrix of dimension NxN^2, consisting of zeros and ones, where "1" indicates the presence of corresponding dimension, and "0" its absence.

So, for three-position system, this matrix has the following form:

$$A^T = \begin{Vmatrix} 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 \end{Vmatrix}, \quad (3)$$

- matrix system condition, taking into account the totality of the processed measurements;

$H^T = \left\|\hat{R}_1, \hat{R}_2 \ldots \hat{R}_N, \hat{R}_{\Sigma 12}, \hat{R}_{\Sigma 21}, \hat{R}_{\Sigma 13}, \hat{R}_{\Sigma 31} \ldots \hat{R}_{\Sigma N(N-1)}\right\|$ - line vector of the estimated parameters (vector row of primary measurements);

$$W = \begin{Vmatrix} \sigma_R^2 & 0 & 0 & 0 \\ 0 & \sigma_R^2 & 0 & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_{R\Sigma}^2 \end{Vmatrix} \quad (4)$$

W is a precision matrix of dimension N^2 x N^2 containing the variance of range-finding errors and the sums of the ranges:

$$\Lambda = \begin{Vmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 1 \end{Vmatrix} \quad (5)$$

Λ is a diagonal matrix of coefficients, the diagonal element of which is equal to one if the measurement is present (or is used for measurements) and equal to zero if the measurement is not used.

Error covariance matrix for the system (2) is defined as [8]

$$K_{\tilde{X}} = \left(A^T W^{-1} A\right)^{-1} \sigma_0^2 = diag\left(A^T A\right)^{-1} \sigma_0^2, \quad (6)$$

where $\sigma_0^2 = \sigma_R^2 = \sigma_\Sigma^2$ - is the variance of measurement error of the range-finding parameter.

Therefore, for range-finding and summarized range-finding systems the expression of angle ranges and variances of the measurement errors obtained with (2) and (6) will take the following form:

- for three positions systems

$$\begin{aligned}
\tilde{R}_1 &= \frac{7}{27}\hat{R}_1 - \frac{2}{27}\hat{R}_2 - \frac{2}{27}\hat{R}_3 + \frac{5}{27}\hat{R}_{\Sigma 12} + \frac{5}{27}\hat{R}_{\Sigma 21} + \frac{5}{27}\hat{R}_{\Sigma 13} + \\
&\frac{5}{27}\hat{R}_{\Sigma 31} - \frac{4}{27}\hat{R}_{\Sigma 23} - \frac{4}{27}\hat{R}_{\Sigma 32} \\
\tilde{R}_2 &= -\frac{2}{27}\hat{R}_1 + \frac{7}{27}\hat{R}_2 - \frac{2}{27}\hat{R}_3 + \frac{5}{27}\hat{R}_{\Sigma 12} + \frac{5}{27}\hat{R}_{\Sigma 21} - \frac{4}{27}\hat{R}_{\Sigma 13} - \\
&\frac{4}{27}\hat{R}_{\Sigma 31} + \frac{5}{27}\hat{R}_{\Sigma 23} + \frac{5}{27}\hat{R}_{\Sigma 32} \\
\tilde{R}_3 &= -\frac{2}{27}\hat{R}_1 - \frac{2}{27}\hat{R}_2 + \frac{7}{27}\hat{R}_3 - \frac{4}{27}\hat{R}_{\Sigma 12} - \frac{4}{27}\hat{R}_{\Sigma 21} + \frac{5}{27}\hat{R}_{\Sigma 13} + \\
&\frac{5}{27}\hat{R}_{\Sigma 31} + \frac{5}{27}\hat{R}_{\Sigma 23} + \frac{5}{27}\hat{R}_{\Sigma 32}
\end{aligned} \quad (7)$$

$$\sigma_{RC}^2 = diag\left(A^T A\right)\sigma_R^2 = diag\begin{Vmatrix} \frac{7}{27} & -\frac{2}{27} & -\frac{2}{27} \\ -\frac{2}{27} & \frac{7}{27} & -\frac{2}{27} \\ -\frac{2}{27} & -\frac{2}{27} & \frac{7}{27} \end{Vmatrix}\sigma_R^2, \quad (8)$$

- for four positions system:

$$\begin{aligned}
\tilde{R}_1 &= \frac{11}{65}\hat{R}_1 + \frac{9}{65}\hat{R}_{\Sigma 12} + \frac{9}{65}\hat{R}_{\Sigma 13} + \frac{9}{65}\hat{R}_{\Sigma 14} - \frac{2}{65}\hat{R}_2 + \frac{9}{65}\hat{R}_{\Sigma 21} - \frac{4}{65}\hat{R}_{\Sigma 23} - \frac{4}{65}\hat{R}_{\Sigma 24} - \\
&- \frac{2}{65}\hat{R}_3 + \frac{9}{65}\hat{R}_{\Sigma 31} - \frac{4}{65}\hat{R}_{\Sigma 32} - \frac{4}{65}\hat{R}_{\Sigma 34} - \frac{2}{65}\hat{R}_4 - \frac{4}{65}\hat{R}_{\Sigma 43} - \frac{4}{65}\hat{R}_{\Sigma 42} + \frac{9}{65}\hat{R}_{\Sigma 41} \\
\tilde{R}_2 &= -\frac{2}{65}\hat{R}_1 + \frac{9}{65}\hat{R}_{\Sigma 12} - \frac{4}{65}\hat{R}_{\Sigma 13} - \frac{4}{65}\hat{R}_{\Sigma 14} + \frac{11}{65}\hat{R}_2 + \frac{9}{65}\hat{R}_{\Sigma 21} + \frac{9}{65}\hat{R}_{\Sigma 23} + \\
&\frac{9}{65}\hat{R}_{\Sigma 24} - \frac{2}{65}\hat{R}_3 - \frac{4}{65}\hat{R}_{\Sigma 31} + \frac{9}{65}\hat{R}_{\Sigma 32} - \frac{9}{65}\hat{R}_{\Sigma 34} - \frac{2}{65}\hat{R}_4 - \frac{4}{65}\hat{R}_{\Sigma 43} + \\
&\frac{9}{65}\hat{R}_{\Sigma 42} - \frac{4}{65}\hat{R}_{\Sigma 41} \\
\tilde{R}_3 &= -\frac{2}{65}\hat{R}_1 - \frac{4}{65}\hat{R}_{\Sigma 12} + \frac{9}{65}\hat{R}_{\Sigma 13} - \frac{4}{65}\hat{R}_{\Sigma 14} - \frac{2}{65}\hat{R}_2 - \frac{4}{65}\hat{R}_{\Sigma 21} + \frac{9}{65}\hat{R}_{\Sigma 23} - \frac{4}{65}\hat{R}_{\Sigma 24} + \\
&+ \frac{11}{65}\hat{R}_3 + \frac{9}{65}\hat{R}_{\Sigma 31} + \frac{9}{65}\hat{R}_{\Sigma 32} + \frac{9}{65}\hat{R}_{\Sigma 34} - \frac{2}{65}\hat{R}_4 + \frac{9}{65}\hat{R}_{\Sigma 43} - \frac{4}{65}\hat{R}_{\Sigma 42} - \frac{4}{65}\hat{R}_{\Sigma 41} \\
\tilde{R}_4 &= -\frac{2}{65}\hat{R}_1 - \frac{4}{65}\hat{R}_{\Sigma 12} - \frac{4}{65}\hat{R}_{\Sigma 13} + \frac{9}{65}\hat{R}_{\Sigma 14} - \frac{2}{65}\hat{R}_2 - \frac{4}{65}\hat{R}_{\Sigma 21} - \frac{4}{65}\hat{R}_{\Sigma 23} + \frac{9}{65}\hat{R}_{\Sigma 24} - \\
&- \frac{2}{65}\hat{R}_3 - \frac{4}{65}\hat{R}_{\Sigma 31} - \frac{4}{65}\hat{R}_{\Sigma 32} + \frac{9}{65}\hat{R}_{\Sigma 34} + \frac{11}{65}\hat{R}_4 + \frac{9}{65}\hat{R}_{\Sigma 43} + \frac{9}{65}\hat{R}_{\Sigma 42} + \frac{9}{65}\hat{R}_{\Sigma 41}
\end{aligned} \quad (9)$$

$$\sigma_{RC}^2 = diag\left(A^T A\right)\sigma_R^2 = diag\begin{Vmatrix} \frac{11}{65} & -\frac{2}{65} & -\frac{2}{65} & -\frac{2}{65} \\ -\frac{2}{65} & \frac{11}{65} & -\frac{2}{65} & -\frac{2}{65} \\ -\frac{2}{65} & -\frac{2}{65} & \frac{11}{65} & -\frac{2}{65} \\ -\frac{2}{65} & -\frac{2}{65} & -\frac{2}{65} & \frac{11}{65} \end{Vmatrix}\sigma_R^2 \quad (10)$$

The rectangular coordinates of an object are determined by solving the system of equations

$$R = \sqrt{\left(x_t - x_i\right)^2 + \left(y_t - x_i\right)^2 + \left(h_t - h_i\right)^2}, \; i = 1 \div N, \quad (11)$$

x_t, y_t, h_t - desired object coordinates;

x_i, y_i, h_i - known coordinates of GTPs.

The accuracy of estimating the location of an object we define by the dependence:

$$\sigma = \sqrt{tr\left(D^T W^{-1} D\right)^{-1}} \quad (12)$$

Where : tr - trace - the sum of diagonal elements of a matrix;

$$D = \begin{Vmatrix} \frac{\partial R_1}{\partial x_t} & \frac{\partial R_1}{\partial y_t} & \frac{\partial R_1}{\partial h_t} \\ \frac{\partial R_2}{\partial x_t} & \frac{\partial R_2}{\partial y_t} & \frac{\partial R_2}{\partial h_t} \\ \vdots & & \\ \frac{\partial R_N}{\partial x_t} & \frac{\partial R_N}{\partial y_t} & \frac{\partial R_N}{\partial h_t} \end{Vmatrix} \text{ - conversion matrix.}$$

III. RESULTS OF CALCULATIONS. SCHEME OF THE EXPERIMENTS

Figures 2 and 3 show, as an example, the values of MSE for the range-finding of the objects driven in a circle with a radius of 200 km around the origin of coordinates. Figure 2 shows the MSE for range-finding for the normal law of error distribution with σ=20 m at zero expectation value, and Figure 3 - for a uniform law of error distribution with a maximum error ΔR=± 60 m.

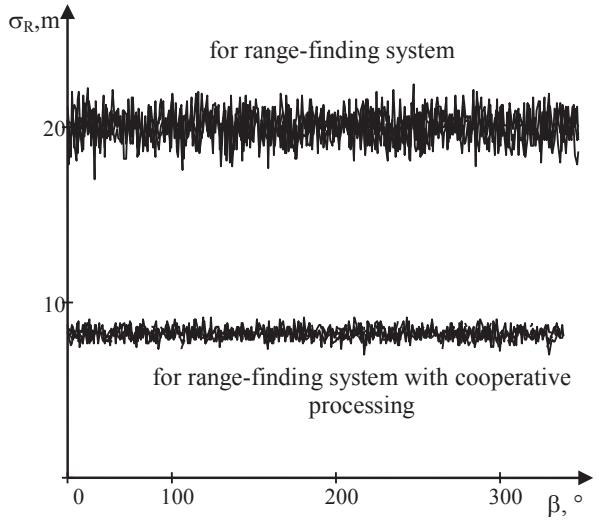


Figure 2 – MSE for range-finding of the object for different methods of measurements processing (for normal distribution)

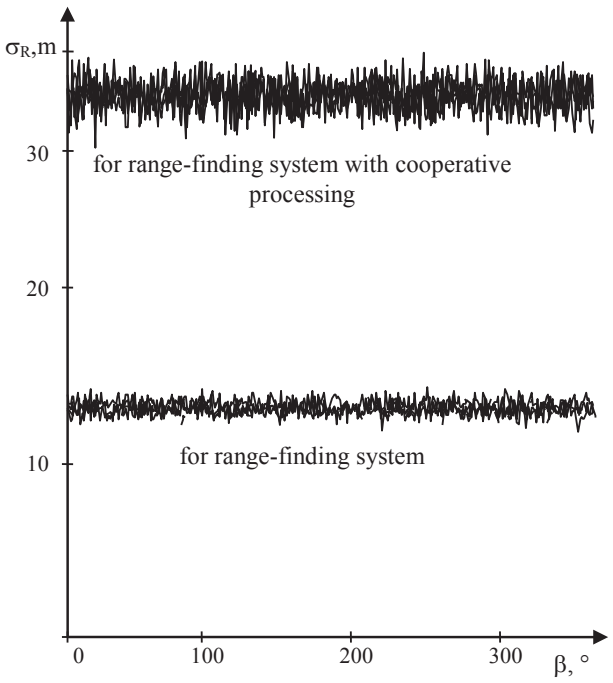


Figure 3 - MSE for range-finding of the object for different methods of measurements processing (for uniform distribution)

Figure 4 shows MSE for positioning of object driven in a circle with a radius of 200 km around the origin of coordinates with ground transceivers stationed as a distance of 20 km from the origin.

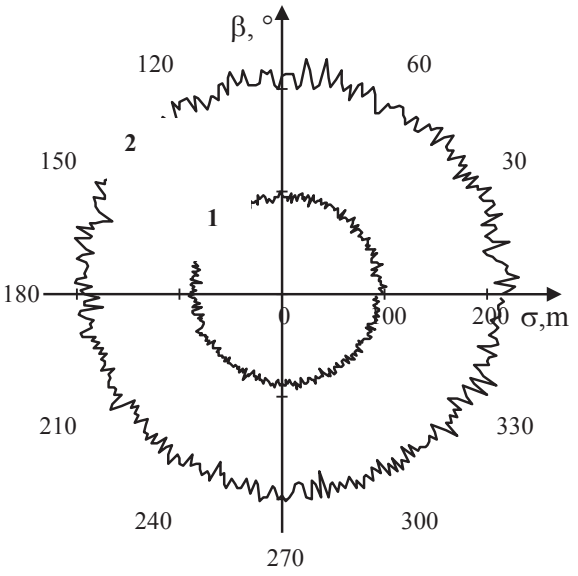


Figure 4 – MSE for positioning of the object for different methods of measurements processing (1 – for range-finding system, 2 – for range-finding system with cooperative processing

Fig. 5 shows MSE for positioning of object driven in a circle at a distance of 20 km around the origin of coordinates with GTPs stationed at a distance of 200 km from the origin.

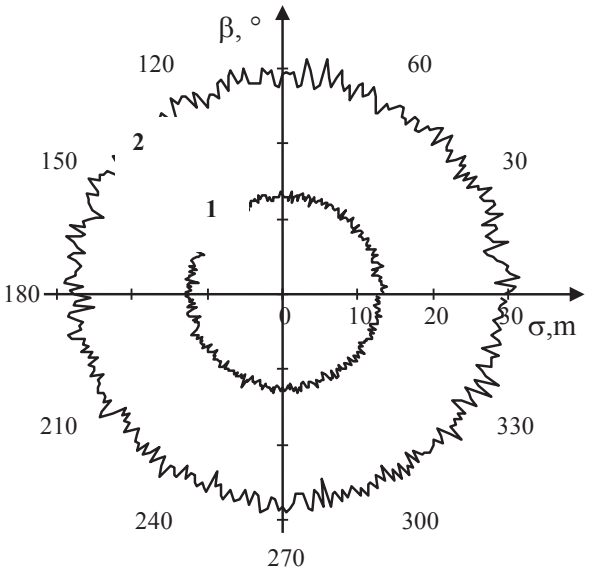


Fig. 5 - MSE for positioning of the object for different methods of measurements processing (1 – for range-finding system, 2 – for range-finding system with cooperative processing)

Fig. 6 shows, as an example, the object positioning MSE for normal distribution of range-finding and sums of ranges when the distance to the object relative to the origin is 200 km and the GTSs are located 20 km from the origin.

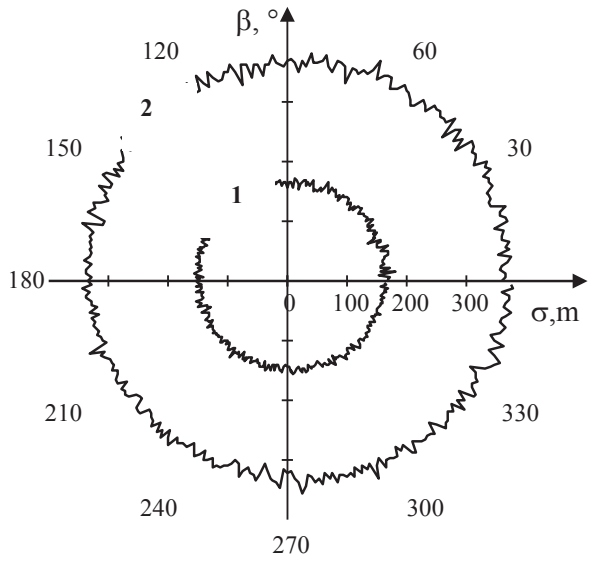


Fig. 6 - MSE for positioning of the object for different methods of measurements processing (1 – for range-finding system, 2 – for range-finding system with cooperative processing)

Fig. 7 shows, as an example, MSE for positioning of the object at uniform distribution of positioning, range-finding and sums of distances errors when the object is 20 km from the coordinates origin, and GTSs are 200 km from the origin.

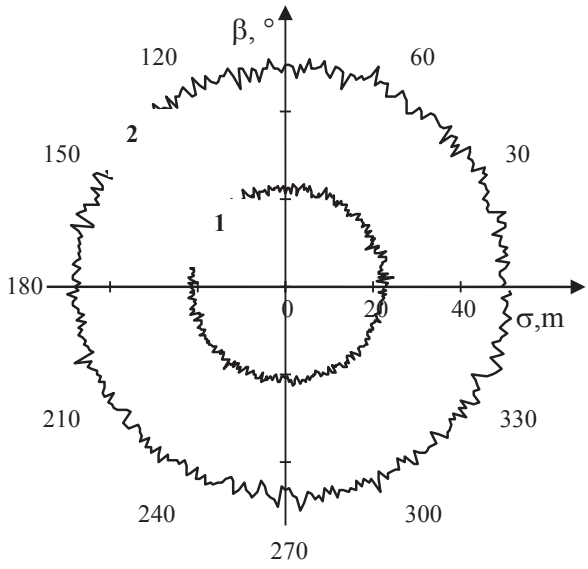


Fig. 7 – MSE for object positioning for different methods of measurements processing (1 – for range-finding system, 2 – for range-finding system with cooperative processing)

Fig. 8 shows MSE for the case of $\sigma_{R\Sigma} = 2\sqrt{2}\sigma_R$. And in case of such MSE values, the errors of positioning grow insignificantly (curves 1 and 2 respectively). Let us assume that from one round of measurements to the other round for some reason or another there is a lack of range measurements and sums of ranges do not exist, i.e. in matrix Λ (formula 5) in 50% of positions we see zeros. Meanwhile, using a conventional range-finding system, indirect measurement of rectangular coordinates in not possible to produce. But in the case of cooperative processing it is still possible to measure rectangular coordinates, but with reduced accuracy and, therefore, to provide algorithmic and informational stability of the system.

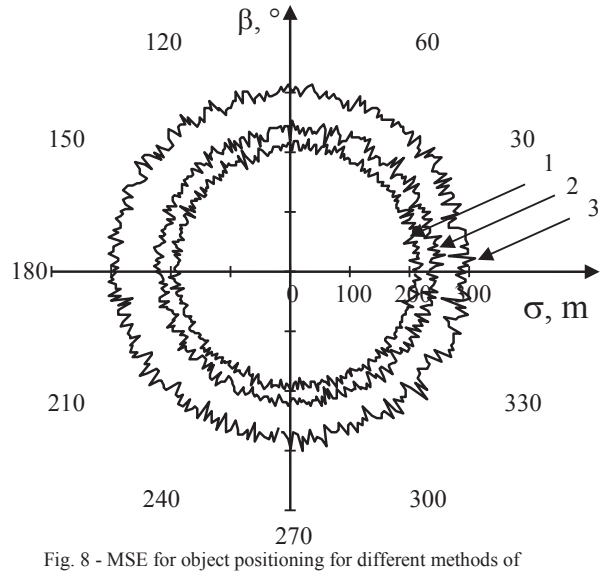


Fig. 8 - MSE for object positioning for different methods of measurements processing (1 – for range-finding system, 2 – for range-finding system with cooperative processing, 3 - for range-finding system with cooperative processing at incomplete set of measurements)

The experiment was made on the NI USRP platform base, which contained four instruments, three of which worked as ground base stations, and the fourth imitated an airborne transponder.

IV. CONCLUSIONS

The aforementioned values of the errors of flying object positioning determination show that cooperative processing of positioning information with respect to the four positional system improves the accuracy of determining the flying object location more than two times in comparison with the conventional range-finding method. In so doing, the high positioning accuracy is achieved by one cycle of data processing within the system. The use of algorithms for filtering trajectory messages will further improve the accuracy of determining the coordinates, without restrictions on the flying object movement hypothesis. This extends the applicability of the proposed options for positioning of maneuvering objects.

This work was done with the financial support of the Ministry of Education and Science of the Russian Federation as a part of an applied scientific research on lot ciphered 2014-14-579-0112 and related to the program "The development of experimental model of multi-position fast deployment self-contained radar system ground infrastructure for landing aircraft on unprepared ground" (application cipher "2014-14-579-0112-030").

REFERENCES

[1] Viktor Sotona, ATM Solutions. (2011, Sept.). Implementation of MLAT/ADS-B Systems. Presented at ICAO/FAA Workshop on ADS-B and Multilateration Implementation. Available: http://www.icao.int/NACC/Documents/Meetings/2011/ADSBMLT/Day01-05-ERA-Sotona.pdf
[2] Thales Air Systems Gmbh. (2012, Nov.). Ground Stations and multilateration. Presented at Asia-Pacific Seminar. Available: http://www.icao.int/APAC/Meetings/2012_SEA_BOB_ADSB_WG8/SP03_Thales%20ADS-B%20Multilateration.pdf
[3] VNIIRA. Saint-Petersburg, Russia. Product: Multilateration System "Mera". In russian language. Available: http://www.vniira.ru/ru/products/790/811/1179?text=basic-purpose
[4] Stephen Shankland. Locata wants to fill holes in GPS location, navigation. (2012, Dec.). Available: http://www.cnet.com/news/locata-wants-to-fill-holes-in-gps-location-navigation/
[5] ICAO. Multilateration. Concept of use. (2007, Sept.). Available: http://www.icao.int/APAC/Documents/edocs/cns/mlat_concept.pdf
[6] V.S. Kondratiev, *Multipositional radio systems*. Moscow, Russia: Radio and communications, 1986, pp. 178-200, 246-251.
[7] V.S. Cheryak. *Multipositional radiolocation*. Moscow, Russia: Radio and communications, 1993, pp. 7-42.
[8] E.G. Borisov, G.M. Mashkov, L.S. Turnetskiy, "Accuracy increase in determing an object's coordinates while co-operative processing is applied in multipositional radiolocation system", *Radiotechnics №5*, pp. 4-9, May, 2013.
The method for determing an object's coordinates in three positional rangefinder radiolocation system, by E.G. Borisov, G.M. Mashkov. (2014, May 10). *Patent RF № 2515571MIIK G01S13/46* [Online]. Available: http://www.freepatent.ru/patents/2515571
[9] J. Mitola II, Z. Zvonar. *Software Radio Technologies*. New York: Wiley- IEEE Press, 2001, pp. 40-47.
[10] J. Mitola II. *Software Radio Architecture: Object-oriented Approaches to Wireless System*. New York: John Wiley & Sons, 2000, pp. 384-436.

Georgiy M. Mashkov was born in 1954, in Uzlovaya, a town and the administrative center of Uzlovsky District in Tula Oblast, Russia. He obtained a Radio Engineer degree in 1977, a Candidate of Sciences degree in 1984, and a Doctor of Technical Sciences degree in 1993. Dr Mashkov is the author of over 150 scientific publications, including 26 of inventions and useful models. Currently he is first Vice Rector (Vice Rector for Academic Affairs) of Saint-Petersburg State University of Telecommunications Named after Professor M. A. Bonch-Bruevich, Saint-Petersburg, Russia. His area of scientific interests include MLAT, and methods of mathematical processing of measurement information of different physical nature.

Evgeny G. Borisov was born in 1967 in Novosibirsk. He obtained the degree of a Radio Engineer in 1990 and a Candidate of Technical Sciences degree in 2000. Presently he is a leading researcher of Saint-Petersburg State University of Telecommunications Named after Professor M. A. Bonch-Bruevich, Saint-Petersburg, Russia.. Dr Borisov is the author of three monographs, more than 120 articles, and 25 patents. Area of scientific interests include MLAT, integration of positioning data, statistical synthesis of radio systems.

Andrey Vladyko obtained his Candidate of Science degree from the Komsomolsk-on-Amur State Technical University, Russia in 2001. Presently he is Head of the Office of Research and Research Training at The Bonch-Bruevich Saint-Petersburg State University of Telecommunications, Saint-Petersburg, Russia. His major interests include control systems, soft computing, network security management. Dr. Vladyko is a Member, IEEE.

Anna Gomonova was born in Saint-Petersburg, Russia in 1991. She obtained her BSc and MSc degrees in radio engineering from Saint-Petersburg State Electrotechnical University, Russia in 2014. Presently she is an engineer at The Bonch-Bruevich Saint-Petersburg State University of Telecommunications, Saint-Petersburg, Russia. Her major interests include MLAT, digital signal processing, software-defined radio.

Editor's Note:
IEEE Communications Society as a Sister Society agreement with HTE (Scientific Association for Infocommunications, Hungary).
The terms of the agreement include re-publication of articles of IEEE Communications Society publications
in HTE's Infocommunications Journal.
The article below has already appeared in IEEE Communications Magazine (Vol. 52, no. 8, 2014, pp. 29–35.)

Opportunities in Mobile Crowd Sensing

Huadong Ma, Dong Zhao, and Peiyan Yuan

ABSTRACT

Mobile crowd sensing is a new paradigm that takes advantage of pervasive mobile devices to efficiently collect data, enabling numerous large-scale applications. Human involvement is one of the most important features, and human mobility offers unprecedented opportunities for both sensing coverage and data transmission. In this article, we investigate the opportunistic characteristics of human mobility from the perspectives of both sensing and transmission, and discuss how to exploit these opportunities to collect data efficiently and effectively. We also outline various open issues brought by human involvement in this emerging research area.

INTRODUCTION

With recent advancements in mobile pervasive sensing and transmission technologies, especially the proliferation of smart phones, we are rapidly entering the era of the Internet of Things (IoT), which aims at sensing and interconnecting various physical objects and their surroundings in the realistic world more comprehensively and on a larger scale [1, 2].

If we still use traditional mote-class sensor networks for large-scale and fine-grained sensing, a large number of sensor nodes must be deployed to guarantee the area coverage and communication connectivity, which is economically infeasible or undesirable. Take the CitySee project, for instance: 100 sensor nodes and 1096 relay nodes are deployed for CO_2 monitoring in an urban area of around 1 km^2 [3]. If this system is extended to a larger urban area, for example, within the 5th ring in Beijing (about 900 km^2), we would need to deploy at least 90,000 sensor nodes and around 1,000,000 relay nodes to maintain full area coverage and communication connectivity. Expensive sensor cost together with the deployment and maintenance cost will make it hard to implement.

Fortunately, recent advancements in mobile pervasive sensing and transmission technologies trigger research in leveraging human-carried everyday devices (e.g., smartphones, wearable devices) or vehicle-mounted sensors (e.g., GPS, OBD-II) to monitor large-scale phenomena that cannot easily be measured by a single individual. This sensing paradigm is popularly called mobile crowd sensing (MCS) [4, 5] or *people/human-centric sensing* [6]. Figure 1 illustrates an urban sensing application scenario: a group of mobile users equipped with various sensors, GPS receivers, and wireless communication modules (e.g., Bluetooth, WiFi) move within a monitoring region, opportunistically take samples, and report sensory data to the monitoring center to build a city-scale sensing map of some phenomenon. This novel sensing paradigm has enabled numerous large-scale applications such as urban environment monitoring, traffic monitoring, road surface monitoring, and street parking availability statistics [4].

Human involvement is one of the most important characteristics of MCS. Compared to traditional sensor networks, human mobility offers unprecedented opportunities for both sensing coverage and data transmission. Lane *et al.* [7] discern two classes of sensing paradigms in MCS:

- *Participatory sensing*: It requires the participants to *consciously* opt to meet the application requests by deciding when, where, what, and how to sense.
- *Opportunistic sensing*: It is fully *unconscious*, namely the application may run in the background and opportunistically collect data without active involvement of users (e.g, continuous Wi-Fi signal sensing only needs to keep the Wi-Fi open).

In this article we mainly focus on opportunistic sensing paradigm, which more easily supports large-scale deployments and application diversity [7].

On the other hand, there are two classes of transmission paradigms in MCS:

- *Infrastructure-based transmission*: It considers users reporting and accessing sensory data through the Internet by cellular networks (e.g., 3G/4G mobile networks).
- *Opportunistic transmission*: It enables opportunistic data forwarding among mobile users through intermittent connections with short-range radio communications (e.g., bluetooth, WiFi).

Most existing MCS applications adopt the infrastructure-based transmission paradigm. However, this paradigm cannot be applied in some scenarios where network coverage is poor or network access is expensive. For example, dead spots of network coverage are commonly found in remote areas and even in some parts of major cities. Moreover, the infrastructure is down in disaster recovery scenarios. In this article, we mainly focus on the opportunistic transmission paradigm, which offers another way to collect and share data. It works well without requiring any centralized server or infrastructure for communication and management, and also reduces

Huadong Ma, Dong Zhao, and Peiyan Yuan are with Beijing University of Posts and Telecommunications.

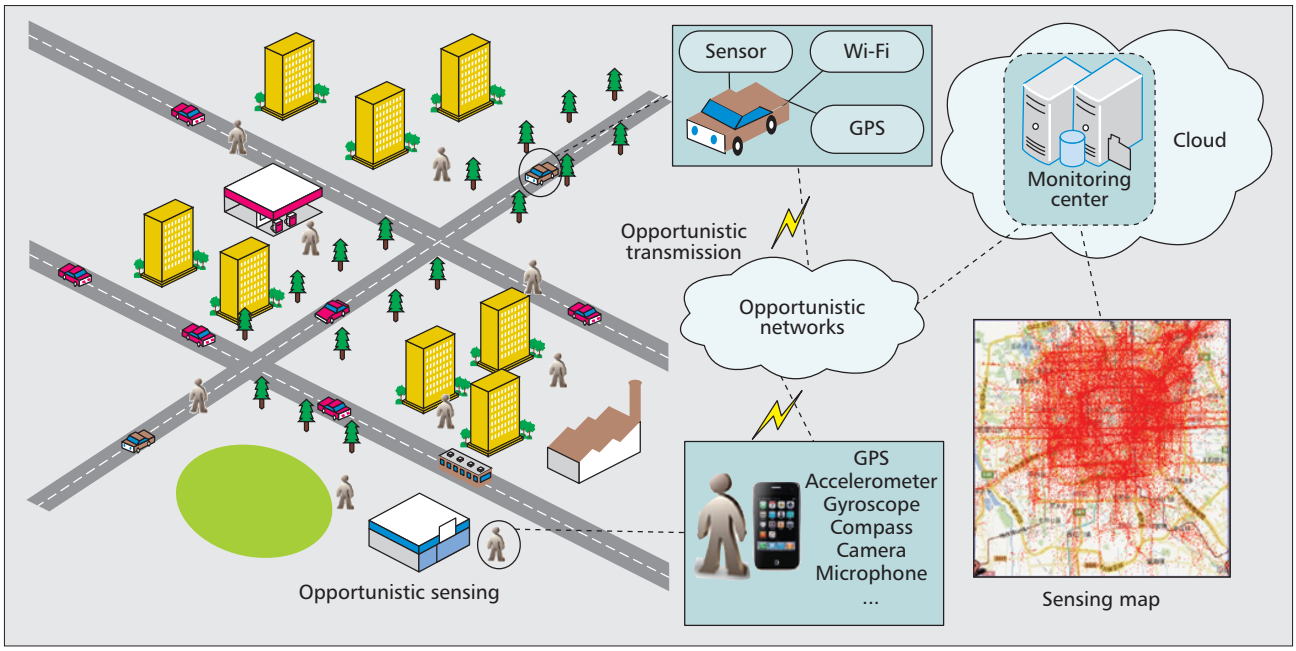


Figure 1. An illustration of opportunistic urban sensing.

the workload of cellular networks in dense areas. Moreover, it is more energy-efficient and less expensive, which is very important because most users hope to save battery energy and data usage on their mobile devices.

In the following, we discuss the opportunities and challenges that human involvement brings to the MCS. We investigate the opportunistic characteristics of human mobility from the perspectives of both sensing and transmission. We discuss how to exploit the opportunities that human mobility offers for sensing and transmitting efficiently and effectively. Finally, we summarize our conclusions and various open issues.

HUMAN INVOLVEMENT: A DOUBLE-EDGED SWORD

In traditional mote-class sensor networks, humans are only the end consumers of the sensory data collected from unattended and autonomous systems. In contrast, one of the most important characteristics of MCS is deeper involvement of humans in the whole loop of the data-to-decision process, including sensing, transmission, analysis of big sensory data, and decision making. This characteristic is a double-edged sword. From a positive perspective, it brings unprecedented opportunities.

- It is easier to deploy the network at lower cost, because millions of mobile devices or vehicles already exist in many cities around the world. Moreover, human mobility can be exploited to improve sensing coverage and data transmission. On one hand, mobile nodes can sense the surroundings wherever their holders arrive opportunistically, which enables building large-scale sensing applications. On the other hand, opportunistic contacts among mobile users can be exploited to deliver sensory data in networks with intermittent connections based on the store-carry-and-forward paradigm [8].
- It is easier to maintain the network, because mobile nodes often have more power supply, stronger computation, and larger storage, and communication capacity. Moreover, mobile nodes are always managed and maintained in good condition by their holders. For example, people could charge their mobile phones as needed every day.
- It is more extensible and flexible, because we only need to recruit more users to adapt to the expansion of the system scale.

From a negative perspective, human involvement also brings many new challenges.

- The number of mobile users, the availability of sensors, and the data quality always change over time due to the randomness of human mobility and the dynamics of human contexts such as the residual battery energy of mobile nodes and people's preferences. All these factors make it more difficult to guarantee reliable sensing quality in terms of coverage, latency, and confidence.
- Human involvement naturally brings privacy concerns. Mobile users may not want to share their sensory data, which may contain or reveal their private and sensitive information (e.g., their current location).
- While participating in MCS, mobile users consume their own resources (e.g., battery and computing power) and have potential privacy threats. Thus, incentive mechanisms are necessary to provide participants with enough rewards for their participation costs. From the sensing perspective, proper incentives must be offered to users for completing specific sensing tasks [9]. From the transmission perspective, users need to be rewarded for forwarding data for each other [10].

OPPORTUNISTIC CHARACTERISTICS OF HUMAN MOBILITY

In traditional sensor networks, nodes are often static, or have random or controlled mobility. In contrast, in MCS human mobility has some unique characteristics such as spatio-temporal correlation, hotspots' effects, and sociality. Identifying these characteristics is beneficial to estimate sensing quality, perform network planning, design efficient sensing and transmission protocols, and develop accurate mobility models.

OPPORTUNITIES IN OPPORTUNISTIC SENSING

At present, little work focuses on the *sensing opportunities* provided by human mobility, which have important impacts on the sensing quality of many MCS applications (e.g., urban environment monitoring applications). Here we consider the application scenario in Fig. 1, assuming that it aims to build a sensing map of some phenomenon (e.g., CO_2 concentration) in a large monitoring region (e.g., within the 5th ring in Beijing) during a time span T (e.g., 6:00–24:00 every day). In fact, there are two basic problems unsolved:

- How can the sensing opportunities and sensing quality be measured?
- How many mobile users can provide enough sensing opportunities to achieve the required sensing quality?

In traditional stationary sensor networks, the coverage is used to measure the sensing quality, which cannot change over time. In contrast, the coverage in MCS is time-variant due to human mobility. Therefore, we propose a new metric called *inter-cover time* to characterize the opportunity with which a subregion is covered [11]. Especially in the time domain, we divide T into multiple sampling periods of T_s, as illustrated in Fig. 2a. In the space domain, we divide the monitoring region into a set of grid cells, as illustrated in Fig. 2b. A grid cell is said to be covered by a mobile user only when a new sampling period arrives and the location of the mobile user is just within the area of the grid cell.The inter-cover time is defined as the time elapsed between two consecutive periods of coverage of the same grid cell. Obviously, shorter inter-cover time results in better sensing quality for a grid cell. In order to explore the pattern of inter-cover times occurring in realistic scenarios, we perform empirical measurement studies on real mobility traces of thousands of taxis collected in Beijing and Shanghai, two of the largest cities in China. According to our analysis results, we find that the distribution of the aggregated inter-cover times follows a *truncated power-law* distribution (it has a power-law tendency at the head part and decays exponentially at the tail) regardless of the size of grid cells and the number of mobile users.

In order to solve the second basic problem, we first use a metric called the *opportunistic coverage ratio* to characterize the relationship between the sensing quality and the number of users. The opportunistic coverage ratio is defined as the expected ratio of grid cells that can be opportunistically covered during a specific time interval. It can be derived as a function of the distribution of the aggregated inter-cover times, which increases monotonically with the number of users and the time interval. Then we formulate this problem as follows: What is the minimum number of mobile users that need to be deployed so that the opportunistic coverage ratio is not less than a threshold during a specific time interval? For example, according to our analysis on the real datasets, we need to deploy at least 5800 and 6300 taxis in Beijing and Shanghai, respectively, so that the opportunistic coverage ratios in a region of 900 km^2 are not less than 90 percent during the time interval of one hour.

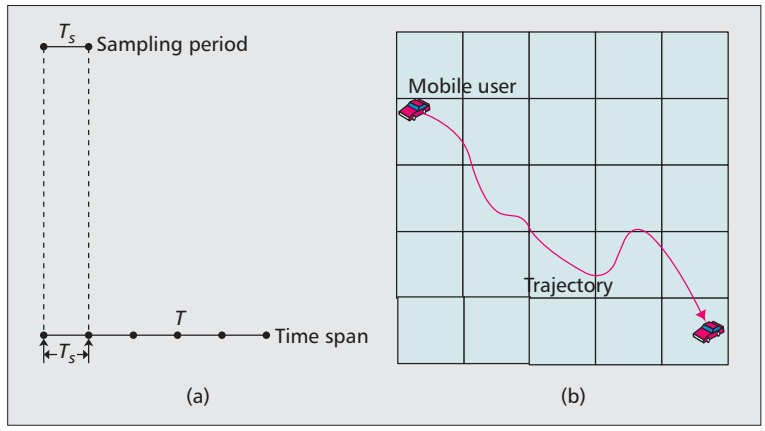


Figure 2. An illustration of discretizing the time-space domain: a) time domain; b) space domain.

OPPORTUNITIES IN OPPORTUNISTIC TRANSMISSION

The transmission opportunities of human mobility and their impacts on the data delivery performance have been intensively studied and relatively well understood in opportunistic networks or delay-tolerant networks (DTNs). The *inter-contact time* is one key metric to characterize the transmission opportunities of the same couple of mobile users. Since the inter-contact time reflects the frequency of opportunities for forwarding messages from one user to another, it directly affects the data delivery performance. Obviously, longer inter-contact time results in longer delivery delay and lower delivery ratio. Now, several empirical results based on human mobility traces have reached a common conclusion that the distribution of the aggregated inter-contact times follows a *truncate power-law* [8].

On the other hand, human sociality has a key impact on human mobility, since it decides the spatial properties of human mobility (i.e., where people move). Thus, human sociality has been considered an important factor affecting the performance of opportunistic forwarding protocols. Recently, several works have mainly focused on social-based forwarding protocols to improve opportunistic transmission performance by leveraging the ample social information encoded in human mobility. The reason behind these protocols is that the underlying social attribute is more stable than the time-variant network topology, and hence can be used for better relay selections. Table 1 classifies these protocols based on the social metrics they exploit.

OPPORTUNISTIC SENSING

Due to human dynamics, it is an important and complex problem to identify the right set of mobile users that can produce the desired data with proper parameters (e.g., sampling rate) to achieve the required sensing quality in energy-efficient ways. Reddy *et al.* considered the participatory sensing paradigm and developed a recruitment framework to enable organizers to identify well-suited participants for data collections based on geographic and temporal availability as well as participation habits [12]. A commonly used method for opportunistic sensing is to make every mobile user sense periodically. However, this method is very inefficient because many redundant data samples may be produced by a large number of mobile users. In order to reduce data redundancy and improve energy efficiency, it is necessary to design a cooperative sensing method to control sensing activities of mobile users such that they produce just enough data samples for the application.

First, we notice that the sensing coverage is spatio-temporal correlative. Let us still consider the discrete time-space model in Fig. 2. We further divide the time span T into multiple coverage periods, where each coverage period contains multiple sampling periods. It is reasonable to assume that a grid cell needs to be covered once or several times within a coverage period, instead of each point in the grid cell being covered at any time. Moreover, frequent samplings make it more likely that a grid cell will be covered at a higher cost. Therefore, it is necessary to design a scheduling mechanism for each mobile user to decide when and where to perform sampling tasks.

Second, since different mobile users always have heterogeneous mobility regions with some randomness, they could make different contributions to the coverage. It is important to design a user selection mechanism to eliminate user redundancy and hence reduce data redundancy.

Based on the above analysis, we design a cooperative opportunistic sensing framework [13], as illustrated in Fig. 3. First, the time-space domain of the monitoring region is discretized according to the application requirements. Then we obtain a set of effectively covered grid cells (the number of times these grid cells can be covered within each coverage period are not less than a specified threshold) and coverage contribution matrices (representing the times that one mobile user covers different grid cells within different coverage periods) for each mobile user, according to the history trajectories of all mobile users. We design two mechanisms to reduce user redundancy and data redundancy:

- The offline *user selection* mechanism. It can select the minimum number of users to achieve the coverage requirements for those effectively covered grid cells based on the coverage contribution matrices of mobile users.
- The online *adaptive sampling* mechanism. It can control the sampling rates of the selected users adaptively. In particular, it sets two tables, a control table and a coverage table, locally stored at each user. The control table is used to decide whether each user needs to participate in sampling tasks within each coverage period according to the results of a user selection mechanism. The coverage table records the number of times that each grid cell has been covered during some coverage period, and then each user decides whether to take samplings according to the current coverage table.

Social metrics	Meanings	Typical protocols
Centrality	The social position of a user	SimBet, PeopleRank, BUBBLE
Similarity	The social distance between two users	SimBet, MobiSpace
Social relationship	Acquaintance, friend, or stranger	SMART
Social structure	Virtual community or physical hotspots	BUBBLE, Hotent

Table 1. A summary of existing social-based opportunistic forwarding protocols.

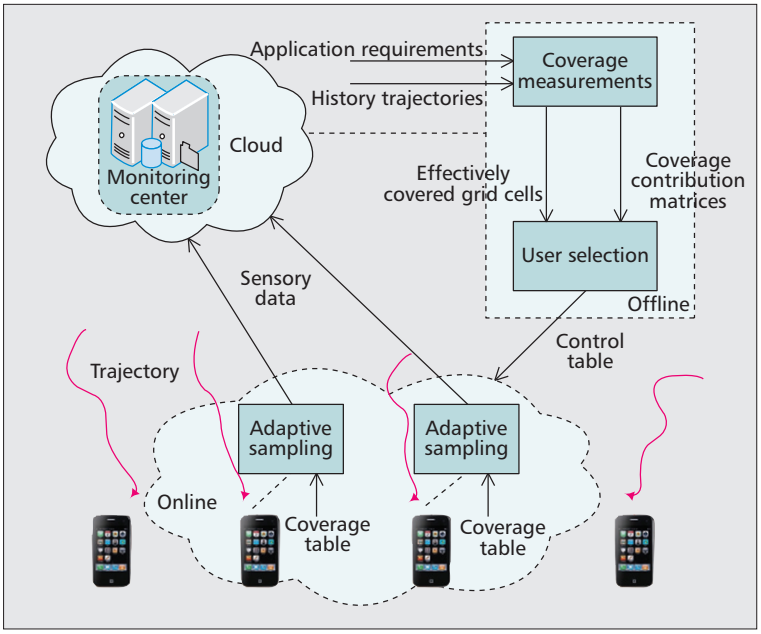


Figure 3. An overview of the cooperative opportunistic sensing framework.

OPPORTUNISTIC TRANSMISSION

In the past few years, many opportunistic forwarding protocols have been proposed in opportunistic networks. Among these protocols, epidemic routing was the first and most generic one without knowing anything about the mobility of users. It tried to grasp each forwarding opportunity, thus resulting in the minimum delivery delay under ideal conditions (unlimited resources such as bandwidth and buffer space) at high cost. In order to reduce the cost, many variants of epidemic routing have been proposed (k-hop schemes, probabilistic forwarding, spray-and-wait, etc.). Some protocols attempted to identify the best relay node to achieve better trade-off between delivery delay and transmis-

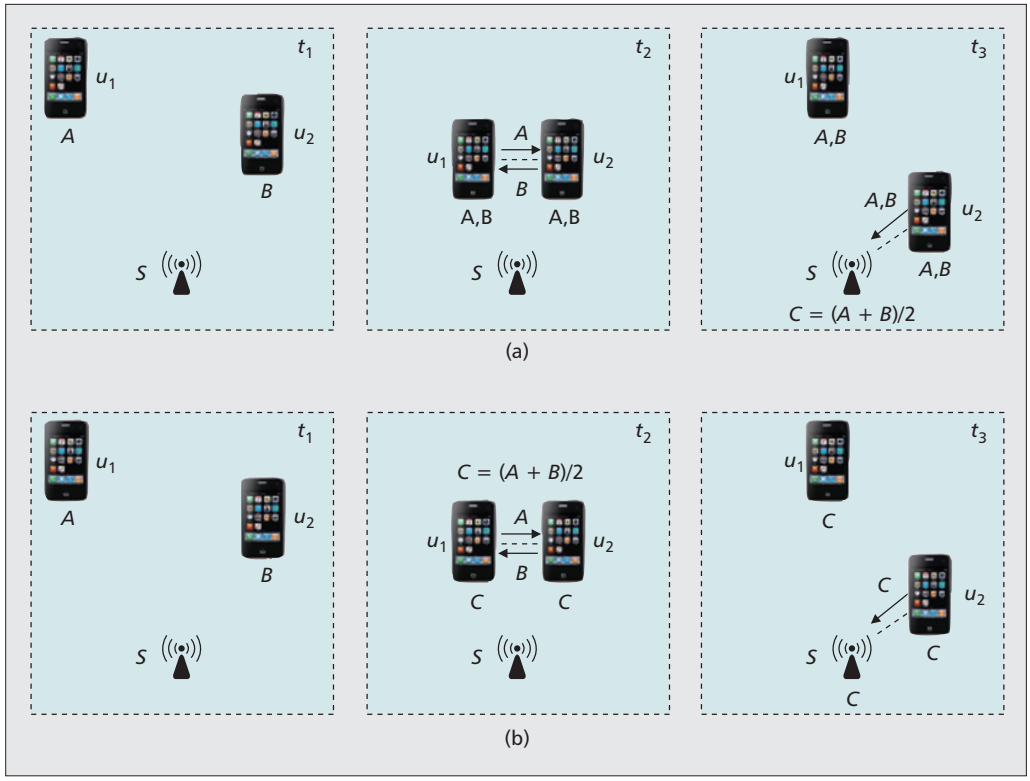


Figure 4. An illustration of data forwarding with or without fusion: a) forwarding sensory data without fusion; b) forwarding sensory data with fusion.

Due to human dynamics, it is an important and complex problem to identify the right set of mobile users that can produce the desired data with proper parameters (e.g., sampling rate) to achieve the required sensing quality in energy-efficient ways.

sion overhead by exploiting the context information (e.g., the mobility of a node and its current energy level). More recently, several works have exploited human sociality for improving opportunistic transmission performance.

However, most existing opportunistic forwarding protocols have focused on the sharing or dissemination of data interesting to individual mobile users instead of the sensory data collection in MCS, and thus failed to consider the spatial-temporal correlation among sensory data and its impact on the network performance. It would be beneficial to combine opportunistic forwarding and in-network processing based on a store-carry-process-and-forward paradigm. On the other hand, most existing social-based opportunistic forwarding protocols use traditional approaches in social networks or ego networks to evaluate social metrics. We argue that these approaches have high spatio-temporal complexity due to transient user contacts and intermittently connected environment, and hence cannot be applied in large-scale opportunistic scenarios. It is necessary to develop a lightweight approach to exploiting human sociality for improving opportunistic transmission performance.

The Impact of Data Fusion

Considering the spatial-temporal correlation among sensory data, it is beneficial to integrate opportunistic forwarding protocols with *data fusion* (or *data aggregation*) for two reasons:

- Users may be interested only in the aggregated results of sensory data (e.g., the average temperature or noise level).
- Sensory data collected in close proximity or time periods may be highly correlated, and data fusion can effectively eliminate redundancy and hence reduce network overhead.

Although many routing protocols supporting data fusion have been proposed in traditional sensor networks, to the best of our knowledge, few works have investigated opportunistic forwarding protocols supporting data fusion in MCS.

We use Fig. 4 to illustrate the importance of coupling between opportunistic forwarding and data fusion. Assume that two users u_1 and u_2 carry two correlated packets (e.g., two samples produced in the same grid cell within the same time period) at time t_1, and opportunistically forward the packets to a sink node S. Finally, S obtains the average of two samples. Figure 4a illustrates the forwarding process without data fusion: when u_1 meets u_2 at time t_2, they forward data to each other; then at time t_3, u_2 meets the sink node S, and delivers both two packets A and B to S; S takes the average of two samples after it has received both of the original packets. Thus, the transmission overhead (i.e., the total number of transmissions) is 4. Next, let us see the forwarding process with data fusion as illustrated in Fig. 4b: when u_1 meets u_2, they forward data to each other, and store the fused result, a new packet $C = (A + B)/2$, instead of two original packets; then at time

Since sensing opportunities are imbalance among different regions, it is important to study where to deploy how many specialized sensor nodes, and how to collaborate with mobile nodes for achieving required sensing quality.

t_3, u_2 meets S, and delivers packet C to S. Thus, the transmission overhead is 3, lower than that of the forwarding process without data fusion.

Although the idea of integrating opportunistic forwarding with data fusion seems simple and straightforward, we still face some new challenges for both performance modeling and protocol design in practice. Previous work on performance modeling of opportunistic forwarding protocols assumed that all packets were propagated individually. However, the packets are spatial-temporally correlated in the forwarding process with data fusion, which causes a more complex propagation process. In our work [14], an ordinary differential equation was derived for modeling the dissemination law of correlated packets, which could serve as fundamental guidelines on integrating opportunistic forwarding with data fusion for achieving tradeoff among various performance metrics.

We designed two novel protocols by leveraging data fusion: Epidemic Routing with Fusion (ERF) and Binary Spray-and-Wait with Fusion (BSWF), and proved that both protocols outperformed those without data fusion (i.e., epidemic routing and spray-and-wait).

EVALUATION AND EXPLOITATION OF HUMAN SOCIALITY

By analyzing GPS traces of pedestrians from the real world, we find three phenomena:

- People always move around a set of popular locations, called *public hotspots*, instead of purely random movements.
- Each individual shows preference for some particular locations, called *personal hotspots*.
- Both types of hotspots have two key features enabling a lightweight opportunistic forwarding protocol: *burstiness*, implying that there are only a small number of hotspots required to exchange among users, and *stability*, implying that only infrequent updating of hotspots is required.

Motivated by the above observations, we exploit hotspots to design a new routing metric, called Hotent (for HOTspot ENTropy) [15]. To reduce the high spatio-temporal complexity of traditional social network analysis technologies, Hotent converts the problem of evaluating social metrics to a similarity matching problem according to the following three steps:

- Using the *inverse symmetric entropy* of personal hotspots of two users to evaluate the *similarity* between them
- Using the *relative entropy* between public hotspots and personal hotspots to evaluate the *centrality* of users
- Using the *law of universal gravitation* to integrate centrality and similarity into the Hotent metric, based on the metaphor of mass for user centrality and distance for the similarity between two users

We have verified that Hotent largely outperformed other state-of-the-art work, especially in terms of packet delivery ratio and average number of hops per packet.

CONCLUSIONS AND OPEN RESEARCH ISSUES

This article discusses the opportunities and challenges in mobile crowd sensing brought on by human involvement. In particular, we have investigated the opportunistic characteristics of human mobility. From the sensing perspective, we use a new metric called *inter-cover-time* to characterize the sensing opportunities, and use another metric called *opportunistic coverage ratio* to evaluate the sensing quality of MCS applications. From the transmission perspective, we review some main results: the distribution of inter-contact-times and human sociality have important impacts on opportunistic transmission performance. We also present some approaches to exploiting the opportunities that human mobility offers for sensing and transmission efficiency and effectiveness.

There are many open issues in this emerging research area, including the following.

Evaluation of sensing quality: This is a complex problem affected by many factors. First, the space distribution of human mobility has important impacts. We can consider an MCS system as an "urban camera," and a large number of mobile devices form the charge-coupled device (CCD) sensor of this camera. An urban camera is able to record urban phenomena in the form of sensing images by measurements from mobile devices. However, different from the commonly used definition of resolution for digital images and cameras, the resolution of an urban camera is not simply the pixel count of a mobile phone camera. This is because the pixels of a digital camera form a fine grid, but the pixels of an urban camera have scattered and dynamic distribution. Thus, it is necessary to redefine the resolution of MCS systems, and investigate the relationship between the resolution and the number of mobile users. Second, mobile users have heterogeneous data quality. It is important to evaluate users' data quality, and get rid of malicious and low-quality data.

Integration of MCS and static sensing: Although traditional sensor networks have higher cost and poorer scalability, they often have more reliable sensing quality, which can be used to compensate for inadequate sensing opportunities provided solely by an MCS system. Since sensing opportunities are imbalanced among different regions, it is important to study where to deploy how many specialized sensor nodes, and how to collaborate with mobile nodes to achieve the required sensing quality.

Integration of opportunistic forwarding and in-network processing: As a starting point, we investigated the integration of some simple data fusion functions (e.g., averaging, summation, voting, and max/min) and two basic opportunistic forwarding protocols. In the future, we need to further explore whether more complex in-network processing approaches can be combined with other opportunistic forwarding protocols (e.g., context-aware and social-based forwarding) and how much performance can be improved.

Adaptive opportunistic forwarding protocols: Although extensive opportunistic forwarding

protocols have been proposed, almost all of them only applied to some specific scenarios rather than every scenario. Since transmission opportunities are imbalanced among different physical regions and virtual communities, it is still an important and challenging problem to study when and where to use which protocols (strong-connectivity-oriented or weak-connectivity-oriented), and how to switch them adaptively.

Context-aware incentive mechanisms: Since the preferences of mobile users always change dynamically with their contexts, we should offer personalized incentives to users for optimizing the system utility by identifying users' contexts, mobility, and social properties.

Balance among sensing quality, incentive, and privacy: We need to consider various factors synthetically and systematically.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China under Grant No.61332005 and No.61133015, the Funds for Creative Research Groups of China under Grant No.61121001, the Specialized Research Fund for the Doctoral Program of Higher Education under Grant No.20120005130002.

BIOGRAPHIES

HUADONG MA [M] (mhd@bupt.edu.cn) is a Chang Jiang Scholar professor and director of the Beijing Key Lab of Intelligent Telecommunications Software and Multimedia, executive dean of School of Computer Science, Beijing University of Posts and Telecommunications, China. He received his Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Science in 1995. From 1999 to 2000, he held a visiting position in the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor. He was a visiting professor at the University of Texas at Arlington from July to September 2004, and a visiting professor at Hong Kong University of Science and Technology from December 2006 to February 2007. His current research focuses on multimedia system and networking, sensor networks, and the Internet of Things, and he has published over 180 papers and four books in these fields. He is a member of ACM.

DONG ZHAO received his Ph.D. degree from the School of Computer Science, Beijing University of Posts and Telecommunications in 2014. He was a visiting Ph.D. student in the Department of Computer Science at Illinois Institute of Technology during 2012–2013. His research interests include sensor networks, opportunistic networks, mobile crowdsourcing, and the Internet of Things.

PEIYAN YUAN received his Ph.D. degree at the School of Computer Science, Beijing University of Posts and Telecommunications in 2014. His research interests include networking and protocol engineering, mobile opportunistic networks, social networks, and others. He is a member of ACM.

REFERENCES

[1] H.-D. Ma, "Internet of Things: Objectives and Scientific Challenges," *J. Computer Science and Tech.*, vol. 26, no. 6, 2011, pp. 919–24.
[2] B. Guo *et al.*, "Opportunistic IoT: Exploring the Harmonious Interaction between Human and the Internet of Things," *J. Network and Computer Applications*, vol. 36, no. 6, 2013, pp. 1531–39.
[3] X. Mao *et al.*, "Citysee: Urban CO_2 Monitoring with Sensors," *Proc. IEEE INFOCOM*, 2012, pp. 1611–19.
[4] R. K. Ganti, F. Ye, and H. Lei, "Mobile Crowdsensing: Current State and Future Challenges," *IEEE Commun. Mag.*, vol. 49, no. 11, 2011, pp. 32–39.
[5] B. Guo *et al.*, "From Participatory Sensing to Mobile Crowd Sensing," IEEE PerCom Workshops (SCI), 2014.
[6] A. Campbell *et al.*, "The Rise of People-Centric Sensing," *IEEE Internet Comp.*, vol. 12, no. 4, 2008, pp. 12–21.
[7] N. Lane *et al.*, "Urban Sensing Systems: Opportunistic or Participatory?," *Proc. HotMobile*, 2008, pp. 11–16.
[8] M. Conti *et al.*, "From Opportunistic Networks to Opportunistic Computing," *IEEE Commun. Mag.*, vol. 48, no. 9, 2010, pp. 126–39.
[9] D. Zhao, X.-Y. Li, and H.-D. Ma, "How to Crowdsource Tasks Truthfully Without Sacrificing Utility: Online Incentive Mechanisms with Budget Constraint," *Proc. IEEE INFOCOM*, 2014, pp. 1213–21.
[10] H. Zhou *et al.*, "ConSub: Incentive-based Content Subscribing in Selfish Opportunistic Mobile Networks," *IEEE JSAC*, vol. 31, no. 9, 2013, pp. 669–79.
[11] D. Zhao *et al.*, "On opportunistic Coverage for Urban Sensing," *Proc. IEEE MASS*, 2013, pp. 231–39.
[12] S. Reddy, D. Estrin, and M. Srivastava, "Recruitment Framework for Participatory Sensing Data Collections," *Proc. Pervasive*, 2010, pp. 138–55.
[13] D. Zhao, H.-D. Ma, and L. Liu, "Energy-Efficient Opportunistic Coverage for People-Centric Urban Sensing," *Wireless Networks*, Jan. 2014, published online.
[14] D. Zhao *et al.*, "COUPON: A Cooperative Framework for Building Sensing Maps in Mobile Opportunistic Networks," *IEEE Trans. Parallel and Distrib. Sys.*, Feb. 2014, published online.
[15] P. Yuan and H.-D. Ma, "Opportunistic Forwarding with Hotspot Entropy," *Proc. IEEE WoWMoM*, 2013.

Since the preferences of mobile users always change dynamically with their contexts, we should offer personalized incentives to users for optimizing the system utility by identifying users' contexts, mobility, and social properties.

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.

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


ITU TELECOM WORLD '15
Budapest 12-15 October
BETTER SOONER
We'll help your ideas to go further and faster
ITU Telecom World 2015 is the global platform to accelerate ICT innovations for social and economicdevelopment. It's where policy makers and regulators meet industry experts, investors, SMEs, entrepreneurs and innovators to exhibit solutions, share knowledge and speed change. Our aim is to help ideas go further, faster to make the world better, sooner.
Visit telecomworld.itu.int to find out more.
#ituworld
telecomworld.itu.int
150 1865 2015

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*