Infocommunications Journal

A PUBLICATION OF THE SCIENTIFIC ASSOCIATION FOR INFOCOMMUNICATIONS (HTE)

December 2012 **Volume IV** **Number 4** **ISSN 2061-2079**

Technically Co-Sponsored by

IEEE COMMUNICATIONS SOCIETY

hte

IEEE HUNGARY SECTION

Indexing information

Infocommunications Journal is covered by INSPEC and Compendex.

The journal is supported by NCA NEMZETI CIVIL ALAPPROGRAM the National Civil Fund.

Infocommunications Journal

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

Editorial Office (Subscription and Advertisements)**:**
Scientific Association for Infocommunications
H-1055 Budapest, Kossuth Lajos tér 6-8, Room: 422
Mail Address: 1372 Budapest Pf. 451. Hungary
Phone: +36 1 353 1027, Fax: +36 1 353 0451
E-mail: info@hte.hu
Web: www.hte.hu

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

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

Publisher: PÉTER NAGY • Manager: ANDRÁS DANKÓ

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

www.infocommunications.hu

Special Issue on Quantum Communications – Guest Editorial

Laszlo Bacsardi, Sandor Imre and Ozgur B. Akan

THIS is the first time when Infocommunications Journal has a special issue on quantum communications. This year's Nobel Prize in Physics winners, David J. Wineland and Serge Haroche had a great contribution in the way to a working quantum computer. Although these computers are going to be the applications of the far future, there are already a few algorithms to solve problems which are very difficult to handle with traditional computers. Quantum computing is based on various quantum effects in physics and offers revolutionary solutions for different problems e.g., prime factorization, searching in unsorted database, key distribution and information coding. The power of quantum parallelism allows us to solve classically complex problems, and the quantum entanglement leads to quantum communication algorithms like teleportation and superdense coding. The quantum cryptography provides new ways to transmit information with unconditional security by using different quantum key distribution protocols

In this Special Issue on Quantum Communications of the Infocommunications Journal, three selected papers highlight the different directions and problems of the quantum communications.

Deep-space optical communication is a key component of the NASA roadmap, with the goal of returning greater data-volumes from Mars and other solar-system encounters in future missions. Conventional optical receivers currently under consideration for deep-space communications employ photon-counting or coherent detection to potentially extract useful information even from a single photon, on the average. However, while quantum mechanics promises greater gains, it fails to specify how these theoretical gains can be achieved in practice. *Quantum Receiver for Binary Coherent-State Signals with Constant-Intensity Local Lasers* by Victor A. Vilnrotter describes the quantum receiver for this type of communication. According to their results, the new receiver concept can be implemented using practical measurements amenable to high data-rate operation, hence it may enable future deep-space optical communications with performance approaching the greatest possible fidelity allowed by the laws of quantum mechanics.

There is a growing interest in providing and improving radio coverage for mobile phones, short range radios and WLANs inside buildings. The recently published methods use any heuristic techniques for finding the optimal Access Point (AP) positions. The *Classical and Quantum Genetic Optimization Applied to Coverage Optimization for Indoor Access Point Networks* by Lajos Nagy introduces the Quantum inspired Genetic Algorithm (QGA) for indoor access point position optimization to maximal coverage and compares with the Classical Genetic Algorithm (CGA).

The Problem of Testing a Quantum Gate by Subhash Kak deals with a problem that has no analogy in the classical world. To test a quantum gate we need certified quantum gates to generate all possible inputs and since such gates are not available at this time how are we going to certify a gate that has been submitted for certification? In the paper, the authors consider the question of testing of quantum gates as a part of the larger problem of communication through circuits that use a variety of such gates.

LÁSZLÓ BACSÁRDI obtained M.Sc. degree in computer engineering at Budapest University of Technology and Economics (BME) in 2006. He holds an associate professor position at the University of West Hungary, where he is the Head of the Institute of Informatics and Economics. He wrote his PhD thesis on the possible connection between space communications and quantum communications at the BME Department of Telecommunications in 2012. His current research interests are in mobile ad hoc communication, quantum computing and quantum communications. He is the Secretary General of the Hungarian Astronautical Society (MANT), which is the oldest Hungarian non-profit space association founded in 1956. He is member of the board of a Hungarian scientific newspaper ('World of Nature') and he is the publisher of a non-profit Hungarian space news portal ('Space World'). Furthermore he is member of IEEE, AIAA and the HTE. He has joined the Space Generation Advisory Council (SGAC) as well, currently active as the Hungarian National Point of Contact.

SÁNDOR IMRE was born in Budapest in 1969. He received the M.Sc. degree in Electrical Engineering from the Budapest University of Technology (BME) in 1993. Next he started his Ph. D. studies at BME and obtained dr. univ. degree in 1996, Ph.D. degree in 1999 and DSc degree from the Hungarian Academy of Sciences in 2007. Currently he is carrying his activities as Professor and Head of Dept. of Telecommunications. He is chairman of Telecommunication Scientific Committee of the Hungarian Academy of Sciences. He participates the Editorial Board of two journals: Infocommunications Journal and Hungarian Telecommunications. He was invited to join the Mobile Innovation Centre as R&D director in 2005. His research interest includes mobile and wireless systems, quantum computing and communications. Especially he has contributions on different wireless access technologies, mobility protocols, security and privacy, reconfigurable systems, quantum computing based algorithms and protocols.

OZGUR B. AKAN (M'00-SM'07) received the B.S. and M.S. degrees in electrical and electronics engineering from Bilkent University and Middle East Technical University, Ankara, Turkey, in 1999 and 2001, respectively, and the Ph.D. degree in electrical and computer engineering from the Broadband and Wireless Networking Laboratory, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, in 2004. He is currently a Professor with the Department of Electrical and Electronics Engineering, and the Director of Next-generation and Wireless Communications Laboratory (NWCL), Koc University, Istanbul, Turkey. His current research interests include wireless communications, acoustic communications, nano communications, quantum communications and information theory. Dr. Akan is an Associate Editor for the IEEE Transactions on Vehicular Technology, the International Journal of Communication Systems (Wiley), the European Transactions on Telecommunications, and the Nano Communication Networks Journal (Elsevier). He served as an Editor for ACM/Springer Wireless Networks (WINET) Journal from 2004 to 2010, as an Area Editor for AD HOC Networks Journal (Elsevier) from 2004 to 2008, as a Guest Editor for several special issues. He currently serves as the General Co-Chair for ACM MobiCom 2012, General Co-Chair for IEEE MoNaCom 2012, and TPC Co-Chair for IEEE ISCC 2012. He is the Vice President of the IEEE Communications Society - Turkey Section. He is a Senior Member of the IEEE Communications Society (COMSOC), and a member of ACM. He is a COMSOC Distinguished Lecturer (2011-2012). He received the IEEE COMSOC Outstanding Young Researcher Award for EMEA Region 2010 (as runner-up), the IBM Faculty Award twice in 2010 and 2008, and the Turkish Academy of Sciences Distinguished Young Scientist Award 2008 (TUBA-GEBIP).

Quantum Receiver for Binary Coherent-State Signals with Constant-Intensity Local Lasers

Victor A. Vilnrotter, *Senior Member, IEEE*

Abstract **– A quantum receiver capable of approaching the fundamental quantum limit on bit error probability is described and evaluated. Conventional optical and abstract quantum mechanical descriptions are provided and the underlying principles derived in both domains, thus providing a bridge to optimum quantum measurements in terms of well-understood optical communications concepts. Receiver performance is evaluated for the case of binary phase-shift keyed modulation, and it is shown that significant gains can be achieved over near-optimum receivers reported previously in the literature. This new receiver concept can be implemented using practical measurements amenable to high data-rate operation, hence it may enable future deep-space optical communications with performance approaching the greatest possible fidelity allowed by the laws of quantum mechanics.**

Index Terms **– Quantum detection, binary coherent-state signals, Helstrom bound.** [1]

I. Introduction

Deep-space optical communication is a key component of the NASA roadmap, with the goal of returning greater data-volumes from Mars and other solar-system encounters in future missions. Conventional optical receivers currently under consideration for deep-space communications employ photon-counting or coherent detection to potentially extract useful information even from a single photon, on the average. However, quantum mechanics promises greater gains, but fails to specify how these theoretical gains can be achieved in practice. When pure states are used to communicate information, such as those obtained from pulsed or phase-modulated lasers, the minimum achievable error probability subject to the laws of quantum mechanics has been determined by C. Helstrom [1], and hence referred to as the "Helstrom bound". So far, only a few schemes have been devised that are capable of achieving the Helstrom bound for a general class of binary signals, including: the Dolinar receiver [1] and the Sasaki-Hirota receiver [2]. The Dolinar receiver was the first structured approach that achieved the Helstrom bound using physically realizable measurements together with real-time optical feedback, however practical implementation at high data-rates was found to be challenging due to the requirement for precise local laser intensity control [3, 4]. A different approach was proposed by Sasaki and Hirota [2], which does not employ optical feedback but achieves the Helstrom bound via unitary transformations and photon counting. However, a practical implementation of the Sasaki-Hirota receiver requires multiphoton nonlinear optical processing, which also leads to complex receiver structures. The receiver structure proposed here overcomes these practical impediments by approaching the Helstrom bound using well-known practical measurements that enable high-speed implementation, while attaining significantly better performance than photon-counting, coherent detection or even near-optimum quantum receivers such as the "Kennedy receiver" which is exponentially optimum and implementable at high data-rates [1, 3, 4].

Manuscript submitted August 11th, 2012, revised November 19th 2012.
Author is with the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, 91109, USA (Victor.A.Vilnrotter@jpl.nasa.gov).

II. Quantum Description of Communications Signals

At any instant of time, the state of a quantum system is completely specified by a state vector $|\psi\rangle$ in a Hilbert space over the field of complex numbers. The state vector, or "ket" $|\psi\rangle$, can be thought of as a column vector of infinite dimensions. An equivalent "row vector" representation of the state vector is denoted by $\langle\psi|$ in Dirac notation.

If $|\psi_1\rangle$ and $|\psi_2\rangle$ are states of a quantum system, then so is their linear combination $|\psi\rangle = a_1|\psi_1\rangle + a_2|\psi_2\rangle$ where a_1 and a_2 are complex numbers. The row-vector representation is $\langle\psi| = a_1^*\langle\psi_1| + a_2^*\langle\psi_2|$. The "overlap" between two states $|\psi\rangle$ and $|\varphi\rangle$ is the complex number $\langle\psi|\varphi\rangle$ or its complex conjugate $\langle\varphi|\psi\rangle$. If the overlap is zero, the states are orthogonal. The state is normalized if $\langle\psi|\psi\rangle = 1$. Thus, for orthonormal states $\langle\psi_m|\psi_n\rangle = \delta_{mn}$, where δ_{mn} is the Kroenecker delta. If $|\psi_1\rangle$ and $|\psi_2\rangle$ are orthonormal and $|\psi\rangle$ is normalized, then their overlap is

$$\begin{aligned}\langle\psi_1|\psi\rangle &= \langle\psi|\psi_1\rangle^* = a_1 \\ \langle\psi_2|\psi\rangle &= \langle\psi|\psi_2\rangle^* = a_2\end{aligned} \tag{1}$$

where $|a_1|^2 + |a_2|^2 = 1$, and we interpret $|a_1|^2$ and $|a_2|^2$ as the probabilities that the system is found to be in states $|\psi_1\rangle$ and $|\psi_2\rangle$, respectively. Generalization to superposition of an arbitrary number of states yields

$$|\psi\rangle = \sum_n a_n|\psi_n\rangle \tag{2}$$

$$\sum_n |a_n|^2 = 1 \tag{3}$$

with the interpretation that $|a_n|^2$ is the probability that the system is found to be in state $|\psi_n\rangle$.

In the classical model of optical communications, information can be incorporated in a laser beam by modulating

the amplitude, phase or polarization of the optical field. In the quantum model, information can be similarly incorporated into coherent states represented by the ket $|\alpha\rangle$ and described in great detail in [6]. Coherent states can be expressed as a superposition of orthonormal number states $|n\rangle$ as

$$|\alpha\rangle = e^{-\frac{1}{2}|\alpha|^2} \sum_{n=0}^{\infty} \frac{\alpha^n}{(n!)^{1/2}} |n\rangle \tag{4}$$

Coherent states are not orthogonal, as can be seen by considering the overlap of two arbitrary coherent states, $|\alpha\rangle$ and $|\beta\rangle$. Orthogonality requires that their overlap vanish, however for distinct coherent states the squared magnitude of their overlap is

$$|\langle\alpha|\beta\rangle|^2 = \left| e^{-(|\alpha|^2+|\beta|^2)/2} \sum_n \sum_m \frac{\alpha^n}{\sqrt{n!}} \frac{(\beta^*)^m}{\sqrt{m!}} \langle n|m\rangle \right|^2 \tag{5}$$

$$= \left| e^{(|\alpha|^2+|\beta|^2)/2} \sum_n \frac{(\alpha\beta^*)^n}{n!} \right|^2 = e^{|\alpha\ \beta|^2}$$

by virtue of the orthogonality of the number states $|n\rangle$. Equation (5) demonstrates that there is always some overlap between coherent states, regardless of how great the average photon count in each state may be [6, 7].

In the state-space interpretation of photon counting developed in [7], two signal states define a plane in Hilbert space. Application of the photon counting projection operators to the signal states generate "measurement states" [1] that span the two-dimensional subspace defined by the signal states, designated as $|w_0\rangle$ and $|w_1\rangle$ in Fig. 1. The squared magnitude of the projection of each signal state onto its associated measurement state is the probability that the signal state will be detected correctly. With this approach, the measurement state for the null hypothesis H_0, $|w_0\rangle$, is taken to be the ground state, corresponding to one of the two binary signals shown in Fig. 1a.

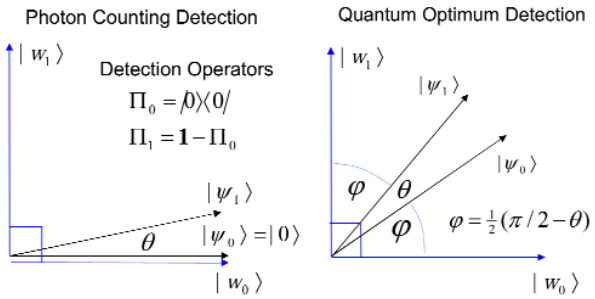


Fig. 1. Measurement state interpretation of binary coherent state detection: a) photon counting; b) optimum quantum measurement achieving the Helstrom bound.

The measurement state for the alternate hypothesis H_1, $|w_1\rangle$, does not in general align with any of the number states, but rather it is a superposition of number states except for the ground state, with coefficients determined by the signal state $|\psi_1\rangle$. A detailed description of this formulation is provided in [7]. The error probability is minimized and the Helstrom bound achieved when the two orthonormal measurement states are rotated symmetrically within the signal subspace, as shown in Fig. 1b. The resulting limit on the error probability has been derived in [7] by evaluating the signal-state projections onto each measurement state, and shown to be exactly equal the Helstrom bound:

$$P(E) = \tfrac{1}{2}\left(1 - \sqrt{1 - 4p_0 p_1 |\langle\alpha|\beta\rangle|^2}\right) \tag{6}$$

The measurement-state approach therefore provides a geometrical interpretation of the optimum quantum measurement, which allows us to relate the abstract quantum optimum measurement to classical measurements that can be carried out in the laboratory.

III. NEAR-OPTIMUM DETECTION OF BPSK SIGNALS

Binary phase-shift keying (BPSK) modulation is particularly well suited to illustrating the key concepts in classical, near-optimum and optimum quantum detection strategies, as well as establishing a correspondence between classical and quantum receiver performance. An example of BPSK signaling is shown in Fig. 2: during each T-second symbol interval, the amplitude of the electric field is taken to be E if the binary data is "1", corresponding to hypothesis H_1, and $-E$ if the binary data is "0", corresponding to H_0. The signal amplitude therefore toggles between $\pm E$ in response to the data, but remains constant during each T-second symbol-interval. Assume that H_0 and H_1 occur with a priori probabilities p_0, p_1 respectively, where $p_0 + p_1 = 1$. The average photon-count within each received symbol interval is $K_\alpha = E^2T = |\alpha|^2$, while the actual photon-count is k. The quantum representation of the binary signals is $|-\alpha\rangle$ when H_0 is true, and $|\alpha\rangle$ when the alternate hypothesis, H_1, occurs.

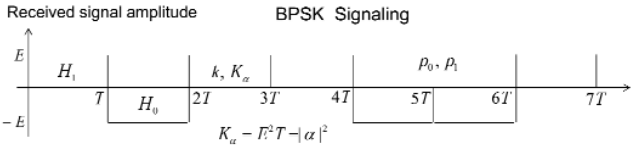


Fig. 2. Classical representation of binary phase-shift keyed (BPSK) data-stream.

For equal a priori probabilities, $p_0 = p_1$, the error probability for coherent detection of BPSK signals is given by the well-known expression $P(E) = \frac{1}{2}[1 - \mathrm{erf}(\sqrt{2K_\alpha})]$, where "erf" is the error function defined as $erf(x) = 2\pi^{-1/2}\int_0^x \exp(-t^2)dt$.

The Kennedy receiver

The displacement operator $D(\gamma)$ shifts any coherent state $|\alpha\rangle$ to a new coherent state $|\alpha+\gamma\rangle$, $D(\gamma)|\alpha\rangle = |\alpha+\gamma\rangle$. A near-optimum detection strategy for binary signals has been devised by R. S. Kennedy in 1974 [1]. The key idea of the Kennedy receiver is to apply the displacement operator $D(\alpha)$ to the coherent states $|-\alpha\rangle, |\alpha\rangle$ before photon-counting

detection, yielding $D(\alpha)|-\alpha\rangle=|0\rangle$ and $D(\alpha)|\alpha\rangle=|2\alpha\rangle$ for the two hypotheses, hence converting the phase-modulated BPSK signals in the classical representation to on-off-keyed signals, but with twice the amplitude and thus four times the pulse energy, since $|2\alpha|^2=4K_\alpha$.

The displaced states are detected using photon counting, yielding an average error probability $P(E)=\frac{1}{2}e^{-4K_\alpha}$ as shown below, corresponding to on-off-keyed signals with average pulse energy $4K_\alpha$. In terms of classical implementation, a constant phase-locked local laser field with amplitude E matched to the received field is first added to the signal using a beam-splitter, followed by conventional photon counting detection. The negative BPSK symbols with amplitude -E (corresponding to the null hypothesis H_0) are therefore converted to zero, whereas the positive symbols with amplitude E are converted to $2E$, as shown in Fig. 3. The detection strategy for the Kennedy receiver is: declare "H_1" if $k>0$, "H_0" if $k=0$.

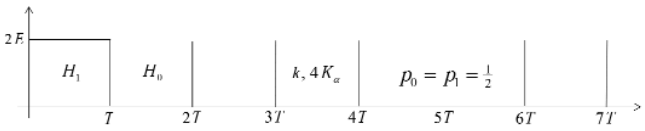


Fig. 3. BPSK signals converted to on-off keyed signals via the Kennedy detection strategy.

The relevant conditional probabilities are given by

$$H_0:\quad p(k=0)=1,\qquad H_1:\quad \begin{cases} p(k=0)=\exp(-4K_\alpha) \\ p(k>0)=1-\exp(-4K_\alpha) \end{cases}$$

$$P(E)=1-P(C)$$

The conditional probabilities of correct detection become $P(C|H_0)=1;\ P(C|H_1)=1-\exp(-4K_\alpha)$ which must be averaged over the a priories to obtain the average probability of correct detection: $P(C)=p_0P(C|H_0)+p_1P(C|H_1)$. With equal a priori probabilities the probability of correct detection is $P(C)\underset{p_0=p_1}{=}\frac{1}{2}\left(1+[1-\exp(-4K_\alpha)]\right)=1-\frac{1}{2}\exp(-4K_\alpha)$. The average error probability is related to the probability of correct detection as $P(E)=1-P(C)$, hence the average error probability of the Kennedy receiver is $P(E)\underset{p_0=p_1}{=}\frac{1}{2}\exp(-4K_\alpha)$.

Approaching the Helstrom bound via signal-state rotation

It is noteworthy that with the Kennedy receiver the cancelled signal always results in correct detection, since no photons can occur when there is no signal pulse. In addition, doubling the signal amplitude for the alternate hypothesis increases the signal energy by a factor of four, greatly reducing the probability of a zero photon-count when a pulse is present: these are the primary the reasons why the Kennedy receiver achieves near-optimum performance. However, photon-counting detection implies that one of the measurement states should be aligned with the ground state, and as we have seen, this is not the condition under which optimum performance is achieved. The state-space representation of optimum detection described in [7] and illustrated in Fig. 1b shows that the measurement states must be symmetrically arranged with the signal states for optimum detection, not asymmetrically as with photon-counting detection. It is therefore natural to ask under what conditions optimum detection could be approached by starting with photon-counting detection, and rotating the signal-states into a more symmetrical configuration with respect to the measurement states.

An approximate state-space representation of photon counting for small signal energies is shown in Fig. 4a, where the measurement states are approximated by the number states $|0\rangle$ and $|1\rangle$, so that $|w_0\rangle=|0\rangle$ and $|w_1\rangle=|1\rangle$. With photon-counting detection, the signal state representing H_0 is aligned with the measurement-state, resulting in $|\psi_0\rangle=|w_0\rangle=|0\rangle$, whereas the alternate state $|\psi_1\rangle$ is rotated in the $(|0\rangle,|1\rangle)$ plane by an angle θ related to the overlap of the signal-states as $\theta=\cos^{-1}(|\langle\psi_0|\psi_1\rangle|)=\cos^{-1}(e^{-\frac{1}{2}|2\alpha|^2})$: for example, with $|\alpha|=0.2$ the angle between the signal states is $\theta=22.6$ degrees.

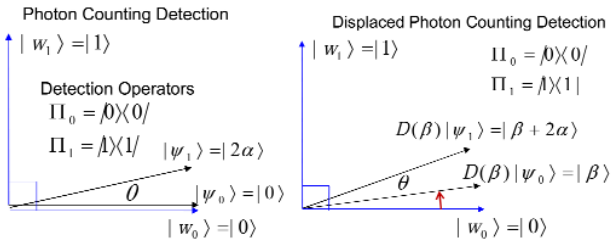


Fig. 4. Small signal energy representation of photon counting and displaced photon counting.

Recall from Fig. 1b that the measurement states should be placed symmetrically around the signal states in the signal subspace, the optimum rotation angle is $\varphi=\frac{1}{2}(\pi/2-\theta)$: for our example, the optimum rotation angle between the signal state $|\psi_0\rangle$ and its associated measurement state $|w_0\rangle$ should be $\varphi=33.7$ degrees, the same as between $|\psi_1\rangle$ and $|w_1\rangle$. From the overlap relation for coherent states, an angle of 33.7 degrees corresponds to an overlap of $\cos(33.7)=0.832=e^{-\frac{1}{2}|\beta|^2}$, yielding a displacement magnitude of $|\beta|=0.61$. This rotation can be accomplished by applying the displacement operator $D(\beta)$ to the signal states as indicated in Fig. 4b, where $|\beta|=0.61$ and $\arg(\beta)=\arg(\alpha)$. After displacement, the probability of finding $D(\beta)|\psi_0\rangle$ projected onto the next higher dimensional state $|2\rangle$, corresponding to a tilt in the signal subspace from the two-dimensional $(|0\rangle,|1\rangle)$ subspace into the three-dimensional $(|0\rangle,|1\rangle,|2\rangle)$ subspace, can be evaluated as $p(k=2)=|\langle 2|\beta\rangle|^2=|\beta|^4\ e^{-|\beta|^2}/2=0.04$. This is small enough to justify the two-dimensional measurement-state model, however this probability increases to

$|\langle 2|\beta+2\alpha\rangle|^2=0.183$ for $D(\beta)|\psi_1\rangle$, which is significantly greater than zero and hence cannot be ignored. Similarly, the probability of finding $D(\beta)|\psi_0\rangle$ projected onto any of the higher-dimensional states $|2\rangle,|3\rangle,\cdots$ is equal to the probability that it is not projected onto the states $|0\rangle$ or $|1\rangle$: $p(k\geq 2)=1-\sum_{k=0}^{1}|\langle k|\beta\rangle|^2$. These probabilities are shown in Fig. 5 as a function of $|\beta|$, from 0 to 1.

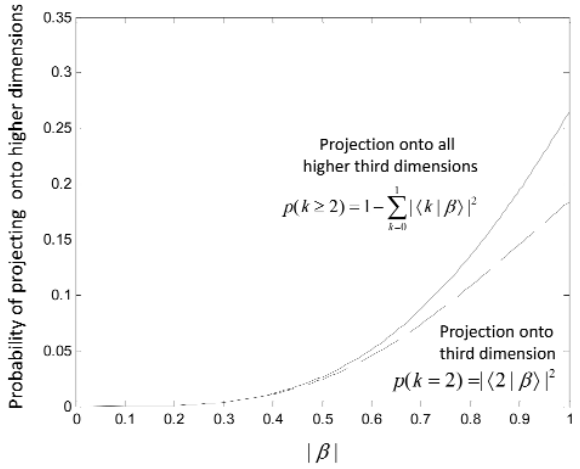


Fig. 5. Probability of finding a displaced ground state projected onto higher number state dimensions.

With the help of Fig. 5, we can argue that as long as the total displacement of the "pulse" state is less than approximately 0.2-0.3 in amplitude, the two-dimensional model should be accurate. For larger displacements, the projection onto third and higher dimensions starts to become significant, effectively tilting the signal subspace out of the two-dimensional $(|0\rangle,|1\rangle)$ subspace, hence the photon-counting interpretation is no longer accurate with larger displacements. This argument helps to explain why displacement followed by photon-counting detection approaches the optimum quantum measurement for small signal energies, but fails to reach it completely. However, the small-energy model still provides theoretical insights into the manner in which displacement followed by photon-counting detection approximates the Helstrom bound for small signal energies, and suggests approaches that may result in better receiver performance when small signal energies are involved.

The Optimized Kennedy Receiver

A displacement-optimized version of the Kennedy receiver, where the displacement does not cancel the null hypothesis exactly, but at the same time provides significant additional energy to the alternate, has been reported in [10], termed the "optimized displacement receiver". Here we provide an alternate derivation and interpretation of this idea. Since the displacement operator can be implemented with a strong local laser and a classical beamsplitter [9], the above discussion suggests that the performance of the Kennedy receiver could be improved in the small signal energy regime by first adding a phase-locked coherent field to the BPSK signals, detecting via photon-counting, then applying the optimum threshold defined in equation (7), which is valid for all displacements. The classical signal model for BPSK signals after displacement is shown in Fig. 6, where E_{LO} takes the place of the coherent state amplitude β.

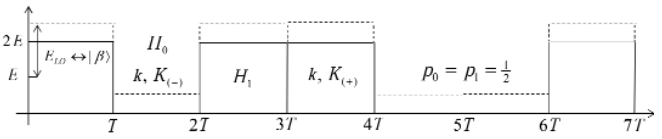


Fig. 6. Classical model of displaced BPSK signals, for the optimized Kennedy receiver.

The optimum value of the displacement for photon-counting detection can be derived by noting that for the small-energy region the value of the optimum threshold is always between 0 and 1. The goal of the optimization is to determine that value of β that maximizes the average probability of correct detection given β, $P(C)=\max_{\beta}P(C|\beta)$, or equivalently minimizes the average probability of error. With no loss in generality, assume that the signal amplitudes α,β are real, and write the conditional probabilities under the two hypotheses H_0 and H_1, given the displacement β, as $P(C|H_0,\beta)=\exp[-(\beta-\alpha)^2]$, $P(C|H_1,\beta)=1-\exp[-(\beta+\alpha)^2]$. Differentiating the conditional probability of correct detection, $P(C|\beta)=p_0P(C|H_0,\beta)+p_1P(C|H_1,\beta)$, with respect to β and solving the resulting transcendental equations numerically, the optimum displacement is that value of $\beta=\beta^*$ that satisfies the following transcendental equation as described in [10]: $p_0(\beta-\alpha)/p_1(\beta+\alpha)=\exp(-4\alpha\beta)$. This result is in contrast to the Kennedy receiver, where the signal fields are either cancelled completely or re-enforced by applying a displacement exactly equal to one of the signal amplitudes.

Applying the optimum displacement operator $D(\beta^*)$ to the binary signals results in the displaced signals $D(\beta^*)|-\alpha\rangle=|\beta^*-\alpha\rangle$ and $D(\beta^*)|\alpha\rangle=|\beta^*+\alpha\rangle$, with corresponding energies $|\beta^*-\alpha|^2=K_\alpha+K_\beta-2\sqrt{K_\alpha K_\beta}\equiv K_{(-)}$ and $|\beta^*+\alpha|^2=K_\alpha+K_\beta+2\sqrt{K_\alpha K_\beta}\equiv K_{(+)}$. It is easily shown that with displaced received fields and photon-counting detection the optimum threshold η is given by

$$\eta=\frac{\log_e(p_0/p_1)+4|\alpha||\beta|}{\log_e(|\beta+\alpha|^2/|\beta-\alpha|^2)} \qquad (7)$$

The optimum decision strategy calls for declaring H_1 if $k\geq\eta$, and H_0 if $k<\eta$. Note that non-zero counts are now possible even under H_0 due to the optimal displacement, unlike with the Kennedy receiver which displaced the signals sub-optimally by completely cancelling one of them. The relevant probabilities under the two hypotheses are given by

$$H_0: \quad p(k<\eta)=\sum_{k=0}^{\eta} K_{(-)}^k \exp(-K_{(-)})/k!$$

$$H_1: \begin{cases} p(k<\eta)=\sum_{k=0}^{\eta} K_{(+)}^k \exp(-K_{(+)})/k! \\ p(k\geq\eta)=1-\sum_{k=0}^{\eta} K_{(+)}^k \exp(-K_{(+)})/k! \end{cases}$$

For any signal energy, with optimal displacement the conditional probabilities of correct detection become $P(C|H_0)=p(k<\eta|K_{(-)});\ P(C|H_1)=1-p(k<\eta|K_{(+)})$, which must be averaged over the a priori probabilities to obtain the average probability of correct detection, finally yielding the average error probability as $P(E)=1-P(C)$. Due to the optimization of the displacement, these calculations are somewhat more involved than for the Kennedy receiver, as the following example illustrates.

Numerical Example

Consider the case $|\alpha|^2=0.2, |\alpha|=0.447$ with $p_0=p_1=\frac{1}{2}$. Solving the transcendental equation for β, we find the optimum displacement magnitude $|\beta|=0.757$ yielding the optimum threshold $\eta=0.86$ from equation (7). The conditional detection probability for the null hypothesis becomes $P(C|H_0)=p(k<\eta|K_{(-)})=\exp(-K_{(-)})=0.933$, whereas for the alternate hypothesis we obtain the following: $P(C|H_1)=1-p(k<\exp\eta|K_{(+)})=1-\exp(-K_{(+)})=0.765$. The average probability of correct detection follows directly as $P(C)=p_0P(C|H_0)+p_1P(C|H_1)=\frac{1}{2}(0.933+0.765)=0.849$, yielding $P(E)=1-P(C)=0.151$. This result is shown in Fig. 7 as the single point labeled "Numerical example" on the "Optimized Kennedy receiver" performance curve, which was computed numerically over the range of values $0<|\alpha|^2<0.8$. Note that the optimized Kennedy receiver described in [10] outperforms both the Kennedy and coherent receivers for all signal energies, including small signal energies where coherent detection actually approaches the Helstrom bound. However, it does not maintain this improvement over the Kennedy receiver for large symbol energies, but rather begins to approach the performance of the Kennedy receiver as the signal energy increases.

The above derivation suggests that by applying the optimum displacement to the BPSK signals prior to photon-counting detection, lower error probabilities will be obtained than possible with the Kennedy receiver, for any signal energy. The measurement-state derivation above also suggests that for the case of small signal energies, the Helstrom bound may be better approached by the optimized Kennedy receiver, since the displacement of the signal states is in the direction of the optimum measurement, where measurement states are placed symmetrically around the signal states in the $(|0\rangle,|1\rangle)$ plane.

As a heuristic check, we note that as the signal energy approaches zero the optimum displacement approaches $|\beta^*|^2=0.5$, or $|\beta|=0.707$ which now projects significantly onto higher number-state dimensions (as can be seen in Fig. 5, where the probability of projecting onto a higher dimension is seen to be 0.09). Hence the measurement-states no longer reside entirely in the $(|0\rangle,|1\rangle)$ plane, and the photon counting interpretation for displaced signal states is not strictly valid.

Nevertheless, displaced photon counting still approximates the optimum quantum measurement in this region, which explains why error probabilities close to the Helstrom bound can be achieved with the application of optimized displacement and photon-counting detection in this region.

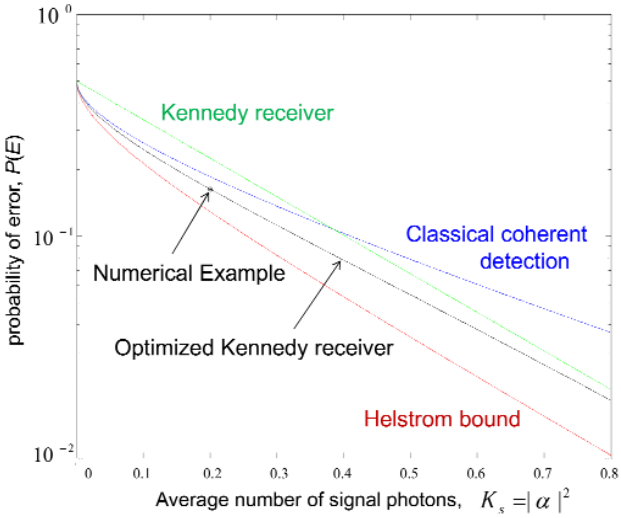


Fig. 7. Error probability performance of coherent, conventional Kennedy, and optimized Kennedy receivers.

Partitioned-Interval Detection Strategy

Consider the partitioned signal detection strategy illustrated in Fig. 8, where the original BPSK symbols have been converted to on-off signaling via matched displacement, as in the Kennedy receiver. Each T-second signal interval is now partitioned into two consecutive disjoint intervals: an initial interval of duration T_1 seconds, and a second interval of duration T_2 seconds. The average photon counts in these two intervals can be denoted as $4K_1$ and $4K_2$ respectively, with corresponding photon counts k_1 and k_2. The first interval is intended to be short, providing a small-energy (hence nearly quantum-optimum) "pre-detection" measurement, whereas the second interval is intended to supply more signal energy to further lower the error probability to acceptable levels for communication.

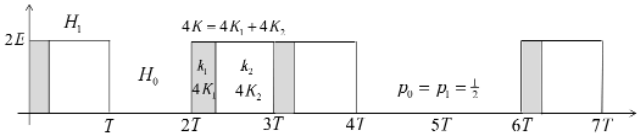


Fig. 8. Signal model for the partitioned interval detection strategy.

Based on the observation that for the Kennedy receiver correct detection occurs whenever the cancelled signal (null hypothesis) is observed, the strategy is to try to "guess" the correct hypothesis with a near-quantum-optimum measurement

in the first interval, and cancel the signal in the second interval by applying the appropriate displacement whenever a signal is pre-detected. If no signal is detected in the first interval, then the receiver continues to count photons in the second interval, without cancellation. This is similar to the approach used by the Dolinar receiver, which however must respond instantaneously to each photon-detection event within the signal interval, whereas here the counting intervals are determined based on predictable average signal energies instead of unpredictable photon occurrence times. Since with equal a priori probabilities roughly half of the original signal intervals contain no signal energy, it follows that any correct detection in the pre-detection interval will lead to more cancelled signals being observed in the second interval than in the original sequence. The final decision is based on the presence or absence of photon counts observed in the second interval, but also takes into account the outcome of the pre-detection measurement. The resulting two-step detection strategy can be summarized with the following algorithm:

If $k_1 > 0$, add 180° to the local field, and continue counting;
if $k_2 = 0$, decide "H_1"

If $k_1 = 0$, continue counting; if $k_2 > 0$, decide "H_1"
If $k_1 = 0$, continue counting; if $k_2 = 0$, decide "H_0"

This detection strategy is equivalent to a "modified sequence" interpretation, where some of the pulses in the second segment have been cancelled due to correct identification of the signal in the first segment. Restricting our attention to the second segment only, we find that this new sequence has more cancelled pulses than the original sequence where the a priori probabilities were equal. Therefore, if we observed only the modified sequence (where some of the original pulses have now been cancelled due to correct "pre-detection" decisions, but no new pulses have been added), then we would assign a higher probability to the occurrence of nulls in the second interval. Based on observing the modified sequence, we would conclude that the a priori probabilities p_0', p_1' of this new sequence were in fact not equal, but rather given by the expressions $p_0' = p_0 P(C|H_0) + p_1 P(C|H_1)$ and $p_1' = 1 - p_0'$. Representations of the modified sequence are shown in Fig. 9 a and b, where the intermediate decisions are shown in a) and the final sequence in b): for example, the second segment in the fourth symbol-interval ($3T < t \le 4T$) has been cancelled due to a correct decision in the first segment of this same interval, because a count $k_1 > 0$ has been observed in the first interval. The modified sequence therefore appears to have more null-hypotheses H_0' and fewer alternatives H_1'.

The decision strategy for the modified sequence shown in Fig. 9b, in terms of the modified a priori probabilities $p_0' > p_0$ and $p_1' < p_1$, can be stated as follows:

If $k_2 > 0$, declare H_1'; If $k_2 = 0$, declare H_0'.

However, now we must keep track of the correct decisions in the pre-detection segment that lead to pulse-cancellations, in order to detect the original message.

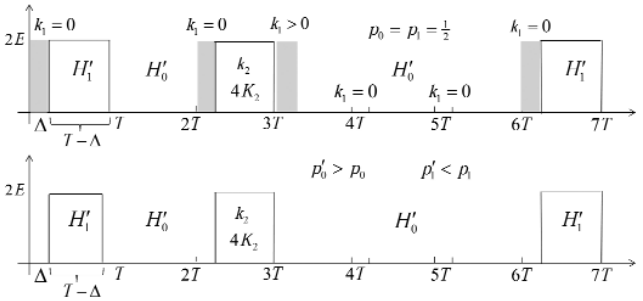


Fig. 9. a) Original and b) modified sequences, after a single correct pre-detection measurement.

We can now write the probability of correct detection and the probability of error for the modified sequence in terms of the modified a priori probabilities, as

$$\begin{aligned} P(C) &= p_0' P(C|H_0') + p_1' P(C|H_1') \\ &= (1-p_1') P(C|H_0') + p_1' P(C|H_1') \\ &= P(C|H_0') - p_1'[P(C|H_0') - P(C|H_1')] \end{aligned} \tag{8}$$
$$P(E) = 1 - P(C) = 1 - P(C|H_0') + p_1'[P(C|H_0') - P(C|H_1')]$$

Note that $P(C|H_0') \neq 1$ in general. Recalling that p_1' represents the probability of making an error in the first segment, and that the modified hypotheses $H_i'; i = 0,1$ refer to the second segment, we can see that the final error probability can always be improved by reducing p_1' if $P(C|H_0') > P(C|H_1')$, which is satisfied by the detection processes considered here, namely the Kennedy and optimized Kennedy receivers in the region of interest. Therefore, we can potentially choose a detection technique in the first segment that closely approaches the Helstrom bound for small signal energies, and perhaps a different detection strategy in the second segment, in order to achieve the desired error probability for communications applications. This result forms the basis of the "partitioned" approach, which we now examine for several cases of interest.

For the case of signal cancellation followed by photon-counting detection, the relevant probabilities for the modified sequence become

$$H_0': \quad p(k_2 = 0) = 1, \qquad H_1': \begin{cases} p(k_2 = 0) = \exp(-4K_2) \\ p(k_2 > 0) = 1 - \exp(-4K_2) \end{cases}$$

which yield the following conditional probabilities of correct detection: $P(C|H_0') = 1$, and for the alternate hypothesis $P(C|H_1') = 1 - \exp(-4K_2)$. Substituting into equation (8) yields $P(C) = p_0' + p_1'[1 - \exp(-4K_2)] = 1 - p_1' \exp(-4K)$, and error probability $P(E) = 1 - P(C) = p_1' \exp(-4K_2)$. Note that if $p_1' = \frac{1}{2}\exp(-4K_1)$, as would be obtained with photon-counting detection over the first segment, then the error probability after observing the entire symbol interval would

simply become $\frac{1}{2}\exp(-4K)$, which is exactly the same as for the Kennedy receiver, hence nothing would be gained. However, it also suggests that if $p_1' < \frac{1}{2}\exp(-4K_1)$ the error probability of the modified sequence will decrease correspondingly, resulting in $P(E) < \frac{1}{2}\exp(-4K)$. This observation provides a means for approaching the Helstrom bound by employing a better detection strategy in the first segment than simple field cancellation followed by photon counting detection, resulting in better overall performance. With $P_K(E)$, $P_c(E)$ and $P_{oK}(E)$ referring to the Kennedy, coherent and optimized Kennedy receivers respectively, the best strategy can be inferred from the ratio of the error probabilities $P_K(E)/P_c(E)$ and $P_K(E)/P_{oK}(E)$ in Fig. 10, which is interpreted as "gain over the Kennedy receiver". Note that in Fig. 10 N refers to the total number of segments used by the partitioned receiver, as explained subsequently in the section on the *Optimized N-segment receiver*.

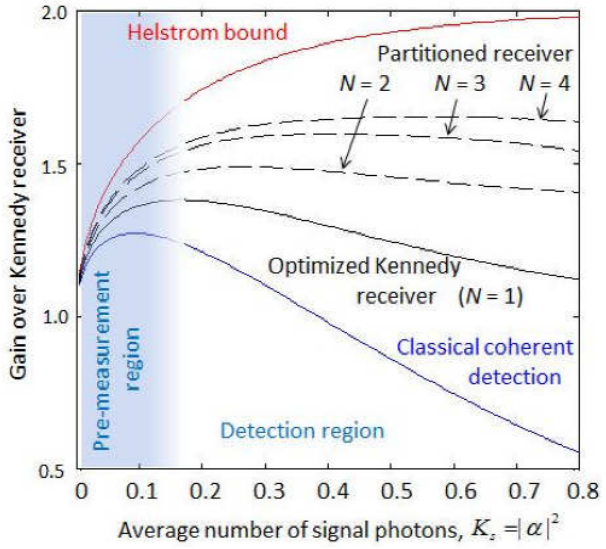


Fig. 10. Gain of coherent, optimized Kennedy, and partitioned receivers over the Kennedy receiver.

In Fig. 10, the coherent receiver peaks at $K_s = 0.095$ attaining a maximum gain of 1.272 over the Kennedy receiver, whereas the optimized Kennedy receiver peaks at $K_s = 0.165$, with a maximum gain of 1.381, after which both gains decrease: the gain of the optimized Kennedy receiver approaches 1 at high signal energies, reverting back to the conventional Kennedy receiver, whereas the gain of the coherent receiver continues to decrease towards zero.

Optimized N-segment receiver

The optimized two-segment detection approach described above can be extended directly to three or more segments, by considering the first N-1 segments of an N-segment receiver to be a "pre-detection" segment whose decision outcome modifies the a priori probabilities of the original sequence, thus improving the fidelity of the final decision. For example, the performance of a three-segment receiver can be evaluated by starting out as a two-segment receiver, but then partitioning the smaller pre-detection interval into two segments and optimizing each before optimizing the error probability for the third segment, further improving receiver performance. This procedure extends directly to an arbitrary number of segments, each step yielding an improvement over the previous step, but also increasing the complexity of the receiver.

The gain over the Kennedy receiver for up to four optimized segments reaches a maximum at slightly higher signal energies as can be seen in Fig. 10, which flatten out as the number of segments increase, effectively maintaining the maximum gain achieved by the pre-detection measurement over the region of interest. For three and four segment optimized receivers the maxima occur at $K_s = 0.38$ and $K_s = 0.62$ average signal photons. The gain curves can be divided roughly into two regions, a "pre-detection" region over which the gains increase rapidly, followed by a "detection" region over which the gains flatten out attempting to maintain maximum gain. The boundary between these two regions is roughly the initial small-energy region of up to approximately 0.2 signal photons, as shown in Fig. 10. This interpretation is in line with our previous conclusion that displacement followed by photon counting is close to the optimum strategy at small signal energies, hence we can interpret any measurement made within this region in a segmented receiver approach to be a pre-detection measurement: the use of multiple segments is merely a means to obtain better pre-detection performance. It should be noted that any other pre-detection strategy that improves upon these initial error probabilities will lead to gains in overall performance. Therefore, other measurement techniques that may be developed in the future could also be used to carry out pre-detection, potentially leading to further improvements in overall performance.

The error probability performance of optimized two, three and four segment receivers are shown in Fig. 11 along with that of the Kennedy receiver, coherent receiver and optimized Kennedy receiver for comparison. The partitioned receiver discussed here outperforms the previously known "near-optimum" approaches such as the Kennedy receiver for all signal energies, with gains of more than 2 dB over the Kennedy receiver at $P(E) = 0.1$, in the region of greatest interest for coded optical communications. This new approach effectively partitions the signal interval into two segments, a pre-detection segment that employs displaced photon counting to closely approach the Helstrom bound at small signal energies, followed by a detection segment that measures the remaining signal energy to achieve the desired communications performance.

For any number of predetermined segments N, the result of the first (N-1) decisions is incorporated into the probability of correct detection p_0' obtained from the first (N-1) segments, which is used as the a priori probability of hypothesis H_0' for the final segment. This strategy can be implemented using a bank of N lasers and switching between them using offset clocks operating at the symbol rate, hence it leads to practically implementable receivers for small values of N, but it also highlights the reason for suboptimal performance when

compared to the Dolinar receiver: since photons arrive randomly within any predetermined decision interval, the signal energy following a photon detection event in any sub-interval is effectively wasted with the partitioned-interval approach, since the contradicting decision could actually have be made as soon as the photon was detected.

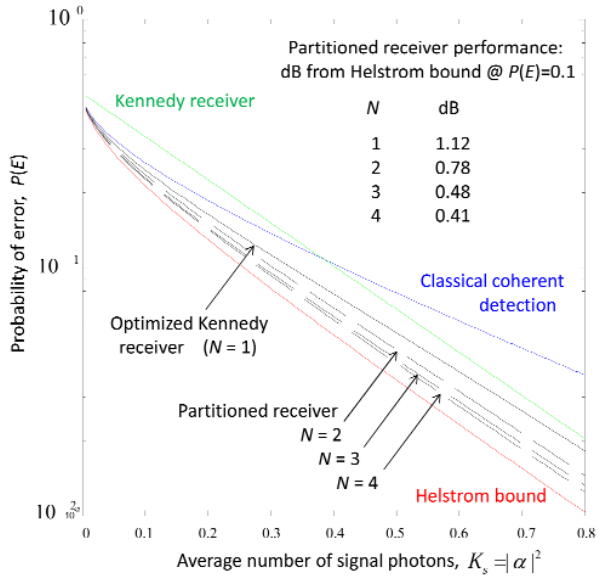


Fig. 11. Error probability performance and comparison of N-segment partitioned receivers.

However, responding instantly to each photon-occurrence requires processing bandwidth far exceeding the signal bandwidth, and hence leads to problems with implementation particularly at high data rates. This is one of the fundamental differences between the two approaches: the Dolinar receiver requires infinitely large processing bandwidths to reach the Helstrom bound, along with precisely controlled continuous-time laser intensities, whereas the partitioned-interval approach switches between a few local lasers with predetermined intensities at the signaling rate, but cannot approach the Helstrom bound arbitrarily closely for small, hence practical, values of N.

IV. Summary and Conclusions

An optical communications receiver concept capable of approaching the quantum limit in the region of interest for coded optical communications from deep-space, has been developed and analyzed in this paper. The key idea is to break up the signal interval into a short "pre-detection" segment followed by a longer validation segment in such a way as to optimize overall performance. This two-interval interpretation was extended to higher complexity N-interval detection by interpreting the processing in the first N-1 intervals as an improved pre-detection measurement, viewing the final N^{th} interval as the validation segment. It is shown that increasing N leads to improved performance for N = 2, 3, and 4 segments, arguably approaching the quantum limit for larger N but only at the cost of greater processing complexity. Therefore, this approach is intended primarily for low-complexity applications where improved receiver performance is deemed necessary. It was shown that with four disjoint segments, performance of the partitioned receiver approaches the Helstrom bound to within 0.41 dB, or equivalently improves upon the Kennedy receiver by 2 dB, at an error probability of 0.1 typically required by modern codes.

Acknowledgment

The research reported in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

[1] C. W. Helstrom, Quantum Detection and Estimation Theory, Mathematics in Science and Engineering, Academic Press, New York, 1976.

[2] M. Sasaki, O. Hirota, "Optimum decision scheme with a unitary control process for binary quantum-state signals," Physical Review A, Volume 54, 1996.

[3] C.-W. Lau, V. A. Vilnrotter, S. Dolinar, J. M. Geremia, and H. Mabuchi, "Binary Quantum Receiver Concept Demonstration," Interplanetary Network Progress Report 42-165, Jet Propulsion Laboratory, May 15, 2006.

[4] R. Cook, P. Martin, J. M. Geremia, "Optical coherent state discrimination using a closed-loop quantum measurement," Nature, vol. 446, 12 April 2007.

[5] A. Acin, E. Bagan, M. Baig, L. Masanes, R. Munoz-Tapia, "Multiple-copy state discrimination with individual measurements," Physical Review A, Volume 71, 2005.

[6] R. J. Glauber, "Coherent and Incoherent States of the Radiation Field, The Physical review, vol. 123, no. 6, 1963.

[7] V. A. Vilnrotter, C-W. Lau, "Quantum Detection and Channel Capacity for Communications Applications," Proceedings of SPIE, Vol. 4635, 2002.

[8] S. M. Barnett, P. M. Radmore, Methods in Theoretical Quantum Optics, Oxford Series in Optical and Imaging Sciences, Clarendon Press, Oxford, 1997.

[9] M. Paris, "Displacement operator by beam splitter," Physics Letters A, Elsevier, 1996.

[10] C. Wittmann, M. Takeoka, K. Cassemiro, M. Sasaki, G. Leuchs, U. Andersen, "Demonstration of Near-Optimal Discrimination of Optical Coherent States," Physical Review Letters, vol. 101, 21 November 2008.

Victor A. Vilnrotter (M'79, SM'02) was born in Kúnhegyes, Hungary, in 1944. He received his Ph.D. in electrical engineering and communications theory from the University of Southern California in 1978 and joined the Jet Propulsion Laboratory, Pasadena, Calif., in 1979, where he is a Principal Engineer in the Communication Architectures and Research Section. His research interests include electronic compensation of large antennas with focal-plane arrays, adaptive combining algorithms for antenna arrays, optical communications through atmospheric turbulence, the application of quantum communications to deep-space optical links, and the development of uplink array calibration and tracking technologies. He has published extensively in conferences and refereed journals, and has received numerous NASA awards for technical innovations, and a NASA Exceptional Service Medal for contributions in signal processing.

Classical and Quantum Genetic Optimization Applied to Coverage Optimization for Indoor Access Point Networks

Lajos Nagy, *Member, IEEE*

***Abstract*— The new focus of wireless communication is moving from voice to multimedia services. There is a growing interest in providing and improving radio coverage for mobile phones, short range radios and WLANs inside buildings. The need of such coverage appears mainly in office buildings, shopping malls, train stations where the subscriber density is very high. The cost of cellular systems and also the one of indoor wireless systems depend highly on the number of base stations required to achieve the desired coverage for a given level of field strength. There are already numerous optimization methods published which can be applied to the optimal design of such indoor networks [7,8,9, 10,11]. The fitness function of the optimization problem has numerous local optimum and therefore gradient based methods can not be applied. The recently published methods use any heuristic technique for finding the optimal Access Point (AP) positions. Common drawbacks of the methods are the slow convergence in a complex environment like the indoor one.**

The complexity of the selection procedure of a classical genetic algorithm is $O(NlogN)$ where N is the size of the population. The Quantum Genetic Algorithm (QGA) exploits the power of quantum computation in order to speed up genetic procedures. While the quantum and classical genetic algorithms use the same number of generations, the QGA outperforms the classical one in identifying the high-fitness subpopulation at each generation. In QGA the classical fitness evaluation and selection procedures are replaced by a single quantum procedure.

The article introduces the Quantum inspired Genetic Algorithm (QGA) for indoor access point position optimization to maximal coverage and compares with the Classical Genetic Algorithm (CGA).

***Index Terms*— optimization, radio network, indoor radiowave propagation**

I. INTRODUCTION

THE new focus of wireless communication is shifting from voice to multimedia services. User requirements are moving from underlying technology to the simply need reliable and cost effective communication systems that can support anytime, anywhere, any device. While a significant amount of traffic will migrate from mobile to fixed networks, a much greater amount of traffic will migrate from fixed to mobile networks. In many countries mobile operators are offering mobile broadband services at prices and speeds comparable to fixed broadband. Though there are often data caps on mobile broadband services that are lower than those of fixed broadband, some consumers are opting to forgo their fixed lines in favor of mobile. [3] There is a growing interest in providing and improving radio coverage for mobile phones, short range radios and WLANs inside buildings. The need of such coverage appears mainly in office buildings, shopping malls, train stations where the subscriber density is very high. The cost of cellular systems and also the one of indoor wireless systems depend highly on the number of base stations required to achieve the desired coverage for a given level of field strength. [12]

The design objectives can list in the priority order as RF performance, cost, specific customer requests, ease of installation and ease of maintenance. The first two of them are close related to the optimization procedure introduced and can take into account at the design phase of the radio network. There are already numerous optimization methods published which can be applied to the optimal design of such indoor networks [7,8,9,11,15].

The recently published methods use any heuristic technique for finding the optimal Access Point (AP) or Remote Unit (RU) positions. Common drawback of the methods are the slow convergence in a complex environment like the indoor one because all of the methods are using the global search space i.e. the places for AP-s are searched globally.

This article presents approaches in optimizing the indoor radio coverage using multiple access points for indoor environments. First the conventional Classical Genetic Algorithm (CGA) and Quantum inspired Genetic Algorithm (QGA) [1,2,16,17] is shortly introduced and applied to determine the optimal access point positions to achieve optimum coverage.

Finally the importance of applying this optimization process is certified by evaluating the indoor coverage area for different AP cardinality.

Manuscript received September 30, 2012.

Lajos Nagy is with the BME, Technical University of Budapest, Department of Broadband Infocommunications and Electromagnetic Theory, Egry József 18, Budapest, Hungary (phone: 36-1-4632790; fax: 36-1-4633289; e-mail: nagy@mht.bme.hu).

II. THE INDOOR PROPAGATION MODEL AND THE BUILDING DATABASE

In our path loss estimation the Motley-Keenan [6] model was used to analyze indoor wave propagation. This empirical type prediction model is based on considering the influence of walls, ceilings and floors on the propagation through disparate terms in the expression of the path loss.

The overall path loss according to this model can be written as

$$L = L_F + L_a \tag{1}$$

where L_F is the free space path loss and L_a is an additional loss expressed as

$$L_a = L_c + \sum_{i=1}^{I} k_{wi} L_{wi} + \sum_{j=1}^{J} k_{fj} L_{fj} \tag{2}$$

where L_c is an empirical constant term, k_{wi} is the number of penetrated i type walls, k_{fj} is the number of penetrated floors and ceilings of type j, I is the number of wall types and J is the number of floor and ceiling types.

For the analyzed receiver position, the numbers k_{wi} and k_{fj} have to be determined through the number of floors and walls along the path between the transmitter and the receiver antennas. In the original paper [6] only one type of walls and floors were considered, in order for the model to be more precise a classification of the walls and floors is important. A concrete wall for example could present very varying penetration losses depending on whether it has or not metallic reinforcement.

It is also important to state that the loss expressed in (2) is not a physical one, but rather model coefficients, that were optimized from measurement data. Constant L_c is the result of the linear regression algorithm applied on measured wall and floor losses. This constant is a good indicator of the loss, because it includes other effects also, for example the effect of furniture.

For the considered office type building, the values for the regression parameters have been found. (Table I)

The Motley-Keenan model regression parameters have been determined using Ray Launching (RL) deterministic radio wave propagation model. These calculations were made for the office-type building floor of the Department of Broadband Infocommunication and Electromagnetic Theory at Budapest University of Technology and Economics (Figure 1-2.). The frequency was chosen to 2450 MHz with a $\lambda/2$ transmitter dipole antenna mounted on the 2m height ceiling at the center of the floor.

The receiver antenna has been applied to evaluate the signal strength at (80x5)x(22x5)=44000 different locations in the plane of the receiver. At each location the received signal strength was obtained by RL method using ray emission in a resolution of 10. A ray is followed until a number of 8 reflections are reached and the receiver resolution in pixels has an area of 0.2*0.2 m2. The receiver plane was chosen at the height of 1.2 m.

TABLE I
THE REGRESSION PARAMETERS

Wall type	Nr. of Layers	Layer widths	Regression Parameter [dB]
Brick	1	Brick – 6 cm	4.0
Brick	1	Brick – 10 cm	5.58
Brick	1	Brick – 12 cm	6.69
Brick+ Concrete	3	Brick – 6 cm Concrete – 20 cm Brick – 6 cm	11.8
Brick+ Concrete	3	Brick 10 cm Concrete – 12 cm Brick – 10 cm	14.8
Brick+ Concrete	3	Brick – 6 cm Concrete 10 cm Brick – 6 cm	9.3
Brick	1	Brick – 15 cm	8.47
Concrete	1	Concrete – 15 cm	6.56
Concrete	3	Concrete – 15 cm Air – 2 cm Concrete – 15 cm	12.47
Glass	3	Glass – 3 mm Air – 10 cm Glass 3 mm	0
Plasterboard	1	Plasterboard – 5 cm	4.5
Wood	1	Wood 6 cm	0.92
Wood	1	Wood – 10 cm	0.17

The wall construction is shown on Fig. 1 made of primarily brick and concrete with concrete ceiling and floor, the doors are made of wood. The coefficients of the model have been optimized on the data gathered by the RL simulation session described above.

Fig. 1.a. Floor view of V2 building at BME

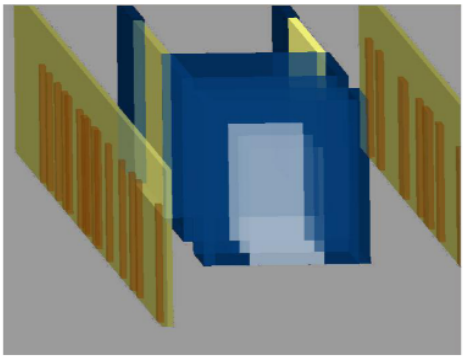


Fig. 1.b. Polygon data base of V2 building at BME

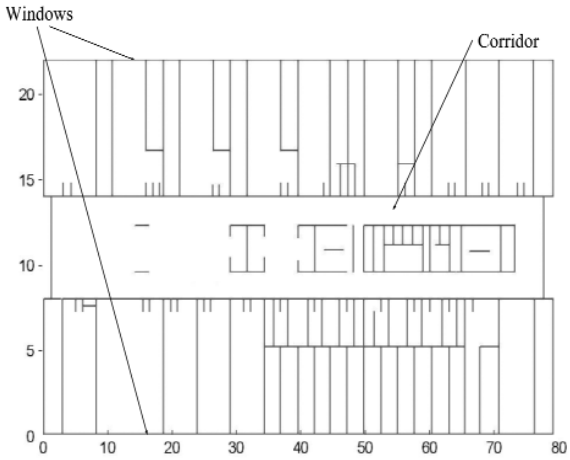


Fig. 2. The building database

The geometrical description of the indoor scenario is based on the concept that the walls has to be partitioned to surrounding closed polygons and every such polygons are characterized by its electric material parameters.

The data base for the ray tracing method in our applications can not contain cut-out surfaces directly, such as windows, doors. Therefore the cut-out surface description is based on surface partitioning of the geometry.

III. Optimization Methods

There are already numerous optimization methods published which can be applied to the optimal design of such radio indoor networks [7,8,9,11,15]. The recently published methods use any heuristic technique for finding the optimal AP positions.

In the paper two global optimization methods the Classical Genetic Algorithm (CGA) and Quantum inspired Genetic Algorithm (QGA) global search algorithm are used with wave propagation solver as can be seen in Fig. 3.

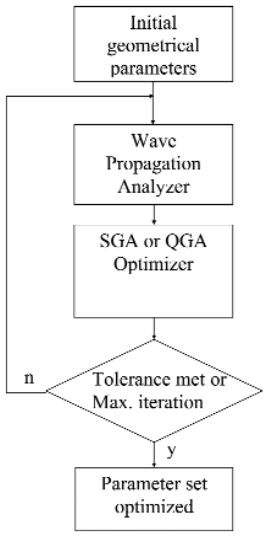


Fig. 3. Diagram of Wave Propagation analyzer and optimizer

Heuristic search and optimization is an approach for solving complex and large problems that overcomes many shortcomings of traditional (gradient type) optimization techniques. Heuristic optimization techniques are general purpose methods that are very flexible and can be applied to many types of objective functions and constraints. Another advantage of heuristic methods is their simplicity because of its gradient-free nature. Gradient free optimization methods are primarily based on the objective function values and are suitable for problems either with many parameters or with computationally expensive objective functions.

A. Optimization Method through Classical Genetic Algorithms (CGA)

Genetic Algorithms are increasingly being applied to complex problems. Genetic Algorithm optimizers are robust, stochastic search methods modeled on the principles and concepts of natural selection. [5,7,10,14] GA are increasingly being applied to difficult optimization problems. GA optimizers are robust, stochastic search methods modeled on the principles and concepts of natural selection. (Fig. 4.)

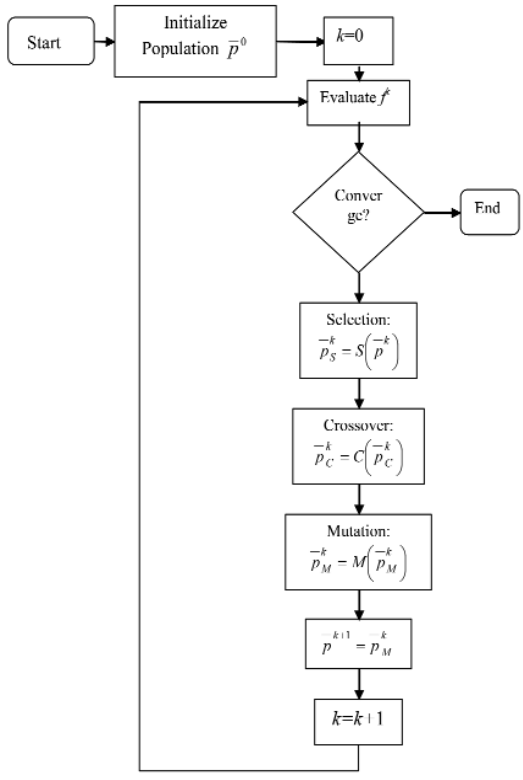


Fig. 4. The flowchart of a simple GA

If a receiver position that is fully described by N_{par} parameters arranged in a vector $x=\{x_i|\ i=1,...,N_{par}\}$ is considered, then the knowledge of x permits the evaluation of the objective function $f(x)$, which indicates the worth of a design (the area coverage percentage). It is assumed that x_i take on either real or discrete values, and that $f(x)$ needs to be maximized.

The GA does not operate on x but on a discrete representation or chromosome $p=\{g_i|\ i=1,...,N\}$ of x, each parameter x_i being described by a gene g_i. Each gene g_i in turn consists of a set of N_{all}^{i} all that are selected from a finite alphabet and that together decode a unique x_i.

The GA does not limit themselves to the iterative refinement of a single coded design candidate; instead the classical GA (CGA) simultaneously acts upon a set of candidates or population

$$\bar{p} = \{p(i)|i = 1,...,N_{pop}\} \tag{3}$$

where N_{pop} is the population size.

Starting from an initial population $\bar{p}^0$, the CGA iteratively constructs populations $\bar{p}^k, k = 1..N_{gen}$, with N_{gen} denoting the total number of CGA generations. Subsequent generations are constructed by iteratively acting upon $\bar{p}^0$ with a set of genetic operators. The operators that induce the transition $\bar{p}^k \to \bar{p}^{k+1}$ are guided solely by knowledge of the vector of objective function values

$$f^k = \left\{f\left(x\left(p^k(i)\right)\right)\middle| i = 1..N_{pop}\right\} \tag{4}$$

and induce changes in the genetic makeup of the population leading to a $\bar{p}^{k+1}$ comprising individuals that are, on average better adapted to their environment than those in $\bar{p}^k$, i.e., they are characterized by higher objective function values.

This change is effected by three operators mentioned in the introduction: selection (*S*), crossover (*C*), and mutation (*M*).

The selection operator implements the principle of survival of the fittest. Acting on $\bar{p}^k$, *S* produces a new population $\bar{p}_S^k = S\left(\bar{p}^k\right)$ again of size N_{pop} that is, on average, populated by the better-fit individuals present in $\bar{p}^k$. Among the many existing schemes tournament selection has been chosen. The crossover operator mimics natural procreation. Specifically, *C* acts upon the population $\bar{p}_S^k$ by mating its members, thereby creating a new population

$$\bar{p}_C^k = \bigcup_{i=1}^{N_{pop}/2} C\left(ch\left(\bar{p}_S^k\right), ch\left(\bar{p}_S^k\right)\right) \tag{5}$$

where the chromosome crossover operator *C* selects a random crossover allele a_{Ncross} between the two chromosomes to be crossed upon which it acts with probability P_{cross}.

The mutation operator generates a new population of size by introducing small random changes into $\bar{p}_C^k$. The action of *M* can be represented in operator form as

$$\bar{p}_M^k = \bigcup_{i=1}^{N_{pop}} M\left(\bar{p}_C^k(i)\right) \tag{6}$$

The cost function of the optimization procedure has been the coverage percentage of the points for which the received power is greater than a given level.

$$f = c\left(P_{rec}\right) = \frac{\text{Number of points}\,(P_{thresh} < P_{rec})}{\text{Total number of test points}} \tag{7}$$

B. Optimization Method through Quantum inspired Genetic Algorithm (QGA)

GQA is based on the concepts of qubits and superposition of states of quantum mechanics.[16,17] The smallest unit of information stored in a two-state quantum computer is called a quantum bit or qubit. A qubit may be in the '1' state, in the '0' state, or in any superposition of the two. The state of a qubit can be represented as

$$|\Psi\rangle = \alpha|0\rangle + \beta|1\rangle \tag{8}$$

where α and β are complex numbers that specify the probability amplitudes of the corresponding states. $|\alpha|^2$ and $|\beta|^2$ gives the probability of finding the qubit in logical value '0' or '1' if the state has been measured. Normalization of the state to unity guarantees

$$|\alpha|^2 + |\beta|^2 = 1 \tag{9}$$

It is possible to use a number of different representations to encode the solutions onto chromosomes in evolutionary computation. The classical representations can be broadly classified as: binary, numeric, and symbolic. GQA uses a novel representation that is based on the concept of qubits. One qubit is defined with a pair of complex numbers, (α,β) and for *m* qubits as

$$\begin{bmatrix} \alpha_1 & \alpha_2 & .. & \alpha_m \\ \beta_1 & \beta_2 & .. & \beta_m \end{bmatrix} \tag{10}$$

This representation has the advantage that it is able to represent any superposition of states.

QGA maintains a population of qubit chromosomes, $Q(t) = \{\mathbf{q}_1^t, \mathbf{q}_2^t, ... \mathbf{q}_n^t\}$ at the generation *t*, where n is the size of population and $\mathbf{q}_j^t$ is a qubit chromosome defined as in (10).

$$\mathbf{q}_j^t = \begin{bmatrix} \alpha_1^t & \alpha_2^t & .. & \alpha_m^t \\ \beta_1^t & \beta_2^t & .. & \beta_m^t \end{bmatrix} \tag{11}$$

The observed values of $Q(t)$ states can be generated taking into account the α_j probabilities. One binary representation of the j-th qubit is $\mathbf{x}_j^t, j = 1,2,\ldots,n$ and the observation of the $Q(t)$ is $P(t) = \{\mathbf{x}_1^t, \mathbf{x}_2^t, \ldots, \mathbf{x}_n^t\}$.

The set of qubit chromosomes $Q(t)$ is updated in each evolution step by applying appropriate quantum gates (Q-Gates) $G(\theta)$, which are evaluated taking into account the best fitness solution and given by the rotational angle selection strategy (Table II) [14]. This step makes the qubit chromosomes converge to the fitter states.

$$G(\theta_i) = \begin{bmatrix} \cos(\theta_i) & -\sin(\theta_i) \\ \sin(\theta_i) & \cos(\theta_i) \end{bmatrix} \quad (12)$$

TABLE II
ROTATIONAL ANGLE SELECTION STRATEGY OF Q-GATE

x_i	b_i	f(x)≥f(b)	$\Delta\theta_i$	$\alpha_i\beta_i>0$	$\alpha_i\beta_i<0$	$\alpha_i=0$	$\beta_i=0$
0	0	0	$10^{-3}\pi$	-	+	±	±
0	0	1	$10^{-3}\pi$	-	+	±	±
0	1	0	0.08π	-	+	±	±
0	1	1	$10^{-3}\pi$	-	+	±	±
1	0	0	0.08π	+	-	±	±
1	0	1	$10^{-3}\pi$	+	-	±	±
1	1	0	$10^{-3}\pi$	+	-	±	±
1	1	1	$10^{-3}\pi$	+	-	±	±

The flowchart of the QGA is showed in the following.

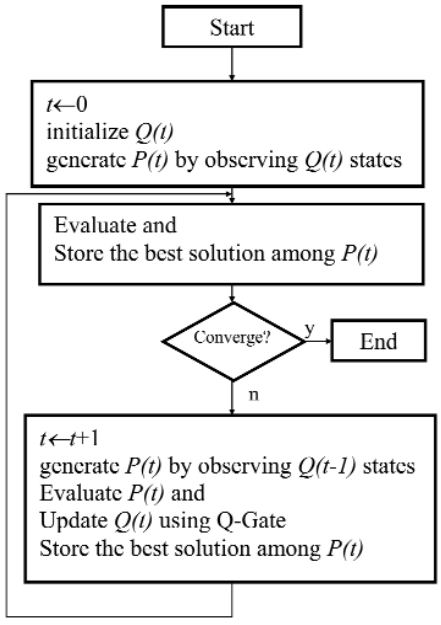


Fig. 5. The flowchart of QGA

Learning the description of the two versions of Genetic algorithms there are two significant differences between a classical and quantum versions or computer and a quantum computer realization.

The first is in storing information, classical bits versus quantum q-bits. Each quantum state represents many possible values of observation therefore the QGA increases the searching space.

The second is the quantum mechanical feature known as entanglement, which allows a measurement on some qubits to effect the value of other qubits.

The last section compares the algorithms and shows results on coverage results.

IV. RESULTS

The optimization procedure characterized as searching space with multiple local optimums and in this first session short investigation will be shown evaluating objective function for one access point in indoor environment.

The Fig. 6 shows the objective function which is the covered area percentage for the (X,Y) points as AP. If for instance the AP position is at (18,11) than the coverage is more than 30%, but if at (45,19) than the coverage is less than 15%. The Fig. 6 illustrate unambiguously the 'good' positions for access points having best coverages.

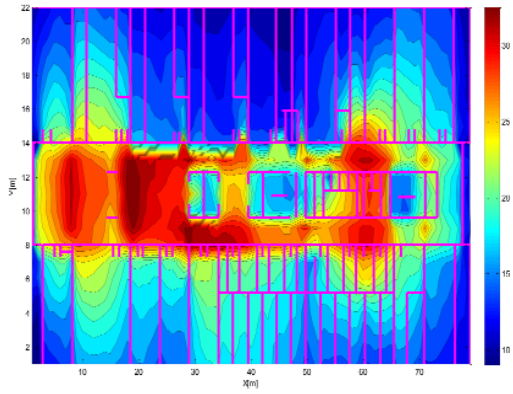


Fig. 6. Objective function for 1 AP

The Fig. 6 clearly shows the multiple local maximums of the objective function and therefore the motivation to apply heuristic optimization methods.

The brute force search which would be a possible optimization search doesn't give the expected result because of the huge computational demand. (TABLE III)

TABLE III
EXHAUSTIVE (BRUTE FORCE) SEARCH

Number of AP-s	Resolution of search space	Computation time	Result of optimization
1AP	1m x 1m grid (1738 points)	5.5 min	33.67% (19;12)
2AP	1m x 1m grid	159 hours (estimated)	
1AP	0.5m x 0.5m grid	22 min	34.57% (18.5;12.5)
2AP	0.5m x 0.5m grid	637 hours (estimated)	

Next the convergence comparison will be introduced for CGA and QGA for one access point. The testing results of the algorithms are shown in Fig. 7.

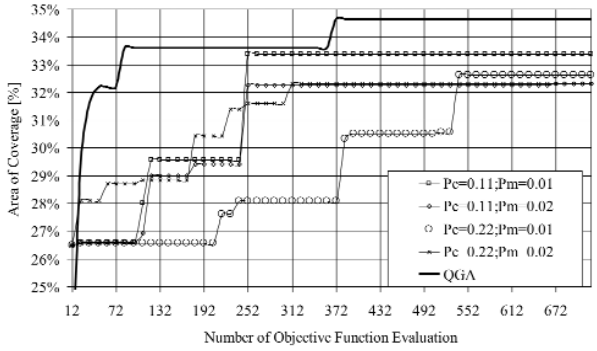


Fig. 7. Comparison of SGA and QGA (best objective function)

For CGA four set of parameters were tested two mutation and two crossover probability sets are evaluated.

Based on the comparison in Fig. 7 and other not detailed evaluations can be stated that for our indoor one access point coverage optimization case the QGA outperforms the classical SGA therefore it worth to investigate and deploy for more complex optimization cases with multiple access points.

In the last part of the results chapter coverage results are shown for 3, 4 and 6 access points.

The first scenario is an optimization on AP positions (circles in Fig. 8) of the half part of the floor. The Fig. 8 shows the original 4 AP positions which were chosen to best coverage in laboratories and the corridor coverage was not an aim.

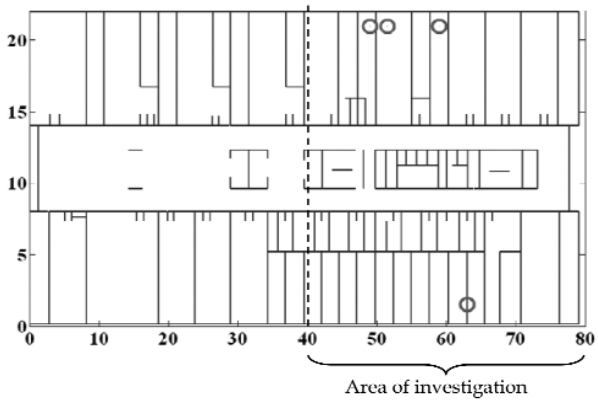


Fig. 8. Original (not optimized) AP positions

The Fig. 9 shows the optimal AP positions using the cost function of (7). The simulated distribution of received power for the optimized geometry is shown in Fig. 10 with the measured results.

To make the measurements we have chosen WLAN APs and the power levels were measured using laptops with external wireless adapter moved on the area of investigation. 90 sampling points in distances of 1 m were chosen on the level and the comparison of Fig. 10 shows a good agreement for the received power distribution.

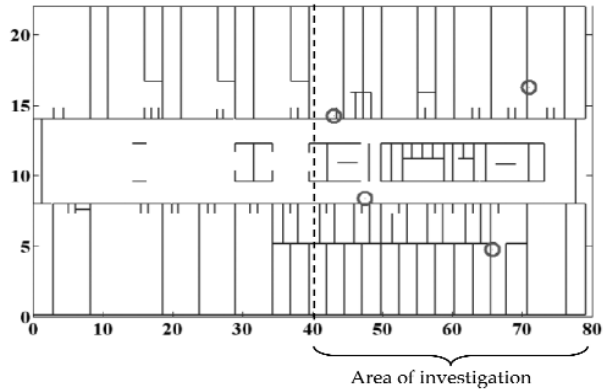


Fig. 9. Optimized AP positions

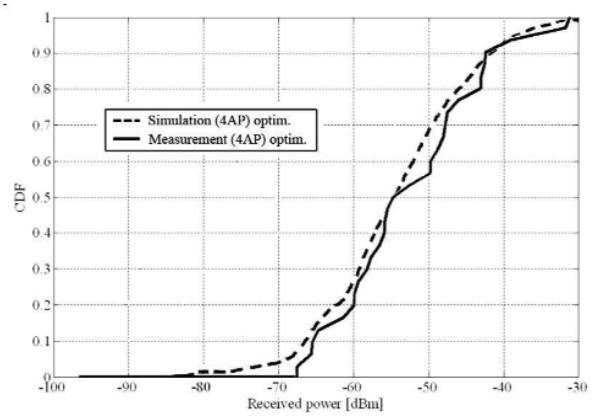


Fig. 10. Cumulative Density Function of received power level (optimized)

The most important change in the distributions of optimized and not optimized cases is increased number of points with proper coverage. (TABLE IV.)

TABLE IV
AREA COVERAGE FOR OPTIMIZED AND NOT OPTIMIZED CASE

Configuration	Not optimized	Optimized
Coverage for P_{rec}>-60dBm (simulation)	40%	75%
Coverage for P_{rec}>-60dBm (measurement)	50%	80%

The second simulation is on the entire floor level and the aim of the simulation is to compare the necessary number of APs for the same area coverage.

The Fig. 11 shows plausible positions of APs and the Fig. 12 the optimized ones.

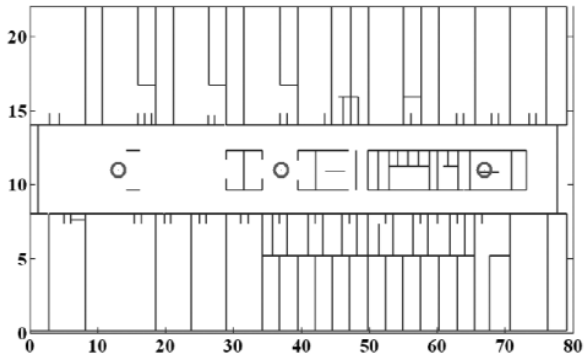


Fig. 11. Plausible AP positions

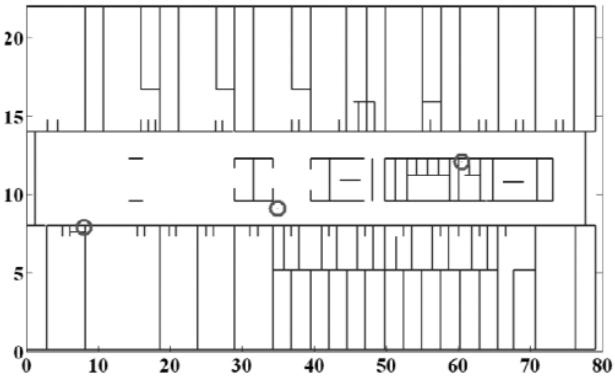


Fig. 12. Optimized AP positions

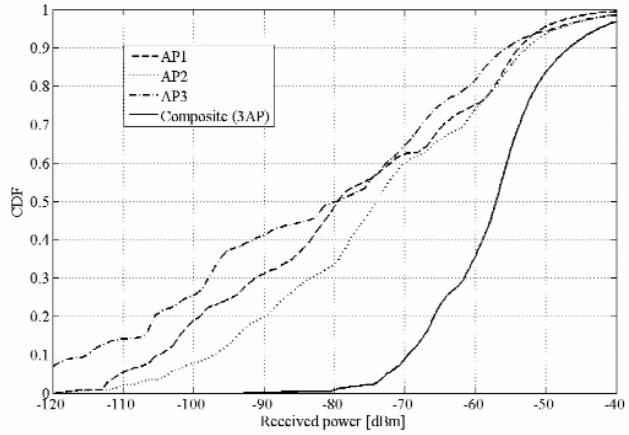


Fig. 13. Independent and composite CDF (optimized AP positions)

The Fig. 14 and TABLE V summarize the importance of APs position of radio network. With the proper choice of the placement the optimized 3 AP network configuration results nearly the same coverage as the configuration 6 AP with APs installed in plausible positions.

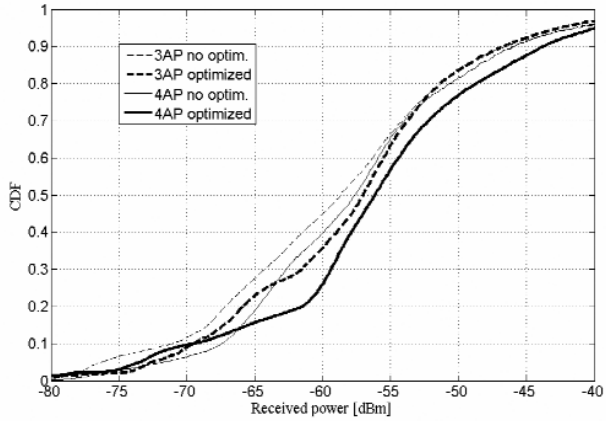


Fig. 14. Optimized and not optimized CDF using 3 and 4 APs

TABLE V
AREA COVERAGE FOR OPTIMIZED AND NOT OPTIMIZED CASE

Configuration	3AP	4AP	6AP
Coverage (not optimized)	40%	60%	66%
Coverage (optimized)	65%	75%	87%

These results (TABLE V) illustrate and justify well the importance of Access Point installation positions in radio networks in order to maximize the wireless coverage. Using the mentioned optimization procedure the network cost can be significantly reduced and the optical distribution network also can be simplified.

V. CONCLUSION

The optimal Access Point position of radio network is investigated for indoor environment. The article illustrates the possibility of optimization of radio network using Genetic Algorithm and Quantum inspired Genetic Algorithm in order to determine positions of APs. The QGA as new approach is introduced to solve the global optimization problem. The methods are introduced and investigated for 1,2, 3 and 6 AP cases. The influence of Genetic Algorithm parameters on the convergence has been tested, the algorithms are compared for the one AP case and the optimal radio network is investigated. It has been shown that for finding proper placement the necessary number of APs can be dramatically reduced and therefore saving installation cost of WLANs.

The results clearly justify the advantage of the method we used but further investigations are necessary for convergence comparison for multiple AP case. Other promising direction is the extension of the optimization cost function with interference parameters of the wireless network part and with outer interference.

REFERENCES

[1] Nielsen, Michael A.; Chuang, Isaac L. (2000). Quantum Computation and Quantum Information. Cambridge, UK: Cambridge University Press. ISBN 978-0-521-63235-5.

[2] Sandor Imre, Ferenc Balazs, Quantum Computing and Communications: An Engineering Approach, ISBN: 978-0-470-86902-4, November 2004, Wiley

[3] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2009-2014, (white paper), 2010 http://cisco.biz/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.pdf

[4] Daniel E. Finkel (2003). DIRECT Optimization Algorithm User Guide, http://www.ncsu.edu/crsc/reports/ftp/pdf/crsc-tr03-11.pdf

[5] E. Michielssen, Y. Rahmat-Samii, D.S. Weile (1999). Electromagnetic System Design using Genetic Algorithms, Modern Radio Science, 1999.

[6] J.M. Keenan, A.J. Motley (1990). Radio Coverage in buildings, BT Tech. J., 8(1), 1990, pp. 19-24.

[7] Lajos Nagy, Lóránt Farkas (2000). Indoor Base Station Location Optimization using Genetic Algorithms, PIMRC'2000 Proceedings, Sept. 2000, London, UK

[8] Liza K. Pujji, Kevin W. Sowerby, Michael J. Neve (2009). A New Algorithm for Efficient Optimization of Base Station Placement in Indoor Wireless Communication Systems, 2009 Seventh Annual Communication Networks and Services Research Conference, Moncton, New Brunswick, Canada, ISBN: 978-0-7695-3649-1

[9] Martin D. Adickes, Richard E. Billo, Bryan A. Norman, Sujata Banerjee, Bartholomew O. Nnaji, Jayant Rajgopal (2002). Optimization of indoor wireless communication network layouts, IIE Transactions, Volume 34, Number 9 / September, 2002, Springer,

[10] Lóránt Farkas, István Laki, Lajos Nagy (2001). Base Station Position Optimization in Microcells using Genetic Algorithms, ICT'2001, 2001, Bucharest, Romania

[11] Portilla-Figueras, S. Salcedo-Sanz, Klaus D. Hackbarth, F. López-Ferreras, and G. Esteve-Asensio (2009). Novel Heuristics for Cell Radius Determination in WCDMA Systems and Their Application to Strategic Planning Studies, EURASIP Journal on Wireless Communications and Networking, Volume 2009 (2009)

[12] R.D. Murch, K.W. Cheung (1996). Optimizing Indoor Base-station Locations, XXVth General Assembly of URSI, 1996, Lille, France

[13] R. E. Schuh, D. Wake, B. Verri and M. Mateescu, Hybrid Fibre Radio Access (1999) A Network Operators Approach and Requirements, 10th Microcoll Conference, Microcoll'99, Budapest, Hungary, pp. 211-214, 21-24 March, 1999

[14] Z. Michalewicz (1996). Genetic Algorithms + Data Structures = Evolution Programs, Springer-Verlag, Berlin, 1996.

[15] Yufei Wu, Samuel Pierre (2007).Optimization of 3G Radio Network Planning Using Tabu Search, Journal of Communication and Information Systems, Vol. 22, No. 1, 2007

[16] Zakaria Laboudi, Salim Chikhi (2012). Comparison of Genetic Algorithm and Quantum Genetic Algorithm, Proceedings of International Arab Journal of Information Technology, 2012, pp. 243-249.

[17] Mohammed, A.M., Elhefhawy, N.A., El-Sherbiny, M.M., Hadhoud, M.M. (2012) Quantum crossover based quantum genetic algorithm for solving non-linear programming, 8th International Conference on Informatics and Systems (INFOS), 14-16 May 2012

Lajos Nagy (M'06) Lajos Nagy received the Engineer and PhD degrees, both from the Budapest University of Technology and Economics (BME), Budapest, Hungary, in 1986 and 1995, respectively. He has been the head of Department of Broadband Infocommunications and Electromagnetic Theory in 2007. He is a lecturer at BME on Antennas and radiowave propagation, Radio system design, Signals and systems. His research interests include antenna analysis and computer aided design, electromagnetic theory, radiowave propagation, communication electronics, signal processing and digital antenna array beamforming. Member of Hungarian Telecommunication Association, Hungarian Committee Secretary of URSI and chapter chair of Hungarian IEEE joint chapter MTT, ED, AP, ComSoc.

The Problem of Testing a Quantum Gate

Subhash Kak

***Abstract*— We consider the question of testing of quantum gates as a part of the larger problem of communication through circuits that use a variety of such gates. We argue that the correct outputs for the basic input values to the two-qubit gate are not sufficient to guarantee satisfactory validation of its workings for all values. We present reasons why these gates may not be error free, and as non-ideal gates they need to be tested for a wide range of probability amplitudes, and we argue that the burden of this additional testing is substantial. Experimental implementations of the controlled-NOT gate have substantial non-unitarity and residual errors. Some analytical results on the complexity of the problem of quantum gate testing are presented.**

***Index Terms*— Non-ideal gates, quantum computing, quantum information, testing a quantum gate**

I. INTRODUCTION

THE problem of testing a quantum gate has no analog in the classical world. Basically, the problem is this: To test a quantum gate we need certified quantum gates to generate all possible inputs and since such gates are not available at this time how are we to certify a gate that has been submitted for certification? Put differently, physical implementations of the gate will be linear only over a restricted range of inputs, whereas quantum computing demands that the gates be completely linear.

To understand the scope of the problem, note that a quantum bit (qubit) $a|0\rangle + b|1\rangle$ is different from a classical bit in the sense that it is associated with arbitrary sets of complex values of *a* and *b* as long as $|a|^2 + |b|^2 = 1$. The variables *a* and *b* are probability amplitudes and their mod squares are the probabilities of the component states associated with the qubit. The process of interaction with the qubit causes it to collapse to either $|0\rangle$ or $|1\rangle$ and, consequently, the amount of information that can be extracted from a single qubit is one bit. Unlike a classical state, a quantum state is a superposition of mutually exclusive component states and an unknown quantum state cannot be cloned. Quantum states are also characterized by the Heisenberg Uncertainty Principle. Communication with qubits is of interest to engineers because it represents the use of polarization states of light seen as a stream of single qubits which is appropriate in certain applications using very weak laser outputs. It is also of interest in the problem of quantum cryptography which promises unbreakable security under certain conditions.

The theory of quantum gates is at the basis of quantum computing [1] and just as the hardware for a classical computer is constituted of circuits of logic gates, the practical implementation of a quantum computer is as a circuit made of quantum gates. The problem of testing a quantum gate is, therefore, of interest beyond quantum communication.

In principle the problem of gate testing need not be too restrictive if one can show that the gate works for the widest range of inputs. But in reality the number of inputs to test for increases exponentially as the assumed granularity of data increases. Thus for the case of the Pauli-X gate $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, it is not enough to show that the input $|0\rangle$ become $|1\rangle$ and vice versa and one should be able to establish that the gate transforms $a|0\rangle + b|1\rangle$ into $b|0\rangle + a|1\rangle$ for *all possible* complex values of a and b. To establish this requires that precise values of can be generated for all *a* and *b* using some other certified gate and then checking the output of the Pauli-X gate being tested. One would need to either use other exact gates to steer the output state to one of the directional bases of the measurement apparatus to confirm this, or to perform a quantum tomographic analysis of the output state [1] that will establish it in a probabilistic sense. It would also be essential to establish that the sum of the squares of the absolute value of the probability amplitudes equals one at the output to ensure that there is no dissipation within the system.

In this paper, we consider the question of testing of quantum gates from the point of view of complexity. We consider implications of a certain desired accuracy at the output of the circuit for the corresponding accuracy of individual gates. Since errors in a linear (or approximately linear) circuit add up, the constraints on individual gates are higher than for the overall circuit. Recent results on the experimental implementations of the controlled-NOT (which is the quantum analog of the classical XOR gate in which the first bit is mod-2 added to the second bit while leaving the first bit unchanged) and the Toffoli gates (CCNOT or universal reversible logic gates) have substantial non-unitarity and residual errors. Since these gates cannot be completely error free, as non-ideal elements they need to be tested for a wide range of probability amplitudes, and the burden of this additional testing is exponential with respect to the number of qubits and quadratic with respect to the desired precision.

Manuscript received September 28, 2012, revised December 4, 2012. This work was supported by National Science Foundation grant #1117068..
Subhash Kak is with the Oklahoma State University, Stillwater, OK, USA (phone: 405-744-6036; fax: 405-744-9097; e-mail: subhash.kak@okstate.edu).

II. QUANTUM COMPUTING USING GATES

Quantum computing has enriched computing theory beyond classical techniques in doing Fourier transform faster (albeit the solution is not fully available), factorization in approximately log n rather than $\sqrt{n}$ steps, search for an item with a specific property in an unordered database in roughly $\sqrt{n}$ rather than $n/2$ steps, solving Pell's equation (x^2-ny^2=1) in polynomial time, and it has applications to cryptography [1]. But even in theory, quantum computing can only solve problems where the solution amounts a rotation in an appropriate space (as in primality testing since integers form a multiplicative group). The allure of finding new problems that could be shown to be solved faster by quantum algorithms and the challenge of building quantum computing machines has made the field popular amongst mathematicians and physicists. It has also spurred research in materials science and in the experimental area for finding good candidates for implementing quantum circuits.

Given this background, it is worthwhile to investigate prospects for practical quantum computing. Soon after proposals for quantum computing were advanced, it was agreed that problems of decoherence and noise precluded physical implementation of these computers [2]. The implementation had to deal with the following dilemma: The computing system should not interact with the environment lest it decohere, and at the same time it should strongly and precisely interact with the control circuitry. The question of direction of flow of information in a quantum circuit is also an issue. The presence of the arrow of time implies that the system is in a state of non-equilibrium and is, therefore, essentially dissipative [3]. Thus quantum computing itself promotes decoherence and the directionality of information flow implies correlation in the resulting noise associated with the process.

To deal with errors, quantum error correction codes have been proposed although they have not yet been physically implemented. The implementation of these codes would depend on the correct working of very many quantum gates, and, therefore, testing and certifying quantum gates is essential to their effectiveness. There are other questions related to the assumptions under which quantum error correction can be effective, but these questions will not be considered here.

Fault-tolerant conceptual schemes assume that a system could be built recursively (e.g. [4], page 40) where each noisy gate is replaced by an ideal gate by the use of error correction circuitry, but such an idea appears to be impractical. Even if it is assumed that the errors are below a certain threshold, correction in the quantum context can work only if it is related to a known state; in contrast, error correction in classical computation works for unknown states. Certain gate faults, which lie outside of the set of assumptions underlying quantum error correction, cannot be corrected [5].

A non-recursive so-called Fibonacci scheme [6] has been proposed as another approach to quantum error correction. But it assumes the existence of perfect gates in the use of teleportation for error correction. It also uses Bell states, that is two-qubit states in which the qubits are maximally entangled, the generation of which requires the existence of perfect gates. The scheme also assumes that there are no gates with leakage or correlated faults.

It is not possible for this to be true for the quantum case because the unknown received state, for example, could be $a|0\rangle + b|1\rangle$ or $c|0\rangle + e^{i\alpha}d|1\rangle$, with unknown *a, b, c, d,* α, and if one did not know which one of the two is the correct state, one cannot know what operation is needed on the received state. It may be argued that the initialized state in a quantum computation is known [7], but that cannot be claimed for the intermediate states in the quantum circuit, which would also need to be provided with error correction.

Even assuming that the noise levels are below the threshold associated with the Threshold Theorem and there are enough error-correction resources and the errors are not correlated, there would be the problem of errors introduced before the error correction process begins. There is the additional problem of the presence of gauge states in quantum computing [8].

III. COMPLEXITY OF TESTING

Consider a circuit with k gates and n qubits at input and output. Let the accuracy desired in the output qubits be expressed by δ quantization levels. This means that the error will be $\varepsilon=1/\delta$. For example, for the Pauli-X gate with a single qubit $a|0\rangle + b|1\rangle$, both a and b must be associated with δ quantization levels making for a total of δ^2 quantization regions.

Figure 1 represents the situation schematically where the input probability amplitudes a and b are replaced at the output by new values b^* and a^*. We cannot assume that the circuit is linear over the entire range and, therefore, testing for each of these regions is necessary.

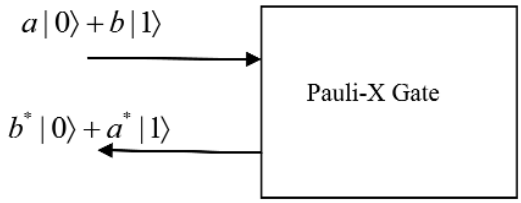


Figure 1. Testing the X gate

Since a real gate will be dissipative to a degree and not be linear over the entire range, it also follows that the square of the estimated probability amplitudes at the output will not equal 1. In other words, $|a^*|^2 + |b^*|^2 \neq 1$.

Theorem 1. The testing of a circuit with n qubits and δ quantization levels is associated with information of n+2 log δ.
Proof. A circuit of n qubits implies a total of 2^n input lines. Since each of these lines has complex input so the total number of quantization levels to be considered is $2^n\delta^2$.

Theorem 2. For a circuit with k stages, the information associated with the testing of each stage is n+2 log kδ.
Proof. Since noise is additive, the accuracy at each stage must be k times better than for the entire circuit.

The amount of testing to be performed at each gate, therefore, is of the order of $2^n k^2\delta^2$, which is exponential in the number of qubits and quadratic in the number of stages.

Thus for a quantum circuit with 20 qubits and 30 stages, and a desired circuit accuracy of one percent, the amount of testing to be done at each stage for the multi-qubit gate will equal $2^{20} \times 30^2 \times 100^2$. Clearly, for the solution of any meaningful problem, which would require thousands of gates, the constraints on the accuracy of the individual units will be well-nigh impossible to achieve.

IV. Experimental Results on Quantum Gate Testing

DeMarco *et al* [9] present the best-known implementation of the controlled-NOT gate in which the qubits are the spin and the internal energy (ground and second excited) states of a $^9Be^+$ ion. Table 1 presents the performance and we can see the probabilities do not sum to unity because these data represent the results of four separate experiments.

The fact that checking the performance of the gate in this limited setting itself requires different experiments with different control conditions is symptomatic of the difficulty of testing and it shows that the physical implementation cannot be taken to be an ideal gate.

Table 1: Experimental results of controlled-NOT gate [9]

	↓	↑
n=0	0.989 ± 0.006	0.050 ± 0.007
n=2	0.019 ± 0.007	0.968 ± 0.007

The authors mention errors in the initial state preparation and "limited gate fidelity" as contributing to the less than perfect performance of the gate. Since the probabilities do not add up to one, one can be sure that the behavior of the gate has a dissipative component.

For ease of comparison with the standard gate transformation, one may rewrite the results of Table 1 by representing the internal energy and the spin states in terms of the same qubits as follows:

$$|00\rangle \xrightarrow{CNOT} 0.989\,|00\rangle + 0.050\,|10\rangle$$
$$|10\rangle \xrightarrow{CNOT} 0.019\,|00\rangle + 0.968\,|11\rangle$$

which may be idealized to

$$|00\rangle \xrightarrow{CNOT} |00\rangle$$
$$|10\rangle \xrightarrow{CNOT} |11\rangle$$

Since we don't expect the idealized transformation to be achieved in practice, testing the gate for only two points for a transformation whose range (in terms of the probability amplitudes) is continuous over complex values in the range (0,1) is unsatisfactory. This is like testing the transformation $y=x^2$ for two values of x and taking that to be true for all values.

The correctness of the transformation for two values cannot be extended to other values because we do not know if the experimental arrangement is fully quantum (as the gate is non-ideal) and the influence of the *controlling circuitry cannot be taken to be the same for all values of the probability amplitudes*.

In another implementation of the controlled-NOT gate, based on a string of trapped ions [10] whose electronic states represent the qubits where the control is exercised by focused laser beams, the fidelity of the gate operation *was in the range 70 to 80%* [11]. These were ascribed to laser frequency noise, intensity fluctuations, detuning error, residual thermal excitation, addressing error, and off-resonant excitation. Clearly, this method is not a promising candidate for creating a precise functioning controlled-NOT gate.

For another (c. 2008) optical fiber quantum controlled-NOT gate [12], the authors claim an average logical fidelity of 90% and process fidelity in the range 0.83 to 0.91. The estimated second range is broader because of the substantial errors in the photon sources. Once again the performance is far from ideal.

It is important to note that although the name of the gate is controlled-NOT, the transformation is not that of one input controlling the other in all cases of the input qubits.

As example, if the input state is $\frac{1}{2}(|0\rangle+|1\rangle)(|0\rangle-|1\rangle)$ the output state is $\frac{1}{2}(|0\rangle-|1\rangle)(|0\rangle-|1\rangle)$, in which the first qubit has been transformed under the influence of the second qubit rather than the other way around. In general, we can consider a contolled-NOT gate to be controlled by both the qubits in some proportion, but this is clearly seen only if different superpositions of the qubits are considered.

To test a *non-ideal* controlled-NOT gate, we should be able to show that:

$$a\,|00\rangle + b\,|01\rangle + c\,|10\rangle + d\,|11\rangle$$
$$\xrightarrow{CNOT} a\,|00\rangle + b\,|01\rangle + c\,|11\rangle + d\,|10\rangle$$

within certain limits of error. In other words, the complex amplitudes (*a, b, c, d*) should map into (*a, b, d, c*) with some small error. It is true if the gate was ideal then testing beyond the two cases of 10 and 11 would not be necessary. But since any physical implementation of the gate would not be ideal, one needs to check the working for the entire range of probability amplitudes to ensure that the errors are within the acceptable threshold.

The testing of the physical gate operating on two qubits requires that various (*a, b, c, d*) amplitudes be generated with perfect fidelity, which requires that a perfectly functioning two-qubit gate (that may be controlled-NOT gate) should already exist. Since each of the values (*a, b, c, d*) is complex and over (0,1) and testing causes collapse to the components along the measurement bases, certification that a quantum gate is in *perfect working order* with no errors whatsoever is *computationally impossible*. It should also be noted that, in contrast, the errors of classical gates are completely correctable.

To test a non-ideal controlled-NOT gate properly, we first need a perfectly functioning controlled-NOT gate to generate a variety of complex superpositions for two qubits. Current experimental evidence on the implementation of the controlled-NOT gate indicate non-unitarity and substantial errors that preclude useful implementations of quantum algorithms.

V. Implementations using Superconducting Circuits

Recently, several studies have investigated the use of superconducting circuits for the implementation of quantum gates. In one of these, the implementation of a Toffoli gate was done with three superconducting *transmon* qubits coupled to a microwave resonator [13] where the *transmon* is a type of superconducting charge qubit that has reduced sensitivity to charge noise.

By exploiting the third energy level of the *transmon* qubits, the authors reduced the number of elementary gates needed for the implementation of the gate (that requires six controlled-NOT gates and ten single-qubit operations), relative to that required in theoretical proposals using only two-level systems. Such reduction should have improved the reliability of the gates, but using full process tomography and Monte Carlo process certification, the authors measured fidelity of only 68.5 ± 0:5 per cent.

Another study using superconducting circuits examines the three-qubit code, which maps a one-qubit state to an entangled three-qubit state [14]. The authors implement the correcting three-qubit gate, the conditional-conditional NOT (CCNOT) or Toffoli gate, with a 85±1% fidelity to the expected classical action of this gate and 78 ± 1% fidelity to the ideal quantum process matrix. They also performed a single pass of both quantum bit- and phase-flip error correction with 76 ± 0.5% process fidelity.

In the above-mentioned studies the results, for the most basic testing of the gate, are thus disappointing. One reason behind this performance may be the statistical constraints that need be satisfied [15].

VI. Comparing Classical and Quantum Paradigms

The difficulties of gate testing are another manifestation of the fact that the quantum computing paradigm is fundamentally different from the classical one. This may be seen in the consideration of quantum teleportation, which is transporting an unknown quantum state to a remote location using Bell states and classical communication. While mathematically it is easy to show that quantum teleportation works, there is no way one can certify that a given unknown state has been teleported correctly. The pair of resource qubits in Bell correlation in the experimental arrangement may not be correctly entangled and the gates may also have errors. Furthermore, a customer may not be sure that a specific quantum object supplied to him by the Certification Authority is not entangled with some other object. The fact that quantum mechanics is not a theory of individual objects but rather of collectives [16] has unexpected consequences for certification of transactions in an information network.

A quantum system composed of several subsystems can be in an entangled state, in which the properties of the full system are well defined but the properties of each subsystem are not well defined. This is also the reason why it is difficult to account for quantum mechanical effects in biological system [17]-[20].

The difficulty of implementing quantum gates (and of testing them) creates substantial burdens in the practical realization of the quantum circuit model of computing. It is no wonder that progress in physical implementations is limited and there is no unanimity on what kind of physical qubits to use [6].

Unlike classical computing, we cannot test the performance of given quantum gates while using imprecise signal generators because the requirement of linearity should hold over all possible values [21]. The time taken for a gate to operate creates another complication in a concatenation of gates. For example, the time taken was 63 ns in the superconducting circuit implementation of a CCNOT gate. Furthermore, the gate times are not constants and thus the propagation of qubits in the quantum circuit will suffer from timing issues [22].

In a classical circuit the accuracy of individual gates can be less than that of the entire circuit and we can create a reliable circuit using relatively less reliable component gates [23]. The extension of this for quantum processing is true in a very restrictive sense. In quantum computing the threshold theorem claims that a noisy quantum computer can accurately and efficiently simulate any ideal quantum computation provided that noise is weakly correlated and its strength is below the critical quantum accuracy threshold. But there is no evidence that errors in quantum gates fall below this threshold and that noise is only weakly correlated [24]-[27].

Given that the sources of errors in the gates are many that interact in a variety of modes, one might suppose that a threshold is associated with an error parameter ε so that below ε_0 the error is linear but above ε_0 it becomes uncontrollable large. Such a consideration is meaningful in classical computing where one can estimate in advance, or otherwise limit, the

range of signals that drive the computation. No such limit can be assigned for quantum computing as the probability amplitudes, in principle, have all possible values – with no constraints on phase values – so long as their absolute value remains less than one.

VII. CONCLUSIONS

We conclude that the representational difference between classical and quantum states has the most profound implications for their circuit implementations. Mapping classical bits into corresponding qubits increases the representation space by virtue of the complex probability amplitudes that get associated with the component states and their superposition. While this provides advantages in the solution of certain problems, it also creates new difficulties in the control and testing of quantum hardware.

If using quantum circuits one can, in principle, solve an exponential number of problems at the same time (although the solution to only one of these may be accessed), it also increases the size of the control space exponentially. The problems of controlling a quantum gate and testing it become correspondingly harder. Quantum computing systems do not scale. What quantum mechanics giveth by one hand, it taketh away by another.

REFERENCES

[1] M.A. Nielsen and I.L. Chuang, Quantum Computation and Quantum Information. Cambridge University Press, 2000.
[2] R. Landauer, The physical nature of information. Phys. Lett. A 217, 188- 193, 1996.
[3] D.K. Ferry, Quantum computing and probability. J. Phys.: Condens. Matter 21 474201, 2009.
[4] P. Aliferis, Level Reduction and the Quantum Threshold Theorem. Ph.D. Thesis, Caltech, 2007; arXiv:quant-ph/0703230v1
[5] S. Kak, Information complexity of quantum gates. Int. Journal of Theoretical Physics 45: 933-941, 2006.
[6] P. Aliferis and J. Preskill, The Fibonacci scheme for fault-tolerant quantum computation. arXiv:0809.5063
[7] S. Kak, The initialization problem in quantum computing. Foundations of Physics 29: 267-279, 1999.
[8] A. Bruno, A. Capolupo, S. Kak, G. Raimondo and G. Vitiello, Gauge theory and two level systems. Mod. Phys. Lett. B, vol. 25, pp. 1661-1670, 2011.
[9] B. DeMarco, A. Ben-Kish, D. Leibfried,V. Meyer, M. Rowe, B.M. Jelenkovic, W.M. Itano, J. Britton, C. Langer, T. Rosenband, and D. J. Wineland, Experimental demonstration of a controlled-NOT wave-packet gate. Phys. Rev. Lett. 89: 267901-1-4, 2002.
[10] J.I. Cirac and P. Zoller, Quantum computations with cold trapped ions. Phys. Rev. Lett. 74: 4091-4094, 1995.
[11] F. Schmidt-Kaler, H. Häffner, M. Riebe, S. Gulde, G. P. T. Lancaster, T. Deuschle, C. Becher, C. F. Roos, J. Eschner and R. Blatt, Realization of the Cirac-Zoller controlled-NOT quantum gate. Nature 422: 408-411, 2003.
[12] A.S. Clark, J. Fulconis, J.G. Rarity, W.J. Wadsworth, J.L. O'Brien, An all optical fibre quantum controlled-NOT gate. arXiv:0802.1676
[13] A. Fedorov, L. Steffen, M. Baur, M.P. da Silva, and A. Wallraff, Implementation of a Toffoli gate with superconducting circuits. Nature 481, 170, 2012.
[14] M. D. Reed, L. DiCarlo, S. E. Nigg, L. Sun, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, Realization of three-qubit quantum error correction with superconducting circuits. Nature 482, 382, 2012.
[15] S. Kak, Statistical constraints in starting a quantum computation. Pramana, Journal of Physics 57: 683-688, 2001.
[16] A. Peres, Quantum Theory: Concepts and Methods. Springer, 1995.
[17] S. Kak, The three languages of the brain: quantum, reorganizational, and associative. In: K. Pribram and J. King (editors), Learning as Self-Organization. Lawrence Erlbaum, Mahwah, 185–219, 1996. .
[18] S. Kak, Active agents, intelligence, and quantum computing. Information Sciences 128: 1-17, 2000.
[19] M.I. Franco, L. Turin, A. Mershin, E.M.C. Skoulakis, Molecular vibration-sensing component in Drosophila melanogaster olfaction. Proceedings of the National Academy of Sciences of the USA 108: 3797–3802, 2011.
[20] P. Ball, The dawn of quantum biology. Nature 474: 272-274, 2011.
[21] S. Kak, Information, physics and computation. Foundations of Physics 26: 127-137, 1996.
[22] S. Ashhab, P. C. de Groot, F. Nori, Speed limits for quantum gates in multi-qubit systems. Phys. Rev. A 85, 052327, 2012.
[23] J. von Neumann, Probabilistic logics and the synthesis of reliable organisms from unreliable components. In C. E. Shannon and J. McCarthy, editors, Automata Studies, volume 3, pages 43–99. Princeton University Press, Princeton, 1956.
[24] R. Alicki, Quantum memory as a perpetuum mobile? Stability v.s. reversibility of information processing. Open Systems & Information Dynamics 19, 2012.
[25] G. Kalai, How quantum computers fail: Quantum codes, correlations in physical systems, and noise accumulation. arXiv:1106.0485
[26] M. Dyakonov, Is fault-tolerant quantum computation really possible? In: Future Trends in Microelectronics. Up the Nano Creek, S. Luryi, J. Xu, and A. Zaslavsky (eds), pp. 4-18, Wiley, 2007.
[27] M. Dyakonov, Revisiting the hopes for scalable quantum computation. arXiv:1210.1782

Subhash Kak is Regents Professor and Head of the Department of Computer Science at Oklahoma State University at Stillwater. Prior to joining Oklahoma State University, he served for many years as the Delaune Distinguished Professor of Electrical and Computer Engineering at Louisiana State University in Baton Rouge. He is the author of several books that include The Nature of Physical Reality (New York, Peter Lang, 1986) and The Architecture of Knowledge (New Delhi, CRC, 2004). His areas of interest include data security, quantum computing, information theory, neural networks, and history of science.

Professor Kak's awards include British Council Fellow (1976), Science Academy Medal of the Indian National Science Academy (1977), Kothari Prize (1977), UNESCO Tokten Award (1986), Goyal Prize (1998), National Fellow of the Indian Institute of Advanced Study (2001), and Distinguished Alumnus of IIT Delhi (2002).

Special Issue on Applied Cryptography – Guest Editorial

Václav (Vashek) Matyáš, Zdeněk Říha, and Petr Švenda

THIS special issue focuses on the area of applied cryptography, bringing up selected papers from Santa's Crypto Get-Together (SantaCrypt), a workshop that runs since 2001 as an annual Czech and Slovak workshop aiming to facilitate closer cooperation of professionals working in the field of applied cryptography and related areas of security.

The first paper "Attacking Scrambled Burrows-Wheeler Transform" of Martin Stanek of a recent proposal for a modification of the Burrows-Wheeler transform (BWT). The BWT is a commonly used transform in lossless compression algorithms. The BWT does not compress the data itself, instead it is usually the first step in a sequence of algorithms transforming an input data into compressed data. The modification – Scrambled Burrows-Wheeler transform is an attempt to combine encryption and data compression. The paper shows that the proposed approach is insecure, presents chosen plaintext and known plaintext attacks and estimates their complexity in various scenarios.

The second paper "Two Improvements of Random Key Predistribution for Wireless Sensor Networks – Revised Version" of Jiri Kur et al. won the student KEYMAKER competition at SantaCrypt, and its first version was presented at the 8th International Conference on Security and Privacy in Communication Networks. This work deals with the area of random key predistribution in wireless sensor networks. Two novel improvements enhancing security provided by the random key predistribution schemes are proposed and analyzed. The first improvement exploits limited length collisions in secure hash functions to increase the probability of two nodes sharing a key. The second improvement introduces hash chains into the key pool construction to directly increase the resilience against a node capture attack.

The third paper "Privacy Scores: Assessing Privacy Risks Beyond Social Networks" of Michal Sramka focuses on the concept of privacy scores that were proposed in the past to provide each user with a score – a measurement of how much sensitive information a user made available for others on a social network website. This paper discusses their shortcomings, and shows several research directions for their extensions. The author also proposes an extension that takes the privacy score metric from a single social network closed system to include background knowledge, and argues for the need to include publicly available background knowledge in the computation of privacy scores in order to get scores that more truthfully reflect the privacy risks of the users.

The last paper "Accelerating Biometric Identification" of David Naccache et al. deals with biometric identification. As opposed to biometric matching, biometric identification is a relatively costly process because it involved a number of template comparisons. The paper discusses the problem of the optimization of biometric identification. The main idea is to test the most probable candidates first.

Václav (Vashek) Matyáš is a Professor at the Masaryk University, Brno, CZ, and serves as a Vice-Dean for Foreign Affairs and External Relations, Faculty of Informatics. His research interests relate to applied cryptography and security, publishing over a hundred peer-reviewed papers and articles, and co-authoring six books. He was a Fulbright Visiting Scholar with Harvard University, Center for Research on Computation and Society, and also worked with Microsoft Research Cambridge, University College Dublin, Ubilab at UBS AG, and was a Royal Society Postdoctoral Fellow with the Cambridge University Computer Lab. Vashek was one of the Editors-in-Chief of the Identity in the Information Society journal, and he also edited the Computer and Communications Security Reviews, and worked on the development of Common Criteria and with ISO/IEC JTC1 SC27. Vashek is a member of the Editorial Board of the Infocommunications Journal and a Senior Member of the ACM. He received his PhD degree from Masaryk University, Brno and can be contacted at matyas AT fi.muni.cz.

Zdeněk Říha is an Assistant Professor at the Masaryk University, Faculty of Informatics, in Brno, Czech Republic. He received his PhD degree from the Faculty of Informatics, Masaryk University. In 1999 he spent 6 months on an internship at Ubilab, the research lab of the bank UBS, focusing on security and usability aspects of biometric authentication systems. Between 2005 and 2008 he was seconded as a Detached National Expert to the European Commission's Joint Research Centre in Italy, where he worked on various projects related to privacy protection and electronic passports. He was involved in the ePassport interoperability group known as the Brussels Interoperability Group. Zdeněk has been working with the WG 5 (Identity management and privacy technologies) of ISO/IEC JTC 1/SC 27. Zdeněk's research interests include smartcard security, PKI, security of biometric systems and machine readable travel documents. Zdeněk can be contacted at zriha AT fi.muni.cz.

Petr Švenda is an Assistant Professor at the Masaryk University, Faculty of Informatics, in Brno, Czech Republic. He received his PhD degree from Masaryk University, working in the area of the cryptographic protocols for restricted environments such as wireless sensor networks, with focus on automatic generation of cryptographic protocols with the help of evolutionary algorithms. In 2008, he worked at TU Dresden on secure logging for the AN.ON anonymity service. He is also interested in practical aspects of security in cryptographic smartcards and their resistance against side-channel attacks and properties of random number generators available on smartcards and mobile devices. Petr can be contacted at svenda AT fi.muni.cz.

Attacking Scrambled Burrows-Wheeler Transform

Martin Stanek

***Abstract*—Scrambled Burrows-Wheeler transform [6] is an attempt to combine privacy (encryption) and data compression. We show that the proposed approach is insecure. We present chosen plaintext and known plaintext attacks and estimate their complexity in various scenarios.**

***Index Terms*—Burrows-Wheeler transform (BWT), scrambled BWT, secure compression, cryptanalysis.**

I. Introduction

The Burrows-Wheeler transform (BWT) [2] is a commonly used transform in lossless compression algorithms. The BWT does not compress the data itself, instead it is usually the first step in a sequence of algorithms transforming an input data into compressed data. The most prominent example of BWT-based compression is bzip2 program [8], which uses basically the following sequence of algorithms: the BWT, the move-to-front transform (MTF), and Huffman coding.

In practice, there are many situations with the simultaneous requirements of data compression and privacy (encryption). A common approach is compress-then-encrypt paradigm with independent algorithms used for compression and encryption. While widely preferred, the algorithms should be properly combined in order to avoid possible attacks [1], [5]. Another approach is to unify compression and encryption. However, designing a secure, "encrypting" variant of compression method is not an easy task. Many attempts were broken successfully [3], [7], [9].

Recently, Oğuzhan Külekci [6] proposed a novel approach – scrambled BWT – to combine data compression with privacy requirement. The scrambled BWT uses a secret lexicographic order of underlying alphabet as a secret key. In order to thwart some weaknesses, the author proposed to accompany the scrambled BWT with modified MTF that uses the secret lexicographic order as well.

a) Our contribution: We show that the proposed scrambled BWT with MTF is completely insecure and can be attacked easily (in the sense of chosen plaintext as well as known plaintext attacks). In case of known plaintext attacks we estimate experimentally the bit security of the scrambled BWT with MTF in various scenarios.

We briefly introduce the "standard" BWT, the MTF, and the scrambled BWT in Section II. Section III contains our analysis of the proposal, and shows chosen-plaintext and known-plaintext attacks on the scrambled BWT with MTF.

II. Preliminaries

Let $\mathcal{A}$ be an alphabet with size $L = |\mathcal{A}|$. Let N denotes a block length. A cyclic rotation of string/block $x = x_0x_1 \dots x_{N-1} \in \mathcal{A}^N$ with offset k is string/block $x^{(k)} = x_kx_{k+1} \dots x_{k+N}$ where all the indices are computed modulo N. The w-th symbol of $x^{(k)}$ is denoted as $x_w^{(k)}$, for $0 \le w < N$.

For real-world scenarios one can expect $L = 256$ (using bytes as an alphabet) and the block size N several hundreds kilobytes (e.g. bzip2 uses default block size 900kB).

M. Stanek is with the Department of Computer Science, Faculty of Mathematics, Physics and Informatics, Comenius University, Slovak Republic. (e-mail: stanek@dcs.fmph.uniba.sk).

A. "Standard" BWT

Let $x = x_0x_1 \dots x_{N-1} \in \mathcal{A}^N$ be an input block. The BWT sorts all cyclic rotations $x^{(0)}, x^{(1)}, \dots, x^{(N-1)}$ of input block x in lexicographic order. Let $j_0, j_1, \dots, j_{N-1}$ be a permutation of $\{0, 1, \dots, N-1\}$ such that

$$x^{(j_0)} < x^{(j_1)} < \dots < x^{(j_{N-1})}.$$

Then the result of the BWT is a string consisting of the last symbols from each of the sorted cyclic rotations:

$$x_{N-1}^{(j_0)}, x_{N-1}^{(j_1)}, \dots, x_{N-1}^{(j_{N-1})}.$$

In order to facilitate the inversion transformation, an additional pointer is used to remember the position of the original string $x = x^{(0)}$, i.e. t such that $j_t = 0$. Since the inversion transformation is not important for our exposition, we will not describe it here.

Example 1: Let $\mathcal{A} = \{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}\}$ and $N = 10$. Let $x =$ `cdababcdaa` be an input block. Applying BWT:

sorted cyclic rotations:

```
aacdababcd
ababcdaacd
abcdaacdab
acdababcda
babcdaacda
bcdaacdaba
cdaacdabab
cdababcdaa
daacdababc
dababcdaac
```

$\Longrightarrow$ result of BWT: `ddbaaabacc`, $t = 7$

The paper [6] uses a variant of the BWT with special symbol denoting the end of block. In that case, there is no need to remember the position of the original block among sorted cyclic rotations. The analysis done in Section III is valid for this variant as well.

B. MTF

The MTF transforms an input string $x = x_0x_1 \dots x_{N-1} \in \mathcal{A}^N$ into sequence of N numbers $\{p_i\}_{i=0}^{N-1}$, where $p_i \in \{0, 1, \dots, L-1\}$. The algorithm uses a table $T[0, \dots, L-1]$ of all symbols from $\mathcal{A}$, initially sorted in lexicographic order.

For each $i = 0, 1, \ldots, N-1$ the symbol x_i is processed as follows:

1) p_i is the position of x_i in table T;
2) T is modified: x_i is moved to the front/top of the table.

It is easy to see that the MTF is invertible. The main idea of MTF is that the recently used symbols are encoded as small integers. This makes its output a suitable data for subsequent compression by simple entropy coders such as Huffman or arithmetic coding.

Example 2: Let $x =$ `ddbaaabacc` be an input string for MTF. The output 3022001130 is computed as follows:

i		0	1	2	3	4	5	6	7	8	9
x_i		d	d	b	a	a	a	b	a	c	c
p_i		3	0	2	2	0	0	1	1	3	0
T	a	d	d	b	a	a	a	b	a	c	c
	b	a	a	d	b	b	b	a	b	b	b
	c	b	b	a	d	d	d	d	d	a	a
	d	c	c	c	c	c	c	c	c	d	d

C. Scrambled BWT with MTF

Oğuzhan Külekci [6] proposed scrambled BWT, where a secret (encryption) key is a secret lexicographic order of symbols in $\mathcal{A}$. One of the claimed advantages is a large key space, for $L = 256$ it is 256! keys. The author observed, that using scrambled BWT is not secure enough, and can be attacked by exploiting known statistical relationships among plaintext symbols (e.g. digram frequencies). Therefore he accompanied the scrambled BWT (sBWT) with MTF, where the secret lexicographic order is applied as well (i.e. the initialization of T depends on this order):

> "...Thus, initializing the alphabet ordering in MTF with the secret lexicographic order used in sBWT provides protection against that statistical attack."

Example 3: Let $x =$ `cdababcdaa` be an input block. We show the result of sBWT and subsequent MTF with various lexicographic orders:

key k	sBWT with k		MTF with k
c < a < d < b	baadbdccaa	($t = 1$)	3203213030
d < c < b < a	ccabaaabdd	($t = 2$)	1033100130
a < c < b < d	dadbbaaacc	($t = 5$)	3113020030

III. SECURITY ANALYSIS

Let us denote the secret lexicographic order as k and the corresponding scrambled BWT as sBWT_k. Similarly, the MTF with secret lexicographic order is denoted as MTF_k. Let x be an input block. The author proposes [6] the same secret key (secret lexicographic order) for both transformations: $\mathrm{MTF}_k(\mathrm{sBWT}_k(x))$.

Let us note that attacks described in this sections will work even for situation with two independent secret keys: $\mathrm{MTF}_{k_2}(\mathrm{sBWT}_{k_1}(x))$. The attack will "undo" the MTF (revealing the value of $\mathrm{sBWT}_{k_1}(x)$), and the sBWT part can be attacked by exploiting the statistical properties of plaintext as suggested in [6].

Remark 1: We can ignore other pre- and post-processing steps in the compression algorithm, since they do not depend on the key.

We base our analysis on the following observations:

1) The scrambled BWT, with secret lexicographic order as a key, keeps the frequencies of symbols intact, i.e. when symbol 'a' appears l times in an input block, then 'a' will appear exactly l times in the output block. One can view sBWT_k as a plaintext-dependent permutation cipher.
2) For any string z and any two lexicographic orders k_1, k_2, performing $y = \mathrm{MTF}_{k_2}^{-1}(\mathrm{MTF}_{k_1}(z))$ can be viewed as a monoalphabetic substitution, i.e. it change the symbols but it does not change the frequencies. For example, 'a' can become 'w', 'b' can become 'f' ... but in such case the number of a's in z is the same as the number of w's in y, the number of b's in z is the same as the number of f's in y etc.

For the rest of the section we denote an input block x and the resulting block $y = \mathrm{MTF}_k(\mathrm{sBWT}_k(x))$. Our analysis is done primarily with single data block (it is sufficient for most scenarios). However, it can be extended to multiple blocks in a straightforward manner, see Section III-C.

A. Chosen plaintext attack

The goal of the attack is to identify a secret key (lexicographic order):

1) Construct input block x, where symbols from $\mathcal{A}$ have unique frequencies.
2) Compute $z^* = \mathrm{MTF}_{k'}^{-1}(y)$ for an arbitrary lexicographic order k'.
3) Pair symbols in x and z^* according their frequencies, and recover the correct "middle" value $z = \mathrm{sBWT}_k(x) = \mathrm{MTF}_k^{-1}(y)$.
4) Reconstruct the key from z and y.

Let $\mathcal{A} = \{a_1, \ldots, a_L\}$. The chosen input block can be, for example, $a_1 a_2 a_2 a_3 a_3 a_3 \ldots$ finishing with L symbols a_L. Then we can reconstruct the key from this single block as long as the block size is at least $\binom{L+1}{2}$ (for $L = 256$ it is 32 896). The complexity of the attack is $\Theta(L^2)$ (counting all its steps).

Example 4: Let us illustrate the attack with the following toy example. Let $\mathcal{A} = \{$a, b, c, d$\}$ and $N = 10$. We choose the input block $x =$ `abbcccdddd`, and observe the output $y =$ `3133001022`. We apply inverse MTF with natural lexicographic order (a < b < c < d): $z^* =$ `dacbbbccab`. Pairing symbols with equal frequencies yields $z =$ `abcdddccbd`. Knowing "middle" value z and the result of $\mathrm{MTF}_k(z)$, i.e. y, we can easily reconstruct the secret lexicographic order: b < d < c < a.

B. Known plaintext attack (single block)

Known plaintext attack extends the previous attack assuming that the attacker cannot control the frequencies of particular symbols in the input data. However, considering the block sizes used in practice (e.g. default 900kB block

size in bzip2), the probability of equal frequencies of symbols is rather low. Moreover, longer blocks usually lead to better compression.

Remark 2: For short blocks, one can assume that the attacker will know the input data for multiple blocks. Therefore he can combine results from these blocks and significantly reduce the complexity of the attack further, see Section III-C.

Let us denote by $\#_v(x)$ the number of symbols $v \in \mathcal{A}$ in the string x. Let $C(x) = \{\#_v(x) \mid v \in \mathcal{A}\}$ be the set of all distinct values of $\#_v(x)$. For each value $r \in C(x)$ we define $\text{size}(x, r)$ to be the number of symbols having exactly r occurrences in x, i.e. $\text{size}(x, r) = |\{v \in \mathcal{A} \mid \#_v(x) = r\}|$.

The attacker proceeds similarly to the chosen plaintext attack, computing $z^* = \text{MTF}_{k'}^{-1}(y)$ for an arbitrary lexicographic order k'. Then he tries to pair symbols in x and z^* according their frequencies to recover the correct "middle" value $z = \text{sBWT}_k(x) = \text{MTF}_k^{-1}(y)$. However, this time there in no guarantee of unique frequencies, therefore the attacker can perform an exhaustive search for all assignments of symbols in groups with equal frequencies. For each assignment, the attacker computes a corresponding lexicographic order and performs an inverse BWT with this order. Comparing the result with the original plaintext gives a confirmation/rejection of particular assignment. The size of the search space is

$$\text{SS}(x) = \prod_{r \in C(x)} \text{size}(x, r)!$$

We measure the complexity of our known plaintext attack as bit security of the cipher, i.e. binary logarithm of the corresponding search space size: $\log_2 \text{SS}(x)$. In order to estimate the bit security, we performed the following experiments:

1) *RandBytes.* We generate the input block as a stream of randomly and independently generated bytes (i.e. $L = 256$) with uniform distribution. Since in real-world one can expect much more "compressible" input block (with greater variation of symbols frequencies), our model simulates the worst situation for the attacker. Therefore, the estimations obtained in this experiment can be viewed as upper bounds for bit security of the cipher.
2) *RandReduced.* This is a similar experiment as the previous one, but this time we restrict possible symbols to the set of $L = 100$ symbols (generated with uniform distribution). The rest of the symbols are not generated.
3) *RandText.* This and the last experiment (RandKernel) are probably the most realistic of our experiments. In RandText we model the input block as a stream of independently generated bytes, where the probabilities of individual symbols correspond to the probabilities of symbols in a novel *Crime and Punishment* [4].
4) *RandKernel.* Similar experiment to RandText. However, this time the probabilities of individual symbols (bytes) correspond to the probabilities of symbols in `kernel32.dll` file in Windows 7.

We gradually increased N in each experiment and computed the average bit security of the cipher. The average value was computed from 1000 samples. It is interesting to see the level of bit security degradation from the theoretical level: $\log_2(256!) \sim 1684$ bits. The values for RandReduced, RandText and RandKernel experiments show that the scrambled BWT with MTF offers practically no security. The results are shown in Table I.

TABLE I
BIT SECURITY FOR SINGLE BLOCK KPA (WITH BLOCK SIZE N)

N	RandBytes	RandReduced	RandText	RandKernel
50 000	378.7	28.6	10.1	65.3
100 000	304.2	20.9	8.4	49.0
150 000	263.7	17.2	8.8	41.4
200 000	236.4	15.2	9.2	36.5

C. Known plaintext attack (multiple blocks)

The known plaintext attack from previous section can be easily extended for multiple blocks, with improved performance. The majority of "ties" (groups of symbols with equal frequencies) in one block can be broken by considering the frequencies in other blocks. In order to illustrate this effect, we performed experiments similar to those in the previous section (RandBytes, RandReduced, RandText, and RandKernel). However, this time we used two blocks of known plaintext and 10 times smaller block sizes. The results are shown in Table II (again, 1000 samples were used to estimate individual values). Even with such artificially small block sizes one can observe further decrease in the bit security of the scrambled BWT with MTF.

TABLE II
BIT SECURITY FOR TWO BLOCKS KPA (WITH BLOCK SIZE N)

N	RandBytes	RandReduced	RandText	RandKernel
5 000	113.3	4.0	8.4	44.6
10 000	61.0	1.9	5.0	21.4
15 000	41.8	1.3	3.6	13.9
20 000	31.6	1.0	3.2	10.4

IV. CONCLUSION

Usually, providing privacy (encryption) by modifying the data compression techniques is not a good idea in practice. We demonstrated the security problems of the scrambled BWT with MTF. Moreover, it seems that these problems cannot be easily fixed. Further extensions of our attacks can be aimed at known plaintext attack with only partial block knowledge, or even ciphertext only attacks. However, the already identified weaknesses of the cipher allow us to conclude that the cipher is insecure.

ACKNOWLEDGMENT

This work was supported by P12/1.

Martin Stanek is an Associate Professor in the Department of Computer Science, Comenius University. He received his PhD. in computer science from Comenius University. His research interests include cryptography and information security.

REFERENCES

[1] Eli Biham and Paul C. Kocher. A Known Plaintext Attack on the PKZIP Stream Cipher. *Fast Software Encryption*, FSE'94, Lecture Notes in Computer Science vol. 1008, Springer, pp. 144–153, 1995.
[2] Michael Burrows and David J. Wheeler A block-sorting lossless data compression algorithm. Technical Report 124, Digital Equipment Corporation, 1994.
[3] John G. Cleary, Sean A. Irvine and Ingrid Rinsma-Melchert. On the insecurity of arithmetic coding. *Computers & Security*, 14(2):167–180, 1995.
[4] Fyodor M. Dostoevsky. *Crime and Punishment*. Project Gutenberg, vol. 2556, English translation by C. Garnett, 2006. http://www.gutenberg.org/ebooks/2554
[5] Tadayoshi Kohno. Attacking and repairing the winZip encryption scheme, *ACM Conference on Computer and Communications Security*, CCS'04, pp. 72–81, 2004.
[6] M. Oğuzhan Külekci. On scrambling the Burrows-Wheeler transform to provide privacy in lossless compression, *Computers & Security*, 31(1):26–32, 2012.
[7] Jen Lim, Colin Boyd and Ed Dawson. Cryptanalysis of Adaptive Arithmetic Coding Encryption Schemes. *Information Security and Privacy, Second Australasian Conference*, ACISP'97, Lecture Notes in Computer Science vol. 1270, Springer, pp. 216–227, 1997.
[8] Julian Seward. bzip2, http://www.bzip.org, 1996-2012.
[9] Jiantao Zhou, Oscar C. Au and Peter Hon-Wah Wong. Adaptive chosen-ciphertext attack on secure arithmetic coding. *IEEE Transactions on Signal Processing*, 57(5):1825–1838, 2009.

Two Improvements of Random Key Predistribution for Wireless Sensor Networks – Revised Version

Jiří Kůr, Vashek Matyáš, Petr Švenda

***Abstract*—Key distribution is of a critical importance to security of wireless sensor networks (WSNs). Random key predistribution is an acknowledged approach to the key distribution problem. In this paper, we propose and analyze two novel improvements that enhance security provided by the random key predistribution schemes. The first improvement exploits limited length collisions in secure hash functions to increase the probability of two nodes sharing a key. The second improvement introduces hash chains into the key pool construction to directly increase the resilience against a node capture attack. Both improvements can be further combined to bring the best performance. We evaluate the improvements both analytically and computationally on a network simulator. The concepts used are not limited to the random key predistribution.**

***Index Terms*—Hash function collision, key management, random key predistribution, security, wireless sensor network.**

I. Introduction

A WIRELESS sensor network (WSN) consists of resource-constrained and wireless devices called sensor nodes. WSNs monitor some physical phenomenon (e.g., vibrations, temperature, pressure, light) and send measurements to a base station. There are several classes of sensor nodes available – ranging from high-end nodes that can easily employ public-key cryptography down to nodes that can barely make use of any cryptography at all. In our work, we consider cheap and highly constrained nodes that can use only symmetric cryptography and their storage is just a few kilobytes.

Key distribution is one of the greatest challenges in WSNs. Since network topology is usually not a priori known, every node should be able to establish a link key with a large portion of other nodes to ensure the connectivity of the network. To achieve this requirement, nodes may pre-share a single network-wide master key and use it to establish link keys. However, if a single node with the master key is captured, then the whole network gets compromised. In an alternative approach, each node pre-shares a unique link key with every node. This offers much better security, yet hits the memory limits as number of nodes in the network increases.

A suitable trade-off between the two approaches comes with the random key predistribution [1]. Every node is preloaded with a fixed number of keys randomly selected from a given key pool. After the network deployment, two nodes establish a link key if they share at least one key from the key pool. The scheme can be extended to require at least q shared keys [2].

In this paper, we propose two improvements of the basic random key predistribution schemes. In the first improvement, we increase a probability that any two nodes establish a link key while maintaining memory requirements fixed. For this purpose we construct the key pool using limited length (e.g., 80-bit) collisions in a secure hash function. We also provide an evidence that such collisions can be found in a reasonably short time on today's personal computers.

The second improvement introduces hash chains into the key pool structure to directly enhance the resilience against a node capture attack. Both the improvements can be further combined together to bring the best performance. These improvements are particularly advantageous for situations in which the attacker manages to capture a significant number of nodes.

The structure of the paper is following – we review the basic random key predistribution schemes and other related work in Section II. We present and evaluate the first improvement in Section III and the second improvement in Section IV. Then we evaluate their combination in the following section of this paper. We provide computational results from a network simulator and proof of concept for the collision search in Section VI, and then the last section concludes the paper. Proofs of selected equations can be found in the Appendix.

II. Background And Related Work

In this section, we provide a background knowledge on the basic random key predistribution schemes and other related work.

Our proposals modify the basic random key predistribution schemes proposed by Eschenauer and Gligor [1] and Chen et al. [2]. We refer to the schemes as to the *EG scheme* and the *q-composite scheme*, respectively, in our paper.

The *EG scheme* works as follows: in the *(i) key setup* phase, a key pool S is created by randomly taking $|S|$ keys from the possible key space. Then, for every node, m keys are randomly drawn from the key pool S without replacement and uploaded into the node. These keys form a key ring for the given node. If $|S|$ and m are chosen properly, any two nodes in the network share at least one key with a high probability. E.g., for a key pool size $|S| = 10,000$ and a ring size $m = 83$ the probability that any two nodes share at least one key is approximately $p = 0.5$.

This is a revised version of our paper from the 8th International Conference on Security and Privacy in Communication Networks that won the SantaCrypt 2012 student KEYMAKER competition, the first author submitted the paper as a PhD student.

J. Kůr, V. Matyáš and Petr Švenda are with the Faculty of Informatics, Masaryk University, Brno, Czech Republic (e-mail: xkur,matyas,svenda@fi.muni.cz).

V. Matyáš undertook final work on this paper as a Fulbright-Masaryk Visiting Scholar at Harvard University.

In the *(ii) shared key discovery* phase, every two neighboring nodes try to identify shared keys among their key rings. This can be done by various methods. E.g., every key can be assigned a short unique identifier that is broadcasted by the nodes that have the corresponding key in their key rings. If a shared key is found, it is used as a link key for the communication between the two nodes. If not, the link key can be established in the *path-key establishment* phase.

The *(iii) path-key establishment* phase is optional. It uses already secured links to establish link keys between two neighboring nodes that could not establish a link key directly as they had no shared key or their keys were compromised [3].

The *q-composite scheme* is a generalization of the EG scheme. In the *shared key discovery* phase, two nodes establish a link key only if they share at least q keys in their key rings. The resulting link key is derived from all the shared keys.

A. Node Capture Resilience

The performance of random key predistribution schemes is evaluated with respect to the node capture resilience [2]. It can be defined as the probability that a given secured link between two uncaptured nodes can be compromised by an attacker using keys extracted from already captured nodes. In other words, the node capture resilience is a fraction of secured links between uncaptured nodes that can be compromised by an attacker.

The node capture resilience is mostly influenced by the following three factors – the ring size m, the key pool size $|S|$ and the probability that any two nodes in the network can establish a link key. These values are to some extent determined by properties of the network concerned. The ring size m is limited by a storage capacity of the network sensor nodes. If we want the network to be connected by secure links, the minimum required probability of a link key establishment is given by the size of the network and by the average number of neighboring nodes (for details see [1]). Given the m and the minimum required probability, the $|S|$ is uniquely determined. Note that in the q-composite scheme also the q influences the node capture resilience and the key pool size $|S|$.

B. Other Related Work

The basic schemes have been modified by Ren et al. [4]. The key pool in their scheme consists of a large number of keyed hash chains where every hash chain element is considered a unique key. Every node is then assigned a number of such keys and a number of whole keyed hash chains represented by their hashing keys and chain starting points. Deterministic and hybrid approaches how to select keys for key rings based on *combinatorial design* are proposed in [5]. These approaches enhance the performance of the basic schemes and similarly to them can be also further improved with our proposals. For other key distribution schemes in WSNs see the survey [6].

Our first proposal is based on hash collisions. Rivest and Shamir took an advantage of hash collisions for a security gain in the MicroMint micro payment scheme, where an electronic coin was represented by a hash collision [7]. However, their scheme relies on the security economics rather than on computational complexity per se. An attacker with a computational power equal to the broker is able to cheat by finding a collision with given properties. In our scheme, an attacker needs to find a pre-image for a given hash.

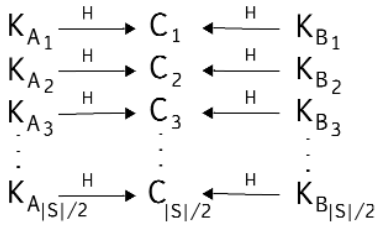


Fig. 1. Key pool structure in the collision key improvement. Colliding keys from the key pool are denoted K_A and K_B. Collision keys are depicted as C. H denotes a secure hash function.

As far as we are aware, the first usage of hash-chains for key agreement appeared in [8].

III. Collision Key Improvement

We propose a modification to the basic random key predistribution schemes. Keeping the key ring size m and the key pool size $|S|$ same as for the basic schemes, this modification additionally increases the number of keys that two nodes may share.

In our scheme, two nodes X and Y can share a key directly as in the basic schemes. Furthermore, an additional shared key C can be constructed if two nodes carry two different, but related keys K_A and K_B such that the condition $C = H(K_A) = H(K_B)$, where H is a suitable cryptographic hash function, is fulfilled. We call such related keys *colliding keys* and the resulting value C a *collision key*. Probability of two randomly chosen values for K_A and K_B being colliding keys is generally very low due to the collision resistance of the hash function. Therefore, we modify the process how keys for the key pool are selected. Instead of randomly selecting $|S|$ keys from the possible key space, $\frac{|S|}{2}$ colliding key pairs are taken to form a key pool S. Thus, the total number of keys in the key pool remains $|S|$ and the key pool gets the structure depicted in Fig. 1.

Colliding keys long enough to withstand a brute-force attack can be efficiently generated due to the birthday paradox. In Section VI, we demonstrate that for key length of $N = 80$ bits thousands of colliding key pairs can be generated with a moderate computational power.

Beside the key pool structure, we also slightly modify the way how keys are selected to a key ring. The keys are still picked from the key pool randomly without replacement, however, we do not allow two colliding keys to be in the same key ring. Thus, if a key is picked, not only itself but also its colliding counterpart is temporarily marked off the key pool. Once the key ring is complete, all keys are put back to the key pool and the next node is processed.

In the shared key discovery phase, similarly to the q-composite scheme, two nodes X and Y can establish a link key if they share at least q keys. The shared keys can be both colliding keys drawn directly from the key rings or collision

keys computed using a hash function H. Since every node has m keys in its key ring, it is also able to establish m collision keys. Thus our improvement significantly increases the effective size of the key rings as evaluated in the following subsection. Therefore, we can expand the key pool accordingly while keeping the probability of a link key establishment at the desired level. This expansion increases the node capture resilience. In the rest of the paper we refer to this improvement as to the *collision key improvement*.

A. Probability of Link Key Establishment

In this subsection we show how to calculate the probability that any two nodes in the network are able to directly establish a link key in the shared key discovery phase. Let us define the following notation to support readability of the subsequent analysis.

Definition 1:

$$\left\{ \begin{array}{c} |S| \\ m \end{array} \right\} = |S| \cdot (|S|-2) \cdot (|S|-4) \cdot \ldots \cdot (|S|-2\cdot(m-2)) \cdot (|S|-2\cdot(m-1)) \cdot \frac{1}{m!}$$

The formula expresses the number of all possible key rings of size m selected from a key pool of size $|S|$ where no colliding key pair is present in the key rings. Thus it can be viewed as $|S|$ choose m, where the choice has to respect the above mentioned constraint. For justification see the Appendix.

The probability that any two nodes in the network share exactly i keys from the key pool S and exactly j collision keys that do not result from the i shared keys can be calculated as follows (see the Appendix for proof):

$$P_{SharedExactly}(i,j) = \frac{\left(\begin{array}{c} m \\ i \end{array} \right) \left(\begin{array}{c} m-i \\ j \end{array} \right) \left\{ \begin{array}{c} |S|-2m \\ m-i-j \end{array} \right\}}{\left\{ \begin{array}{c} |S| \\ m \end{array} \right\}} \tag{1}$$

Two nodes can establish a link key if they share at least q independent keys, no matter whether these are colliding or collision keys. The collision keys are counted only if their pre-images are not. Thus the probability that any two nodes in a network are able to establish a link key is, among m and $|S|$, dependent also on the parameter q and can be calculated as:

$$P_{LinkEstablishI} = \sum_{i=0}^{m} \sum_{j=0}^{m} P_{SharedExactly}(i,j) \tag{2}$$

where $i+j \geq q$ and $i+j \leq m$.

B. Resulting Node Capture Resilience

In this subsection we evaluate the collision key improvement with respect to the node capture resilience. The resilience is dependent on the number of captured nodes x, the key pool size $|S|$, the key ring size m and the desired probability of a link key establishment $P_{LinkEstablish1}$. We assume that an attacker has selected the nodes to capture in a random fashion and calculate the node capture resilience as

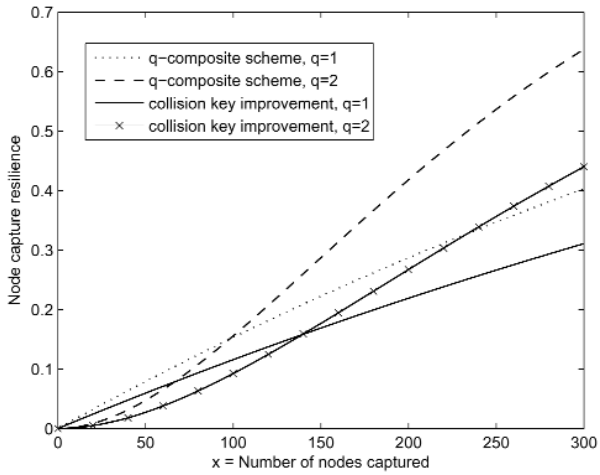


Fig. 2. Node capture resilience after x randomly selected nodes have been captured, key ring size $m = 200$, probability of link key establishment $P_{LinkEstablishI} = 0.33$.

$$P_{LinkComprI} = \sum_{i=0}^{m} \sum_{j=0}^{m} (1-(1-\frac{m}{|S|})^x)^i (1-(1-\frac{2m}{|S|})^x)^j \cdot \frac{P_{SharedExactly}(i,j)}{P_{LinkEstablishI}} \tag{3}$$

where $i+j \geq q$, $i+j \leq m$. For proof see the Appendix.

Fig. 2 presents a comparison of the q-composite scheme and the collision key improvement. It is clear that our improvement provides a better node capture resilience for both values of q. E.g., if $q = 2$ and 50 nodes are captured, the resilience of the q-composite scheme is 4.7%, whereas the collision key improvement gives us the resilience of 2.7%.

IV. KEY-CHAIN IMPROVEMENT

The second proposed modification of the basic random key predistribution introduces hash chains into the key pool construction. We will refer to this modification as to the *key-chain improvement*. The key-chain improvement was loosely inspired by previous work of Ren et al. [4], but our scheme utilizes hash chains in a different manner. Furthermore, we employ basic hash chains instead of the keyed ones.

In our scheme, the key pool consists of $|S|$ non-colliding hash chains of a length L and every value in the chains is considered as a potential key. Thus, we refer to the hash chains as to the key-chains. The structure of the key pool is depicted in the Fig. 3.

In the key-setup phase, every node is randomly assigned a key from m randomly selected key-chains. If two nodes were assigned keys from the same key-chain, they are able to calculate a shared key. A node with a value closer to the beginning of the key-chain can traverse the chain downwards to find the shared key carried by the second node.

In the shared key discovery phase, two nodes can establish a link key when sharing at least q independent keys.

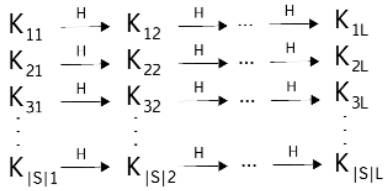


Fig. 3. Key pool structure in the key-chain improvement. The knowledge of a key K_{ij} enables one to compute keys K_{ik} for every $k \geq j$.

The actual size of the key pool is $|S| \cdot L$, although in the subsequent analysis we shall consider the number of key-chains $|S|$ as the key pool size. The length of a key-chain L shall be taken as an independent parameter that influences the scheme security. The key-chain improvement is a generalization of the basic q-composite scheme for which the key-chain length $L = 1$. The key-chain improvement can be further combined with the collision key improvement to get even better performance. The combination is considered in Section V.

A. Probability of Link Key Establishment

The probability that any two nodes share exactly i independent keys is equal to the same probability for the basic EG and q-composite schemes. To calculate the probability we can use the formula from [2]:

$$P_{SharedExactly}(i) = \frac{\binom{|S|}{i}\binom{|S|-i}{2(m-i)}\binom{2(m-i)}{m-i}}{\binom{|S|}{m}^2} \qquad (4)$$

Note that if the key-chains are non-colliding the probability is independent of their length L as the length influences only the node capture resilience provided by the scheme. The probability of a link key establishment for a given q is then

$$P_{LinkEstablish11} = \sum_{i=q}^{m} P_{SharedExactly}(i) \qquad (5)$$

B. Resulting Node Capture Resilience

The node capture resilience is in this case dependent (among other parameters) also on the key-chain length L. To evaluate the node capture resilience, we first calculate the probability that a key from a given key-chain is compromised after an attacker captured x random nodes as follows (see the Appendix for proof):

$$P_{ChainCompr} = \sum_{i=1}^{L} \frac{2 \cdot i - 1}{L^2}\left(1 - \left(1 - \frac{m}{|S|}\frac{i}{L}\right)^x\right) \qquad (6)$$

Assuming an attacker has selected the nodes to capture in a random fashion, the node capture resilience can be calculated as

$$P_{LinkComprII} = \sum_{i=q}^{m} (P_{ChainCompr})^i \frac{P_{SharedExactly}(i)}{P_{LinkEstablishII}} \qquad (7)$$

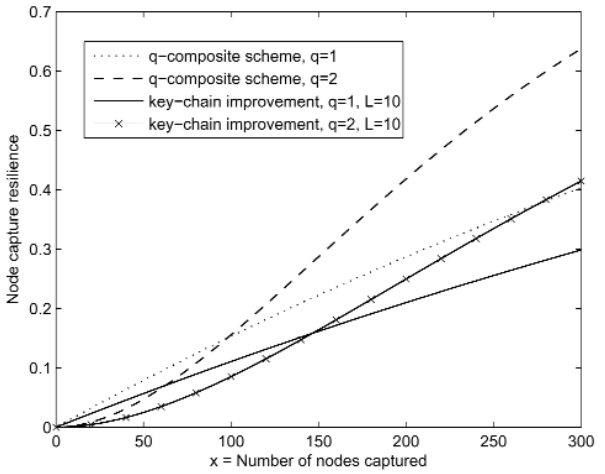


Fig. 4. Node capture resilience after x randomly selected nodes have been captured, key ring size $m = 200$, probability of link key establishment $P_{LinkEstablishII} = 0.33$, effective key-chain length $L = 10$.

Fig. 4 presents a comparison of the q-composite scheme and the key-chain improvement. Again, our improvement provides a better node capture resilience for both values of q. E.g., if $q = 2$ and 50 nodes are captured, the resilience of the q-composite scheme is 4.7%, whereas the key-chain improvement provides the resilience of 2.5%. Such a resilience is even better than the resilience provided by the collision key improvement proposed in Section III.

C. Key-Chain Length

An important security parameter of the key-chain improvement is the length of the key-chains. It holds that the longer the key-chain, the better the node capture resilience. However, as the length of the key-chain increases, the security gain obtained for a single unit increment decreases rapidly as demonstrated in Fig. 5. Also, when evaluating the node capture resilience, we have to consider the *effective length* of the key-chain, not the actual one. The effective length of a key-chain is the number of different keys from the key-chain that are actually assigned to some key ring. The effective length is dependent on the number of nodes in the network n, the size of a key ring m and the size of the key pool $|S|$. The average effective key-chain length cannot exceed $\frac{n \cdot m}{|S|}$, which is the expected number of nodes that will be assigned a key from a given key-chain. If we set the actual length to be equal to this number, the average effective key-chain length will be shorter. We can get close to the bound by setting the actual length artificially long. Yet this would increase the computational complexity of the key establishment as nodes would need to perform more hashing to establish a shared key. In practice, it is not necessary to reach the maximum length available due to the steep decrease in additional gain demonstrated in Fig. 5.

For most networks, a practical value of the effective key-chain length would be around $L = 10$. We base our recommendation on the shape of the curve in Fig. 5. Furthermore, such

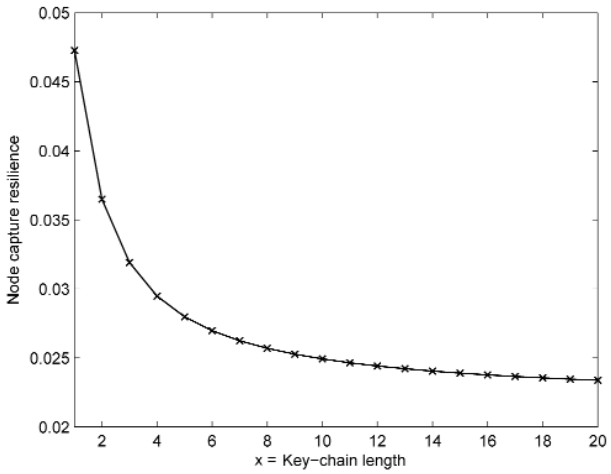


Fig. 5. Relationship between an effective length of a key-chain and node capture resilience. Key ring size $m = 200$, probability of link key establishment $P_{LinkEstablishII} = 0.33$, $q = 2$ and the number of captured nodes $x = 50$.

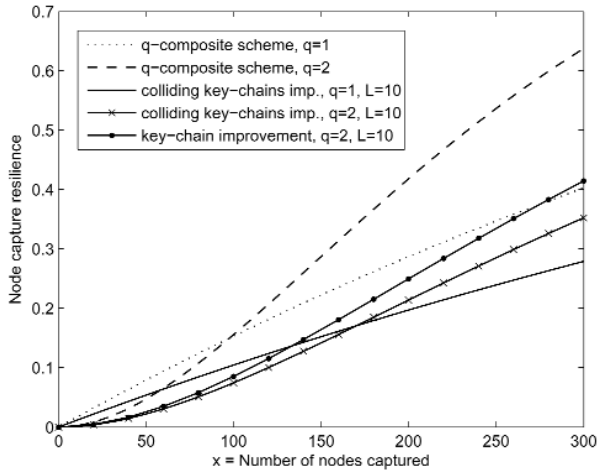


Fig. 6. Node capture resilience of the colliding key-chains improvement. Key ring size $m = 200$, probability of link key establishment $P_{LinkEstablishI} = 0.33$, effective key-chain length $L = 10$.

an effective length is achievable with only a slightly higher actual key-chain length for sufficiently large networks. E.g., for $n = 10,000$, $P_{LinkEstablishII} = 0.33$, $q = 2$, $m = 200$ and the actual key-chain length of 10, the minimum effective key-chain length is 8 and the average one is 9.97, as calculated by our network simulator.

V. Colliding key-chains improvement

The key-chain improvement can be directly combined with the collision key improvement into the *colliding key-chains improvement*. This combination achieves even better results with respect to the node capture resilience. To obtain the colliding key-chains, the end points of the key chains should be the colliding keys. This requirement can be easily fulfilled due to the nature of the parallel collision search algorithm. If the collision is found, also the two hash chains that precede it are obtained, see Section VI.

The *probabilities of a link key establishment* between any two nodes in the network are calculated similarly as in the case of the collision key improvement using Equations 1 and 2. The size of the key pool $|S|$ is in this case defined as the number of key-chains. Thus $|S|$ has a similar meaning as in the key-chain improvement.

For given arguments, one obtains the same probability of a link key establishment for both the collision key improvement and for the colliding key-chains improvement. Yet there is a difference in the achieved node capture resilience, which is higher for the combined solution. The difference is dependent on the effective length L of the key-chains.

The *node capture resilience* of the colliding key-chains improvement can be calculated as

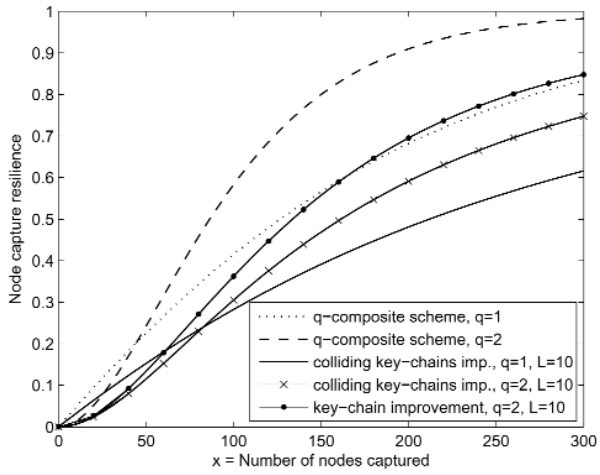


Fig. 7. Node capture resilience of the colliding key-chains improvement. Key ring size $m = 100$, probability of link key establishment $P_{LinkEstablishI} = 0.5$, effective key-chain length $L = 10$.

$$P_{LinkComprIII} = \sum_{i=0}^{m}\sum_{j=0}^{m}(P_{ChainCompr})^i(1-(1-\frac{2m}{|S|})^x)^j \cdot \frac{P_{SharedExactly}(i,j)}{P_{LinkEstablishI}} \tag{8}$$

where $i + j \geq q$, $i + j \leq m$.

Fig. 6 demonstrates that the colliding key-chains improvement outperforms the q-composite scheme and the key-chain improvement. E.g., if $q = 2$ and 50 nodes are captured, q-composite scheme scores 4.7%, collision key improvement 2.7%, key-chain improvement 2.5% and the colliding key-chains improvement 2.2%. The comparison gets even better for the colliding key-chains improvement as the number of

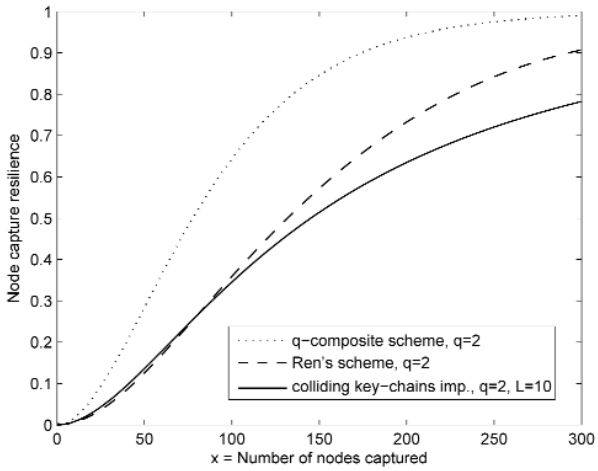


Fig. 8. Comparison of node capture resilience of the q-composite scheme, the Ren's scheme and the colliding key-chains improvement. Key ring size $m = 90$, probability of link key establishment $P_{LinkEstablishI} = 0.5$, effective key-chain length $L = 10$.

captured nodes grows. Fig. 7 shows the performance of the colliding key-chains improvement for values of key ring size $m = 100$ and probability of link key establishment $P_{LinkEstablishI} = 0.5$.

The scheme of Ren et al. [4] provides a slightly better node capture resilience than our colliding key-chains improvement for a small number of captured nodes. However, as this number grows the colliding key-chains improvement starts to outperform the Ren's scheme. The comparison is depicted in Fig. 8. For $P_{LinkEstablishI} = 0.5$, $q = 2$ and $m = 90$, the turning point is around 80 of captured nodes. Since we were not able to fully reproduce Ren's analytical results (and did not get any response from the contacted authors), we did the comparison only for the parameters used in their paper. The curve showing the performance of the Ren's scheme in Fig. 8 was taken from Fig. 9 in [4].

The communication overhead of the shared-key discovery phase, when the colliding key-chains improvement is used, is dependent on the discovery procedure itself. For some procedures it is similar to the overhead of the basic random key predistribution schemes. E.g., if the pseudo-random predistribution [9] is used, identifiers of keys assigned to a particular node can be calculated from the node ID. These identifiers can carry all the information necessary to discover shared keys – the key's position in a hash-chain and the hash-chain identifier. Additionally, the shared collision key can be figured out through the hash-chain identifiers when these identifiers (assigned to the colliding hash-chains) differ only in the least significant bit. Thus the communication overhead only covers transmission of the node IDs. Another advantage of the pseudo-random approach is that the nodes do not need to store their own key identifiers as these can be computed when actually needed.

Another interesting information concerns the composition of link keys established, e.g., what fraction of keys is based

TABLE I

PROBABILITIES THAT TWO NODES SHARE EXACTLY i KEYS FROM HASH CHAINS AND j ADDITIONAL COLLISION KEYS. VALUE i IS DEPENDENT ON THE ROW AND VALUE j ON THE COLUMN OF THE TABLE. $P_{LinkEstablishI} = 0.33$, $q = 1$, $m = 200$.

$i \backslash j$	0	1	2	3
0	0.67	0.1344	0.0134	0.0009
1	0.1344	0.0267	0.0026	0.0002
2	0.0134	0.0026	0.0003	0
3	0.0009	0.0002	0	0

solely on the collision keys or solely on the keys from the hash chains. Such information can be calculated using Equation 1. The equation gives us the probability that two nodes share exactly i keys from hash chains and j additional collision keys. The sample probabilities for $P_{LinkEstablishI} = 0.33$, $q = 1$, $m = 200$ and different combinations of i and j are summarized in Table I. E.g., the probability that a link key is based on exactly two keys from hash chains and a single collision key is given in the row 2, column 1. The probability that two nodes do not share any key is in the row 0, column 0. Note that the table is symmetric, i.e., both types of keys are used with an equal probability. The table is not complete, yet the probabilities of other combinations of i and j are negligible.

VI. COMPUTATIONAL RESULTS

Analytical results presented in the previous sections were computationally verified using our network simulator. We have simulated the q-composite scheme and all the proposed improvements using various settings for critical parameters and networks of different sizes and topologies. For every setting, an average over 10 different simulation runs was taken as a result. The reference simulated network had 10,000 nodes, though we have tested also other sizes. It showed that obtained analytical and simulated results for node capture resilience are consistent. The simulator and its source codes are available for download along with sample configuration scripts that enable the verification of results [10].

The important part of the collision key and the colliding key-chains improvement is a search for collisions of the cryptographic hash function. This search can be efficiently performed due to the birthday paradox. In order to find an N-bit collision in a cryptographic hash function, one needs to perform approximately $2^{\frac{N}{2}}$ hashing operations. Furthermore, to find c^2 such collisions for $1 \leq c \leq 2^{\frac{N}{2}}$, one needs to perform "only" approximately $c \cdot 2^{\frac{N}{2}}$ hashing operations [7]. Thus, once the first collision is found, additional collisions can be found increasingly efficiently. If we assume 80-bit keys are used, to create the key pool for $|S| = 2^{16}$ one needs to find 2^{15} collisions which requires approximately $2^{47.5}$ hashing operations. This can be reached with a moderate computational power. Note that 80-bit keys can be still considered as secure and appropriate for use in wireless sensor networks as attacker needs to try approximately 2^{79} values to brute force the key.

We have conducted our collision search using the parallel collision search method proposed by van Oorschot and Wiener [11]. This method is based on Hellman's time-memory

trade-off approach and calculates long hash-chains. We have searched for 80-bit collisions of the SHA-256 hash function, 80-bit values were taken as an input and 80 most significant bits of the SHA-256 function as an output. We used the Gladman's implementation [12] of the hash function. The aggregate time to find over 5,000 suitable collisions was approximately 19,000 hours on a single 3GHz CPU core. The search was distributed using the BOINC infrastructure [13] to around 1,000 CPU cores so the search took less than a day. Approximately 2^{23} hash chains with an average length of 2^{24} were computed, thus about 2^{47} hashing operations were performed. The time spent and resources invested are moderate and within reasonable bounds since this procedure takes place only once in a network lifetime. The speed of the search could be significantly increased using GPUs or special purpose hardware like FPGA.

VII. Conclusions

The key distribution stands among the most critical security issues for wireless sensor networks. In this paper, we proposed and analyzed two improvements (and their combination) of the random key predistribution schemes. The first improvement exploits limited length collisions in secure hash functions to increase the probability of two nodes sharing a key. The second improvement introduces hash chains into the key pool construction to directly enhance the node capture resilience. Both these improvements can be further combined to bring the best performance. Our analytical results were supported by simulations.

Our improvements are particularly advantageous for networks where the attacker manages to capture a significant number of nodes. However, the benefits of our improvements are not limited to the basic random key predistribution schemes. The improvements could be employed, e.g., in the deterministic or hybrid approach proposed in [5]. We leave the investigation of such combination for the future work. Another challenge is to analyze the improvements face to face with a clever attacker who does not capture the nodes in a random fashion. Yet the impact of such an attacker could be limited by a deterministic selection of keys to be placed into key rings.

Appendix – Proofs and calculations

In this appendix we provide proofs of the selected equations and also justify the following statement that relates to the Definition 1. The formula in Definition 1 expresses the number of all possible key rings of size m selected from a key pool of size $|S|$ where no colliding key pair is present in the key rings.

Proof: We have $|S|$ possibilities how to select the first key for a key ring. After this selection, we mark the selected key and its colliding key off the key pool. Thus we have only $|S|-2$ possibilities how to select the second key. The keys are selected in this fashion until we select the m-th key for which only $|S|-2\cdot(m-1)$ possibilities remain. Since the order in which the keys were selected is not important, we divide the result by the number of permutations $m!$. ■

Proof of Equation 1: We have $\left\{ {|S| \atop m} \right\}$ possibilities how to select m keys into a key ring for any given node. Given these m keys, we have $\binom{m}{i}$ ways to select the i shared keys. Similarly, once these i shared keys have been picked, we have $\binom{m-i}{j}$ ways to select j shared collision keys that do not result from the i shared keys. Finally, we have to pick the remaining $m-i-j$ keys for the second key ring that are not the keys from the first key ring nor their colliding counterparts. Hence we pick them from the key pool without m colliding key pairs ($2m$ keys). This can be done by $\left\{ {|S|-2m \atop m-i-j} \right\}$ ways. Thus the number of key ring assignments for two nodes such that they share exactly i keys and are able to calculate exactly j collision keys excluding the collision keys resulting from the i shared keys is $\left\{ {|S| \atop m} \right\}\binom{m}{i}\binom{m-i}{j}\left\{ {|S|-2m \atop m-i-j} \right\}$. The total number of key ring assignments for any two nodes is $\left\{ {|S| \atop m} \right\}^2$. Thus the resulting probability is the fraction of these two values. ■

Proof of Equation 3: We follow and extend the proof from [2]. Since every node contains m keys out of $|S|$, the probability that an attacker obtains a particular key after a single node is captured is $\frac{m}{|S|}$. The probability that the attacker does not obtain the particular key after x nodes have been captured is thus $(1-\frac{m}{|S|})^x$. Finally, the probability that the attacker compromises a link key that is based on exactly i shared keys is $(1-(1-\frac{m}{|S|})^x)^i$.

Similarly, the probability that the attacker obtains a particular collision key after a single node is captured is $\frac{2m}{|S|}$, because we have only $\frac{|S|}{2}$ distinct collision keys and every node is able to calculate exactly m such keys. Hence, the probability that the attacker compromises a link key based on exactly j collision keys is $(1-(1-\frac{2m}{|S|})^x)^j$.

Assuming a link is secured with a link key, the probability that the link key is based on exactly i shared keys and j collision keys is $\frac{P_{SharedExactly}(i,j)}{P_{LinkEstablish1}}$. ■

Proof of Equation 6: Assume that two nodes were assigned keys from a given key-chain, then the probability that they establish i-th key of the key-chain as a shared key is $\frac{2\cdot i-1}{L^2}$. Furthermore, the probability that i-th key of a given key-chain was compromised after a random node was captured is $\frac{m}{|S|}\frac{i}{L}$. The probability that an attacker has compromised i-th key of a given key-chain after he captured x nodes is $1-(1-\frac{m}{|S|}\frac{i}{L})^x$. ■

Acknowledgments

We are grateful to the anonymous reviewers and to Luděk Smolík for their suggestions that improved the paper and to Tobiáš Smolka for his help with the BOINC infrastructure. This work was supported by the Czech research project VG20102014031, programme BV II/2 - VS. The first author is additionally supported by the project GD102/09/H042 "Mathematical and Engineering Approaches to Developing Reliable and Secure Concurrent and Distributed Computer Systems" of the Czech Science Foundation. The second and third authors are also supported by the research center Institute for Theoretical Computer Science (ITI), project No. P202/12/G061. The access to computing and storage facilities owned by parties and

projects contributing to the National Grid Infrastructure MetaCentrum, provided under the programme "Projects of Large Infrastructure for Research, Development, and Innovations" (LM2010005) is highly appreciated/acknowledged.

REFERENCES

[1] Eschenauer, L., Gligor, V.D.: A key-management scheme for distributed sensor networks. In: 9th ACM conference on Computer and Communications Security, CCS '02, pp. 41–47. ACM, New York (2002)
[2] Chan, H., Perrig, A., Song, D.: Random key predistribution schemes for sensor networks. In: Symposium on Security and Privacy, 2003, pp. 197–213. IEEE, (2003)
[3] Švenda, P., Sekanina, L., Matyáš, V.: Evolutionary design of secrecy amplification protocols for wireless sensor networks. In: 2nd ACM conference on Wireless network security, pp. 225–236. ACM, New York (2009)
[4] Ren, K., Zeng, K., Lou, W.: A new approach for random key predistribution in large-scale wireless sensor networks. Wireless Communications and Mobile Computing, 6(3):307–318, (2006)
[5] Camtepe, S.A., Yener, B.: Combinatorial design of key distribution mechanisms for wireless sensor networks. IEEE/ACM Transactions on Networking, 15(2):346–358, (2007)
[6] Xiao, Y., Rayi, V.K., Sun, B., Du, X., Hu, F., Galloway, M.: A survey of key management schemes in wireless sensor networks. Computer Communications, 30(1112):2314–2341, (2007)
[7] Rivest, R.L., Shamir, A.: PayWord and MicroMint: two simple micropayment schemes. Security Protocols, vol. 1189, pp. 69–87. Springer Berlin Heidelberg, Berlin, Heidelberg, (1997)
[8] Leighton, T., Micali, S.: Secret-key agreement without public-key cryptography. In: Advances in Cryptology – CRYPTO 93, LNCS 773, pp. 456–479. Springer Berlin Heidelberg, Berlin, Heidelberg, (1994)
[9] Di Pietro, R., Mancini, L.V., Mei, A.: Random key-assignment for secure wireless sensor networks. In: 1st ACM workshop on Security of ad hoc and sensor networks (SANS'03), pp.62–71. ACM, New York (2003)
[10] Kůr, J., Matyáš, V., Švenda, P. (2012, Sep.) Two Improvements of Random Key Predistribution for Wireless Sensor Networks, SecureComm 2012 – supplement data. [Online]. Available: http://www.fi.muni.cz/~xsvenda/papers/SecureComm2012/. Last accessed 20th Nov 2012.
[11] van Oorschot, P.C., Wiener, M.J.: Parallel collision search with cryptanalytic applications. Journal of Cryptology, 12(1):1–28, (1999)
[12] Gladman, B. (2007, Jan.) SHA1, SHA2, HMAC and Key Derivation in C. [Online]. Available: http://gladman.plushost.co.uk/oldsite/cryptography_technology/sha/index.php. Last accessed 20th Nov 2012.
[13] Anderson, D.P.: BOINC: A system for public-resource computing and storage. In: 5th IEEE/ACM International Workshop on Grid Computing, pp. 4–10. IEEE Computer Society (2004)

Jiří Kůr is a PhD student at the Laboratory of Security and Applied Cryptography, Faculty of Informatics, Masaryk University, in Brno, Czech Republic. He received his Masters degree in Computer Science at the Masaryk University, in 2008. His research fields are security and location privacy in wireless sensor networks. He is also interested in applied cryptanalysis. He was involved in several R&D projects focused on security of cryptographic smart cards. Jiří can be contacted at xkur AT fi.muni.cz.

Vashek Matyáš is a Professor at the Masaryk University, Brno, CZ, and serves as a Vice-Dean for Foreign Affairs and External Relations, Faculty of Informatics. His research interests relate to applied cryptography and security, publishing over a hundred peer-reviewed papers and articles, and co-authoring six books. He was a Fulbright Visiting Scholar with Harvard University, Center for Research on Computation and Society, and also worked with Microsoft Research Cambridge, University College Dublin, Ubilab at UBS AG, and was a Royal Society Postdoctoral Fellow with the Cambridge University Computer Lab. Vashek was one of the Editors-in-Chief of the Identity in the Information Society journal, and he also edited the Computer and Communications Security Reviews, and worked on the development of Common Criteria and with ISO/IEC JTC1 SC27. Vashek is a member of the Editorial Board of the Infocommunications Journal and a Senior Member of the ACM. He received his PhD degree from Masaryk University, Brno and can be contacted at matyas AT fi.muni.cz.

Petr Švenda is an Assistant Professor at the Masaryk University, Faculty of Informatics, in Brno, Czech Republic. He received his PhD degree from Masaryk University, working in the area of the cryptographic protocols for restricted environments such as wireless sensor networks, with focus on automatic generation of cryptographic protocols with the help of evolutionary algorithms. In 2008, he worked at TU Dresden on secure logging for the AN.ON anonymity service. He is also interested in practical aspects of security in cryptographic smartcards and their resistance against side-channel attacks and properties of random number generators available on smartcards and mobile devices. Petr can be contacted at svenda AT fi.muni.cz.

Privacy Scores: Assessing Privacy Risks Beyond Social Networks

Michal Sramka

***Abstract*—Assessing privacy risks arising from publishing private information on social networks is challenging for the users. Privacy Scores were proposed in the past to provide each user with a score – a measurement of how much sensitive information a user made available for others on a social network website. We present the privacy scores, discuss their shortcomings, and show several research directions for their extensions. We propose an extension that takes the privacy score metric from a single social network closed system to include background knowledge. Our examples and experimental results show the need to include publicly available background knowledge in the computation of privacy scores in order to get scores that more truthfully reflect the privacy risks of the users. We add background knowledge about users by means of combining several social networks together or by using simple web search for detecting publicly known information about the evaluated users.**

***Index Terms*—Privacy, Risk analysis, Inference algorithms.**

I. INTRODUCTION

RECENTLY there was an explosion of popularity of web sites that allow users to share information. These sites – social-network sites, blogs, and forums such as Facebook, Twitter, LinkedIn, and others – attract millions of users. The users publish and share information about themselves by creating online profiles, posting blogs and comments. Such information often contains personal details. Quantifying the individuals' privacy risk due to these information-sharing activities of the individuals is a challenging task.

Securing individuals' privacy in these environments and protecting users against threats such as *identity theft* and *digital stalking* becomes an increasingly important issue. Both users and service providers recognize the need for users' privacy. The sites may provide some privacy controls. However, the users are faced with too many options and too many controls, and lack the understanding of privacy risks and threats or are unable to accurately asses them. This all contributes to the confusion for the users, and often results in skipping the complicated and time-consuming tasks of setting the privacy controls that should protect them.

Even with properly configured privacy settings for a user profile, some privacy concerns remain. Take for example discussion forums, where tenths or hundreds contributions to multiple discussions of various topics are written by a user. Although the user is careful not to disclose any personally identifiable information in his/her individual posts, personal, sensitive, and private information may be disclosed by looking at the set of all posts by the user. From the cumulative set of all posts, it may be then possible to profile the user, infer the user's opinions or even identity.

Michal Sramka is with the Institute of Computer Science and Mathematics, Faculty of Electrical Engineering and Information Technology, Slovak University of Technology, Ilkovicova 3, 812 17 Bratislava, Slovakia, e-mail: sramka@stuba.sk

Privacy Scores by Liu and Terzi [1] were proposed to quantify the privacy risks of individuals posed by their profiles in a social-network site. Focus here is on privacy risks from the individuals' perspective. In the proposed framework, each user in a social network is assigned a privacy score based on the information in his/her profile compared to all available information in all profiles. The score then measures the user's potential privacy risk due to having his/her profile available on the social-network site.

The main drawbacks of this proposal of privacy scores are the concentration only on users' profiles and inconsideration of other publicly available information about the users on the same social network and beyond it. In particular, *background knowledge* about a user is not included in the computation of the privacy score. Background knowledge (sometimes referred to as external knowledge or auxiliary information) is some information about an individual that by itself is not a privacy disclosure, but combined with other information it becomes one.

A. Our Contribution

We propose a new concept for privacy scores. We explore the idea of presenting users with a new privacy score that measures their overall potential privacy risk due to available public information about them. Compared with the original privacy scores by Liu and Terzi, we overcome the drawbacks of concentrating only on users' profiles in a single social network, and we include publicly available background knowledge in computation of the new privacy scores. Our new privacy scores metric better represents the potential privacy risks of users and thus helps them make better decision in managing their privacy.

Our results are twofold. Firstly, in Section II-A we discuss the shortcomings of the privacy scores. We present several opportunities for extending the original privacy scores. With the extension of including background knowledge in mind, we identify some background knowledge that is publicly available but that cannot be easily extracted by computers in an automatic manner. Secondly, we proposed an extension of the privacy score metric that takes it from a closed system evaluating privacy over a single social network to a metric that includes information about the users that comes from outside the social network. In Section III, we present examples and experimental results showing paradoxes that may happen when the computation is over only a single social network. Next, in Section IV, we extend the computation of privacy scores to

include two or multiple social networks. Our final proposal, in Section IV-B, uses web searches to include all available public indexed human knowledge in the computation of the privacy score of a user. Thus, our new privacy score reflects the privacy risks of combining user's profile information with available knowledge about the user represented by the web.

Our proposed method for making web search inferences while scoring the privacy risks of individuals can also be seen as a privacy attack. However, we do not explore this direction, as there are already too many attacks, some of them referenced later in Section I-B. Our contribution rather focuses on helping users achieve their privacy needs and lower their privacy risks. The extended privacy score helps the users to make more informed decisions about their online activities.

B. Related Work

Our work is influenced by the approach by Liu and Terzi [1], which provides users with a quantification of privacy risks due to sharing their profiles in a social network. Each user is assigned a privacy score based on their and all other users' profile items. The proposal is for a single social-network site, that is, a closed system evaluation of privacy that lacks the consideration and inclusion of background knowledge in computation of the privacy scores. We overcome this shortcoming by including background knowledge in the computation of privacy scores, see Section III and IV.

PrivAware [2] is a Facebook application that scores privacy settings of a user based on the user's profile and profiles of his/her direct friends, which are implicitly available to any Facebook application. The score represents individuals' privacy risks arising from using third-party Facebook applications. In addition to [1] and [2], there exist several other scoring systems that somehow evaluate and rank users in social networks, but not their privacy. For some of them, please refer to [1]. However, none of these systems measures privacy risks for individuals.

The privacy risks of social-network sites are summarized in [3]. Several papers present privacy attacks in social networks [4], [2], [5], [6], [7], [8] or try to lower privacy risks and prevent privacy attacks in social networks [9], [10]. In addition, there are privacy risks from being tracked while browsing these websites [11].

Some form of background knowledge is usually considered in privacy attacks and is very likely available to attackers. Absolute privacy is impossible, because there will be always some background knowledge [12]. Inference techniques can then be used to attack or to help protect private data. In particular, web-based inference detection [13], [14] has been used to redact documents and prevent privacy leaks.

II. PRIVACY SCORES

Privacy Scores by Liu and Terzi [1] were proposed to quantify the privacy risks of individuals posed by their profiles in a social-network site. The privacy score is a combination of each one of user's profile items, labelled $1, \ldots, n$, for example, real name, email, hometown, land line number, cell phone number, relationship status, IM screen name, etc. The contribution of each profile item to privacy score is based on sensitivity and visibility. The *sensitivity* β_i depends on the item i itself – the more sensitive the item is, the higher is the privacy risk of it being revealed. The *visibility* of an item i belonging to a user j is denoted $V(i,j)$ and captures how far this item is known in the network – the wider the spread in the network, the higher the visibility.

The privacy score of an item i belonging to a user j is simply $\text{PR}(i,j) = \beta_i \times V(i,j)$. The overall *privacy score* for a user j with n items is then computed as

$$\text{PR}(j) = \sum_{i=1}^{n} \text{PR}(i,j) = \sum_{i=1}^{n} \beta_i \times V(i,j) \ . \qquad (1)$$

To keep the privacy score PR a non-decreasing function, in order for it to be a nicely behaving score, both the sensitivity β_i and visibility $V(i,j)$ must be non-negative functions. In practice, the sensitivity and visibility are determined from an $n \times m$ matrix R that represents n items for m users of a single social network. The value of each cell $R(i,j)$ describes the willingness of the user j to disclose the item i. In the simplest case, the value of $R(i,j)$ is 0 if the user j made the item i private and 1 if the item i is made publicly available. From this, the (observed) visibility can be defined as $V(i,j) = R(i,j)$. In a more granular approach, the matrix R can be defined by $R(i,j) = k$, representing that the user j disclosed the item i to all the users that are at most k jumps away in the graph of the social network. Regarding the sensitivity of an item, β_i can be computed using Item Response Theory (IRT) [1]. The IRT can be also used to compute the true visibility of an item for a user.

The privacy score is computed for each user individually. It is an indicator of the user's potential privacy risk – the higher the score of a user, the higher the threat to his/her privacy.

A. Shortcomings and opportunities of privacy scores

The privacy score is no doubt a useful metric for each and every user of a social network. Nevertheless, there exist several shortcomings of the originally proposed privacy scores. We list a few of them here. Some of these were already noticed and identified by the authors of the privacy scores, others are just observations, and some are our proposals for further exploration, research, and enhancements of privacy scores.

Regarding the items of a user profile, one can immediately notice hardship in quantifying the items themselves:

- The granularity of profile items is of particular concern. For example, the profile item "personal hobbies" can cover a range of non-private and private information and so its true sensitivity cannot be really established for the general case required by the privacy scores.
- Different profile items have different life-cycles. Some profile items may have a time attribute attached to them – for example, a cell/mobile phone number or an address are temporary information, while the date of birth or the mother's maiden name are permanent for life. The proposed privacy score, as defined, ignores these facts. We believe that implicit time relevance should be taken

into account for more precise evaluation of a user's privacy.

- Impossibility or hardness of including all, possibly private, information in privacy score computation. For example, consider photos: It may be hard or impossible in some cases to (automatically or even by a human involvement/assessment) establish relationships from photos. Or whether a person is drinking alcohol in a photo. Another example are discussion forums: Information is exhibited in natural language form. Determining a political orientation of a user from a single post may not be possible, yet looking at the cumulative set of the user's posts, private information can be inferred about the user (see Section III).

Of more concern and interest is the computation and use of sensitivity β_i for item i. As proposed, sensitivity is computed from the matrix R, that is, sensitivity is based only on the users and items in the single social network. When considering a single social network represented by a matrix R, it is easy to get a wrong perception of privacy due to the limited information about the users.

- The sensitivity β_i computed for an item i would reflect the true real-world sensitivity of this item only if the distribution of people in the social network would mirror the real-world distribution. Obviously, many social networks are not like this, and so a paradoxes are likely because this fact. For example, take a date of birth that most people consider a sensitive and private information. However, if everybody in a social network reveals his/her date of birth, then this item will be considered as not sensitive at all (because everybody reveals it). Paradoxes on the other side of the spectrum are possible, too. For example, if an item in a social network is filled only by one or a few users, because the other users are too lazy to fill it in, then the item will be considered sensitive (by the computation of sensitivity), although the item is far from being considered sensitive or private in the real life.
- No background knowledge inclusion, and so no inference detection or control: A privacy metric should include "background knowledge" (auxiliary information or external knowledge) in establishing a score for a user. Speaking more generally, a single social network or any closed system evaluation is not sufficient for real and proper privacy evaluation of a user.
  For privacy scores, this means that the computation of the score should not depend only on the matrix R coming from a single social network. Several extensions of the original privacy score metric are possible based on the background knowledge type and source. In Section IV we propose a new method to compute privacy scores, one that considers information about users beyond the ones in the social network, namely from a second/other social networks or more generally from the web.

Finally, it needs to be mentioned that the proposed privacy score metric measures only some aspect of privacy, namely attribute (item) disclosure and identity disclosure arising from the attribute disclosure. There are several other aspects that may be of concern to the users of a social network, such as:

- the risks of identity disclosure that is not based on attribute disclosure – for example, based on behavioral observations,
- the risk of identity theft,
- the risk of link or relationship disclosure,
- the risk of group membership disclosure, or
- the risk of digital stalking.

How to measure these risks and help the users making informed decisions by presenting them a score reflecting these risks is still an open problem.

III. Discussion forum

The computation of privacy scores proposed by Liu and Terzi [1] introduced in Section II assumes the analyzed information to be readily available for inclusion in the matrix R. As we noted in Section II-A, non-structured information cannot be always easily included for analysis. It may be either information that is hard to extract – for example, relationships from photos – or previously not defined information – for example, non-structured text in natural language may contain multiple private items some of which may not be pre-defined as items of the matrix R.

Together with my Master's student Ján Žbirka we performed a few experiments [15], where simple natural language analysis was used to determine if some private information has been included in discussion comments on a news website. Since the users usually post multiple comments, they may contain multiple private information that must be looked-up for inclusion in the privacy scores. In our experiments, shown in the next Section, we concentrated on information about political orientation before election and religious believes.

A. Experimental results

Discussions of the Slovak news web site www.sme.sk were analyzed just before the government election in March 2012. From all the users that posted comments on the website, 5,268 users who posted more than 500 comments over the lifetime of the website were considered as the most active ones. In the three weeks before the election, these 5,268 most-active users posted 43,035 comments that were analyzed. Almost 20% of the analyzed users revealed in their comments which political party in particular they were or were not going to vote for.

Summary of the findings are in Table I and all the other details about the experiment can be found in the Master's thesis of Ján Žbirka [15].

Since discussions on this website about religion and church are very heated, we also analyzed whether it is possible to find out the faith/religious beliefs of the users from their comments. The experiment that was done on the same sample of the users and comments have shown that simple natural language analysis can determine faith, although the users were more conservative in revealing their religious believes compared to the political orientation. In total, 133 (2.5%) users were found to disclose their religion, and 106 (2.0%) users were found to disclose that they are atheists.

TABLE I
THE NUMBER OF THE USERS (FROM THE TOTAL OF 5,268) WHO WERE FIND TO DISCLOSE THIS INFORMATION IN DISCUSSION COMMENTS

Users who will	vote		not vote	
at all	763	(14.5%)	173	(3.3%)
for a right wing party	209	(4.0%)	194	(3.7%)
for a left wing party	59	(1.1%)	46	(0.9%)
for a particular party	688	(13.1%)	335	(6.4%)

IV. PRIVACY SCORE EXTENSION

The biggest disadvantage of the privacy scores that were outlined in Section II is the non-consideration of background knowledge. *Background knowledge* (sometimes referred to as external knowledge or auxiliary information) is some information about an individual that by itself is not a privacy disclosure, but combined with other information it becomes one. We propose two possible extensions of the original privacy score metric that take public background knowledge into account.

It needs to be noted that the reason to include of background knowledge in the computation of the privacy score is two-fold. On the one hand, such extended privacy score will more precisely present users with privacy risks arising from publishing their information. On the other hand, using background knowledge also reduces another shortcoming of the original privacy scores. Namely, the more background knowledge is considered, the closer is the sensitivity of items to the true sensitivity. In other words, adding background knowledge to the privacy score computation also reduces or eliminates sensitivity paradoxes – see Section II-A.

Our extended privacy score metric uses the same formula as in the equation (1) with sensitivity and visibility as the original privacy scores, but the information that is used to compute these – the matrix R – is extended by additional knowledge. We discuss two instances of this extension. The first one, presented next, combines information from two or several social networks when evaluating privacy risk of a user. The second instantiation of the extended privacy score metric, which we present in Section IV-B, uses "all the human knowledge" in privacy risk evaluation.

Our proposal of a simple inclusion of additional information in the privacy score computation is based on users' information (items) from multiple social networks. Let N be the number of considered social networks, and let R_t be as the already defined matrix R for a social network t, with $t = 1, \ldots, N$. Hence, R_t is a $n \times m$ matrix, where $R_t(i, j)$ represents the publicity of an item i for a user j – that is, non-disclosure when $R_t(i, j) = 0$ or disclosure when $R_t(i, j) > 0$ and possibly how far from the user j is the item public in the (graph of the) social network t.

It is likely in practice that not all the users are in every social network and that every item is in each of the corresponding profiles. Here we assume that the range of the items $i = 1, \ldots, n$ and the range of the users $j = 1, \ldots, m$ are the supersets over all the social networks, and so $R_t(i, j) = 0$ if an item i or user j do not exist in the social network t. We define the matrix R used for sensitivity and visibility computation as $R(i, j) = \max_t R_t(i, j)$ and use the formula from the equation (1) to compute the privacy score. Such privacy score better estimates the risk of privacy disclosure.

Together with my Master's student Lucia Maringová we performed a few experiments [16], where the same users on two social networks were evaluated for their privacy risks. The two social networks were of different type, so it was expected that the users would behave differently and therefore would disclose different amount of information about themselves on each social network. In our experiments, shown in the next Section, we focused on computing privacy scores from each social network individually and then comparing the behavior of people in the terms of private information disclosure on two social networks.

A. Experimental results

The purpose of the experiment is to show that privacy risks, as measured by the original and extended privacy score, are higher when two social networks are combined. Specifically, this means that some users tend to be conservative in one social network while publicly disclose private information in another social network.

For the experiments, profiles from the same users on two social networks were downloaded and analyzed. The social networks (websites) were Pokec.sk and Zoznamka.sk. They both belong to the same content provider, and so use the same user authentication, which facilitated the pairing of the users from the two networks. Zoznamka.sk is a dating website, where a profile can contain up to 5 items: age, body type, weight, height, and contact. Pokec.sk is a website about chatting, messaging, and picture sharing. A profile on Pokec.sk can contain up to 34 items.

A sample of 3,923 users was selected. From all these users, there are only 23 users (<1%) who completely filled all profile items on both websites. These people can be considered very open minded and/or not understanding or ignoring the risks of disclosing private information. Roughly 32% of the users shared the same information on both sites.

Because of the nature of the website, users on the dating website Zoznamka.sk revealed more personal information about themselves. This is likely due to the fact that the users tried to create interest and attract the users who viewed their profiles. No user had less than 2 items (out of 5) filled on Zoznamka.sk. Conversely, many users on Pokec.sk left their profiles empty. What is of interest to us are the users who had empty profiles on Pokec.sk and non-empty profiles on Zoznamka.sk. Table II summarizes these users. All the details can be found in the Master's thesis of Lucia Maringová [16].

In the terms of the privacy score, the users on Pokec.sk who had empty profiles would receive the score of 0, because they do not share or disclose anything. However, this would be awfully wrong in any privacy risk analysis, because private information about these users is publicly available and linkable to these users. At least two additional items can be learned about roughly 44% of the users with empty profiles on Pokec.sk when considering Zoznamka.sk, so the extended privacy score computed over both networks for these users

TABLE II
THE NUMBER OF USERS WHO SHARED NOTHING ON POKEC.SK, BUT HAD NON-EMPTY PROFILES ON ZOZNAMKA.SK. NOTE THAT MINIMUM ITEMS FILLED IN ON ZOZNAMKA.SK WAS 2.

On Pokec.sk	On Zoznamka.sk	# of users
0 items	2 items	542
0 items	3 items	107
0 items	4 items	961
0 items	5 items	119
0 items	> 0 items	1,727
		44%

will be non-zero. This simple experiment itself shows the need to extend the original privacy scores from analyzing information over one social network to analyzing also auxiliary information.

B. Using all the human knowledge in privacy score computation

Extending the original Privacy Scores by Liu and Terzi [1] to multiple social networks certainly helps in privacy risk evaluation. The selection of social networks included in the extended privacy score computation, presented above, strongly impacts the quality and truthfulness of the score. The most truthful privacy risk evaluation can be achieved if all the human knowledge is used in the computation of the privacy score.

Including "all the human knowledge" in any computation is obviously impossible, so an approximation would have to suffice for all practical purposes. To effectively include the knowledge, we need to be able to quickly search for particular information or relation. Thus, we should use all the *indexed* human knowledge. Private databases and the "deep web" are believed to contain much more information than what is publicly available. In general, private information is out of reach for privacy adversaries as well as for privacy evaluators. Hence, we foresee to use all the *indexed public* human knowledge in the privacy score computation. Currently, the best instance and the best source of all the indexed public human knowledge is (Google) web search. In fact, there exists a proposal, namely Web-Based Inference Detection [13], [14], which takes advantage of the assumption that the web search is the proxy for all human knowledge.

Our idea is as follows: If an item of a user is not disclosed in the social network, we want to determine if the item has been disclosed elsewhere by using an inference detection based on the other disclosed items for the user. Our inference detection method is heavily influenced by the Web-Based Inference Detection [13]. So, our idea rewritten in the terms of inference detection is: If there is a privacy-impacting inference detected for an undisclosed item, then this detected inference should be included in the privacy score computation.

More formally, we propose the following method to compute the privacy score:

Consider a social network of m users each having a possibility to fill a profile of n items. Let R be, as before, the $n \times m$ matrix over $\{0, 1\}$ with $R(i, j)$ representing whether the user j has (or has not) disclosed the item i. Let P be $n \times m$ array of strings with $P(i, j)$ being the value of the item i for the user j, in case this value has been disclosed. Let the set D_i be the domain of the item i. Finally, let β, γ, and δ be positive integers, where β and γ are parameters of the proposed algorithm that control the search depth, and δ is the parameter that controls the number of the most frequent words to be considered. Then the algorithm to extend R and determine the users' disclosures outside the social network is as follows:

For each user j, $j \in \{1, \ldots, m\}$

1) Let $S_j = \{k \mid R(k, j) = 1,\ k = 1, \ldots, n\}$ be the set of all disclosed items for the user j.
2) For each undisclosed item i, that is, for all $i \in \{1, \ldots, n\}$ with $R(i, j) = 0$
   a) Let T be an empty multiset.
   b) Take every subset $S'_j \subseteq S_j$ of size $|S'_j| \leq \beta$.
   c) For every such subset $S'_j = \{i_1, \ldots, i_\ell\}$ with $\ell \leq \beta$
      i) Use a web search engine to search for keywords $P(i_1, j), \ldots, P(i_\ell, j)$
      ii) Retrieve the top γ most relevant documents containing these keywords
      iii) Extract the top δ most frequent words from all these γ documents
      iv) Add the top δ most frequent words to T together with their frequencies
   d) Take the most frequent word from T that is also in D_i, if it exists.
   e) If there is such word, let $R(i, j) = 1$.

After this, the newly enhanced matrix R contains the users' disclosures not just from the social network itself, but also from the web. Visibility and sensitivity values can be then computed from this matrix R, and the privacy score can be computed for each user using the equation (1).

The parameters β, γ, and δ can be tuned to achieve different trade-offs between the running time of the algorithm and the completeness and quality of the disclosure detection. In fact, these values can be different for different users, perhaps based on the number of items disclosed in the social network. Additional tuning can be achieved by performing the steps of the algorithm only for those users that have disclosed a "sufficient" number of items in the profile that would allow the web search to identify additional items.

V. CONCLUSIONS

As more and more users are joining and using social-network web sites, they become more heavily used and their owners look for new ways to share different content, including private information and information that may lead to unwanted privacy leakages. It becomes increasingly difficult for individuals to control and manage their privacy in the vast amount of information available and collected about them.

Privacy Scores were proposed as a metric that presents users with a score that reflects their privacy risks arising from disclosing information in their profiles on a social network. We presented several shortcomings of the privacy scores as research opportunities for extending the privacy score metric.

Next, we supported the need for extensions by experimental results from different websites and social networks. Finally, we proposed two extensions of the privacy score metric that consider additional background information about the users in the computation of the scores. Our approach provides a better decision support for individuals than the original privacy scores. Based on our extended privacy score metric, the users can compare their privacy risks with other fellow individuals and make informed decisions about whether they share too much potentially private and sensitive information.

Michal Sramka has received PhDs in Mathematics and Applied Mathematics from Florida Atlantic University and Slovak University of Technology, and also holds MSc in Computer Science. He has worked as a researcher at universities in the USA, Canada, Spain, and Slovakia. His main research interests are in information security, specifically in data privacy and cryptology. Michal Sramka is the holder of the Werner von Siemens Excellence Award and a lifelong member of the Pi-Mu-Epsilon honorary society.

REFERENCES

[1] K. Liu and E. Terzi, "A Framework for Computing the Privacy Scores of Users in Online Social Networks," in *Proceedings of the Ninth IEEE International Conference on Data Mining, ICDM 2009*, 2009, pp. 288–297.

[2] J. Becker and H. Chen, "Measuring Privacy Risk in Online Social Networks," in *Web 2.0 Security & Privacy 2009 Workshop of 2009 IEEE Symposium on Security and Privacy, W2SP 2009*, 2009.

[3] D. S. Rosenblum, "What Anyone Can Know: The Privacy Risks of Social Networking Sites," *IEEE Security & Privacy*, vol. 5, no. 3, pp. 40–49, 2007.

[4] A. Narayanan and V. Shmatikov, "De-anonymizing Social Networks," in *Proceedings of the 30th IEEE Symposium on Security and Privacy, S&P 2009*, 2009, pp. 173–187.

[5] B. Krishnamurthy and C. E. Wills, "On the leakage of personally identifiable information via online social networks," in *Proceedings of the 2nd ACM Workshop on Online Social Networks, WOSN 2009*, 2009, pp. 7–12.

[6] E. Zheleva and L. Getoor, "To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles," in *Proceedings of the 18th International Conference on World Wide Web, WWW 2009*, 2009, pp. 531–540.

[7] A. Korolova, R. Motwani, S. U. Nabar, and Y. Xu, "Link privacy in social networks," in *Proceedings of the 17th ACM Conference on Information and Knowledge Management, CIKM 2008*, 2008, pp. 289–298.

[8] L. Backstrom, C. Dwork, and J. M. Kleinberg, "Wherefore Art Thou R3579X?: Anonymized Social Networks, Hidden Patterns, and Structural Steganography," in *Proceedings of the 16th International Conference on World Wide Web, WWW 2007*, 2007, pp. 181–190.

[9] A. Felt and D. Evans, "Privacy Protection for Social Networking Platforms," in *Web 2.0 Security & Privacy 2008 Workshop of 2008 IEEE Symposium on Security and Privacy, W2SP 2008*, 2008.

[10] E. Zheleva and L. Getoor, "Preserving the Privacy of Sensitive Relationships in Graph Data," in *Proceedings of the First ACM SIGKDD International Workshop on Privacy, Security, and Trust in KDD, PinKDD 2007*, 2007, pp. 153–171.

[11] K. McKinley, "Cleaning Up After Cookies," iSEC Partners, Technical report, 2008. [Online]. Available: https://www.isecpartners.com/files/iSEC_Cleaning_Up_After_Cookies.pdf

[12] C. Dwork, "Differential Privacy," in *Proceedings of the 33rd International Colloquium on Automata, Languages and Programming, ICALP 2006*, 2006, pp. 1–12.

[13] J. Staddon, P. Golle, and B. Zimny, "Web-Based Inference Detection," in *Proceedings of the 2007 USENIX Annual Technical Conference, USENIX 2007*, 2007, pp. 71–86.

[14] R. Chow, P. Golle, and J. Staddon, "Inference detection technology for Web 2.0," in *Web 2.0 Security & Privacy 2007 Workshop of 2007 IEEE Symposium on Security and Privacy, W2SP 2007*, 2007.

[15] J. Zbirka, "Privacy risks arising from publishing private information on the web (in Slovak)," Master's Thesis, Advisor: Michal Sramka, Slovak University of Technology, 2012.

[16] L. Maringova, "Privacy risks arising from publishing private information in social networks (in Slovak)," Master's Thesis, Advisor: Michal Sramka, Slovak University of Technology, 2012.

Accelerating Biometric Identification

David Naccache, Zdenek Riha

***Abstract*—By opposition to biometric matching, biometric identification is a relatively costly process. Let $B = \{b_1, \ldots, b_n\}$ be a database of n biometric templates and let b be a given individual biometric acquisition. The biometric identification problem consists in finding the most likely b_i corresponding to b. This paper assumes the existence of an oracle $\mathfrak{A}$ taking as b and b_i, and responding with true or false. Considering $\mathfrak{A}$ as an atomic operation, any system-level optimization must necessarily minimize the number of calls to $\mathfrak{A}$ per identification session. This is the parameter that we optimize in this paper.**

***Index Terms*—biometrics, biometric identification, correlation**

I. Introduction

By opposition to biometric matching, biometric identification [2], [3] is a relatively costly process. Let $B = \{b_1, \ldots, b_n\}$ be a database of n biometric templates and let b be a given individual biometric acquisition. The biometric identification problem consists in finding the most likely b_i corresponding to b [1].

Whilst in reality matching algorithms return a score compared to a threshold, for the sake of simplicity this paper assume the existence of an oracle $\mathfrak{A}$ taking b and b_i as input, and responding with true or false:

$$\mathfrak{A}(b, b_i) \in \{\mathsf{T}, \mathsf{F}\}$$

Considering $\mathfrak{A}$ as an *atomic* operation, any system-level optimization must necessarily minimize the number of calls to $\mathfrak{A}$ per identification session. This is the parameter that we attempt to optimize in this paper.

For doing so, we assume that every user i has a collection of k additional biometric parameters $m_{i,1}, \ldots m_{i,k}$. An $m_{i,j}$ can be either derived from the template b_i (*i.e.* $m_{i,j} = \text{function}_j(b_i)$) or measured independently. For instance if b_i is a fingerprint then $m_{i,7}$ can be the density of minutiae (the number of minutiae per unit of surface) or an additional parameter, such as the person's height, which is not correlated to b_i.

We will use the $m_{i,j}$ to accelerate identification by applying $\mathfrak{A}$ to the *most probable candidates first*. We denote by σ_j the standard deviation of the $m_{i,j}$'s, for all users i.

The proposed identification process is:

1) Acquire the biometric candidate information b and the additional information $m_1, \ldots, m_k$.
2) Compute for every user i the score:

$$t_i = \sum_{j=1}^{k} \frac{(m_j - m_{i,j})^2}{{\sigma_j}^2} \tag{1}$$

3) Try $\mathfrak{A}(b, b_i)$ by order of *increasing* t_i values.

Given that $\mathfrak{A}$ will be applied to the most promising candidates first (the ones with the lowest t_i), this is likely to result in a significantly faster identification procedure.

However, the comparison of the t_i's assumes that the $m_{i,j}$ are independent. This is not always the case. For instance a tall person is likely to be heavier. In other words, height (*e.g.* $m_{i,2}$) and weight (*e.g.* $m_{i,5}$) are *correlated*.

The process described so far did not take such correlations into account.

II. Analysis of the Procedure

We start by analyzing the proposed procedure without taking correlations into account.

The computation of the t_i's as given by equation (1) rests on the assumption that the measurements m_j each follow an independent normal distribution. More precisely, assuming that every measurement m_j follows a normal distribution with mean μ_j and standard deviation σ_j, the density function can be expressed as:

$$f_{m_j}(x) = \frac{1}{\sigma_j \sqrt{2\pi}} \exp\left(-\frac{(x-\mu_j)^2}{2{\sigma_j}^2}\right)$$

When the m_j's are independently distributed, the probability density of all measurements m_j for $1 \leq j \leq k$ can be expressed as a k-dimensional multivariate distribution:

$$f_{\vec{m}}(\vec{x}) = \prod_{j=1}^{k} f_{m_j}(x_j) = \frac{1}{(2\pi)^{k/2} \prod_{j=1}^{k} \sigma_j} \exp\left(-\sum_{j=1}^{k} \frac{(x_j - \mu_j)^2}{2{\sigma_j}^2}\right)$$

where $\vec{x} = (x_1, \ldots, x_k)$.

Note that in the previous equation μ_j and σ_j are the mean and standard deviation of m_j for *all* users i. For a measurement m_j corresponding to a specific user i, we can also assume that m_j follows a normal distribution with mean $\tilde{\mu_j} = m_{i,j}$ and standard deviation $\tilde{\sigma_j}$; we also assume that the standard deviation $\tilde{\sigma}_j$ around $m_{i,j}$ is the same for all users. In this case, the measurement m_j corresponding to user i has the following distribution:

$$f_{\vec{m}}(\vec{x}) = \frac{1}{(2\pi)^{k/2} \prod_{j=1}^{k} \tilde{\sigma}_j} \exp\left(-\sum_{j=1}^{k} \frac{(x_j - m_{i,j})^2}{2{\tilde{\sigma}_j}^2}\right)$$

D. Naccache is a researcher at the École normale supérieure's Cryptography Group and a professor at the University of Paris II (email: david.naccache@ens.fr).

Z. Riha is with the Masaryk University, Faculty of Informatics, Brno, Czech Republic (email: zriha@fi.muni.cz).

Additionally, we assume that the standard deviation $\tilde{\sigma}_j$ of m_j around $m_{i,j}$ is proportional to the standard deviation σ_j of m_j when all users are considered, *i.e.* we assume $\tilde{\sigma}_j = \alpha \cdot \sigma_j$ for all $1 \leq j \leq k$ for some $\alpha \in \mathbb{R}$. In this case, the probability density function of the m_j's for user i can be written as:

$$f_i(\vec{m}) = \frac{1}{(2\pi)^{k/2}\alpha^k \prod_{j-1}^{k} \sigma_j} \exp\Big(-\sum_{j=1}^{k} \frac{(m_j - m_{i,j})^2}{2\alpha^2 {\sigma_j}^2}\Big) =$$

$$= \frac{1}{(2\pi)^{k/2}\alpha^k \prod_{j-1}^{k} \sigma_j} \exp\Big(-\frac{t_i}{2\alpha^2}\Big)$$

where t_i is precisely the quantity given by equation (1). The probability to obtain measurements m_j from user i is thus a decreasing function of t_i. Given $\vec{m}$, the most probable candidate is hence the one with the lowest t_i.

III. Taking Correlation into Account

The comparison of the t_i's assumes that the different biometric measurements $m_{i,j}$ are independent. This is not necessarily the case since (for instance) a tall person is likely to be heavier; in other words, height and weight are correlated. In this section we the definition of t_i to take correlation into account.

A. Multivariate Normal Distribution

We denote by Σ the covariance matrix of the measurements m_j, defined as follows:

$$\Sigma = \mathrm{var}(\vec{m}) = \mathrm{var}\begin{pmatrix} m_1 \\ \vdots \\ m_k \end{pmatrix} =$$

$$= \begin{pmatrix} \mathrm{var}(m_1) & \mathrm{cov}(m_1 m_2) & \cdots & \mathrm{cov}(m_1 m_k) \\ \mathrm{cov}(m_1 m_2) & \ddots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(m_1 m_k) & \cdots & \cdots & \mathrm{var}(m_k) \end{pmatrix}$$

where $\mathrm{cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$ and $\mathrm{var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2$.

We assume that the measurements m_j follow a k-dimensional multivariate distribution with mean $\vec{\mu}$ and covariance matrix Σ; the probability density function can then be expressed as:

$$f_{\vec{m}}(\vec{x}) = \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} \exp\Big(-\tfrac{1}{2}(\vec{x} - \vec{\mu})'\Sigma^{-1}(\vec{x} - \vec{\mu})\Big)$$

where $|\Sigma|$ is the determinant of Σ. Note that mean $\vec{\mu}$ is a k-dimensional vector and Σ is a $k \times k$-matrix.

Note that in the previous equation $\vec{\mu}$ and Σ are the expected value and covariance matrix of measurements m_j for *all* users i. As in Section II, for measurements m_j's corresponding to a specific user i, we also assume that the m_j's follow a k-multivariate normal distribution with mean $\tilde{\mu}_j = m_{i,j}$ and covariance matrix $\tilde{\Sigma}$; we also assume that $\tilde{\Sigma}$ is the same for all users. In this case, the measurement $\vec{m}$ for user i follows the multivariate distribution:

$$f_{\vec{m}}(\vec{x}) = \frac{1}{(2\pi)^{k/2}|\tilde{\Sigma}|^{1/2}} \exp\Big(-\tfrac{1}{2}(\vec{x} - \vec{m_{i,\cdot}})'\tilde{\Sigma}^{-1}(\vec{x} - \vec{m_{i,\cdot}})\Big)$$

As in Section II we additionally assume that the covariance matrix satisfies $\tilde{\Sigma} = \alpha \cdot \Sigma$ for some $\alpha \in \mathbb{R}$. In this case, the probability density function can be written as:

$$f_{\vec{m}}(\vec{x}) = \frac{1}{(2\pi\alpha)^{k/2}|\Sigma|^{1/2}} \exp\Big(-\tfrac{1}{2\alpha}(\vec{x} - \vec{m_{i,\cdot}})'\Sigma^{-1}(\vec{x} - \vec{m_{i,\cdot}})\Big)$$

which gives:

$$f_{\vec{m}}(x) = \frac{1}{(2\pi\alpha)^{k/2}|\Sigma|^{1/2}} \exp\Big(-\frac{t_i}{2\alpha}\Big)$$

where:

$$t_i = (\vec{m} - \vec{m_{i,\cdot}})'\Sigma^{-1}(\vec{m} - \vec{m_{i,\cdot}}) \tag{2}$$

Therefore we obtain that equation (2) is a generalization of equation (1) when taking correlations into account.

B. The New Identification Procedure

The new algorithm is:

1) Collect from the user the biometric information b and the additional information $m_1, \ldots, m_k$.
2) Compute for every user i the value:
$$t_i = (\vec{m} - \vec{m_{i,\cdot}})'\Sigma^{-1}(\vec{m} - \vec{m_{i,\cdot}})$$
3) Sort the t_i's by increasing values and apply $\mathfrak{A}(b, b_i)$ to user i by increasing t_i values.

C. Bivariate Case

To illustrate the algorithm we first restrict ourselves to the bivariate case. In this case, the covariance matrix between variables X and Y can be written:

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$$

where $\mathrm{var}(X) = \sigma_x^2$, $\mathrm{var}(Y) = \sigma_y^2$ and $\mathrm{cov}(X,Y) = \rho\sigma_x\sigma_y$ where ρ is the correlation between X and Y. In this case, we have:

$$\Sigma^{-1} = \frac{1}{1-\rho^2}\begin{bmatrix} \frac{1}{\sigma_x^2} & \frac{-\rho}{\sigma_x\sigma_y} \\ \frac{-\rho}{\sigma_x\sigma_y} & \frac{1}{\sigma_y^2} \end{bmatrix}$$

and the probability density function can be written:

$$f(x,y) =$$

$$\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} - \frac{2\rho xy}{\sigma_x\sigma_y}\right]\right)$$

In this case, equation (2) gets simplified as follows:

$$t_i = \frac{(m_1 - m_{i,1})^2}{\sigma_1^2} + \frac{(m_2 - m_{i,2})^2}{\sigma_2^2} -$$

$$- \frac{2\rho(m_1 - m_{i,1})(m_2 - m_{i,2})}{\sigma_1\sigma_2}$$

where $\sigma_1 = \mathrm{var}(m_1)$, $\sigma_2 = \mathrm{var}(m_2)$ and ρ is the correlation between m_1 and m_2.

D. Illustration

We illustrate this with a set of simulated measurements: height, weight and number of collected minutiae, for 13 users.

User	1	2	3	4	5	6	7
Height	178	165	190	176	174	192	182
Weight	71	66	82	80	76	85	76
Minutiae	14	15	14	27	15	25	14

User	8	9	10	11	12	13
Height	162	168	175	187	195	168
Weight	65	80	77	68	92	72
Minutiae	22	23	24	23	19	25

We obtain the following correlation matrix:

$$\Sigma = \begin{bmatrix} 104.9 & 52.9 & -5.2 \\ 52.9 & 56.3 & 3.9 \\ -5.2 & 3.9 & 22.8 \end{bmatrix}$$

which can be written as:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 \\ \rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 \end{bmatrix}$$

where $\sigma_1 = 10.2$, $\sigma_2 = 7.5$, $\sigma_3 = 4.8$, and $\rho_{12} = 0.688390$, $\rho_{13} = -0.107015$, $\rho_{23} = 0.109587$.

Since ρ_{13} and ρ_{23} are small, for simplicity we consider only correlations between the first and second variables (height and weight). More precisely we consider the simplified covariance matrix:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \\ & & \sigma_3^2 \end{bmatrix}$$

with the same previous values of σ_1, σ_2, σ_3 and $\rho = \rho_{12}$. This gives:

$$\Sigma^{-1} = \begin{bmatrix} \frac{1}{(1-\rho^2)\sigma_1^2} & \frac{-\rho}{(1-\rho^2)\sigma_1\sigma_2} & \\ \frac{-\rho}{(1-\rho^2)\sigma_1\sigma_2} & \frac{1}{(1-\rho^2)\sigma_2^2} & \\ & & \frac{1}{\sigma_3^2} \end{bmatrix}$$

This gives the following formula for t_i which takes into account correlations between height and weight:

$$t_i = \frac{(m_1 - m_{i,1})^2}{(1-\rho^2)\sigma_1^2} + \frac{(m_2 - m_{i,2})^2}{(1-\rho^2)\sigma_2^2} - \\ - \frac{2\rho(m_1 - m_{i,1})(m_2 - m_{i,2})}{(1-\rho^2)\sigma_1\sigma_2} + \frac{(m_3 - m_{i,3})^2}{\sigma_3^2}$$

IV. Conclusions

In the paper we have presented an approach to accelerate the biometric identification process. The algoritm is based on the basic principle of testing the most probable candidates first. We started with assumption that set of measurements of a user are considered to be independent and later we introduced correlations into the scheme.

One drawback of the previous technique is that given a measurement $\vec{m} = (m_1, \ldots, m_k)$ the t_i's must be computed for all users i. A possible speed-up could be to select only those users i for which $|m_1 - m_{i,1}|$ is relatively small. This can be done efficiently if the values $m_{i,1}$ are pre-sorted. Another refinement consists computing all the t_i's simultaneously (*i.e.* compute j-wise rather than i-wise), progressively delay the computation of "heavier" t_i's and start the comparison of the "lighter" ones as soon as these become available.

References

[1] Michael E. Schuckers. Computational Methods in Biometric Authentication. Springer, London, 2010. ISBN 978-1-84996-201-8.
[2] Herve Jarosz and Jean-Christophe Fondeur. Large-Scale Identification System Design. In *Biometric Systems Technology, Design and Performance Evaluation*. Springer, 2005. ISBN: 978-1-85233-596-0.
[3] Michael Brauckmann and Christoph Busch. Large Scale Database Search. In *Handbook of Face Recognition*. 2011, pp 639-653. ISBN: 978-0-85729-931-4.

David Naccache is a researcher at the École normale supérieure's Cryptography Group and a professor at the University of Paris II. His research interests include public-key cryptography and mobile code security. Naccache has a PhD in cryptology from the École nationale supérieure des télécommunications Paris. Contact him at david.naccache@ens.fr.

Zdenek Riha a is an Assistant Professor at the Masaryk University, Faculty of Informatics, in Brno, Czech Republic. He received his PhD degree from the Faculty of Informatics, Masaryk University. In 1999 he spent 6 months on an internship at Ubilab, the research lab of the bank UBS, focusing on security and usability aspects of biometric authentication systems. Between 2005 and 2008 he was seconded as a Detached National Expert to the European Commission's Joint Research Centre in Italy. Zdenek can be contacted at zriha@fi.muni.cz.

DRCN 9th International Conference on Design of Reliable Communication Networks

March 4-7, 2013
Budapest, Hungary

www.drcn2013.org

DRCN is a well established biennial forum for scientists, engineers, designers and planners from industry, government and academia who have interests in the reliability and availability of communication networks and services. The conference covers topics from equipment and technology for survivability to network management and public policy, through theory and techniques for survivable and robust networks and application design. The aim of the conference is to bring together people from industry, government and academia in those disciplines in a lively forum.

To guarantee the high visibility of the conference, the proceedings will be available through IEEE Xplore. The best papers will be invited to leading related Journals: IEEE Transactions on Reliability (ToR) and Elsevier Optical Switching and Networking (OSN).

The 9th DRCN will be held in the attractive city of Budapest, Hungary. The conference will be enriched by a set of tutorials and invited talks. Additionally, two IFIP best paper awards will be granted. We seek papers that address theoretical, experimental, systems-related and regulatory issues in the area of dependability and survivability of communication networks, end-systems and infrastructure. Topics of interest include, but are not limited to the following areas:

- Operational aspects:
  - Fault management, monitoring, and control
  - Methodologies, equipment and technology for network survivability
  - Survivability of optical and multi-layer networks
  - Reliability of wireless access and mesh networking
  - Resilient wired access networks
  - Dependability of cellular/mobile networks including horizontal handover
  - Resilience of multi-domain connections in the Internet
  - Reliability of emerging technologies (e.g. network virtualization, thin client architectures)
  - Survivability in grid and distributed computing
  - Network dependability in cloud computing
  - Management of survivable networks
- Theory and modelling:
  - Network reliability analysis
  - Methods and theory for survivable network and systems design, analysis and operation (including scalability and complexity)
  - Planning and optimization of reliable networks, systems, and services
  - Simulation techniques for network resilience
- Services:
  - Reliability requirements and metrics for individual users, businesses, and the society
  - Restoration of services under various types of failures
  - Service differentiation based on recovery methods
  - Dependability of networked applications
  - Recovery of overlay and peer-to-peer networks
  - Application and service-specific survivability techniques
  - Survivability of multimedia networks including voice over IP, IPTV, and content delivery
  - Reliability and resilience of data centre networks
  - Robustness of compound services
- Broad context:
  - Telecommunication networks as an element of critical national infrastructures
  - Public policy issues for survivability and resilience
  - Standardization of network resilience and reliability
  - Network resilience combined with economics and commercial issues
  - Quality of experience and network survivability
  - Security issues in networks and their relation to survivability
  - Dependability and energy consumption trade-offs
  - Risk and reliability in the Internet and enterprise networks
  - New and emerging threats

STEERING COMMITTEE

Piet Demeester, *(Steering Committee Chair), Ghent University - IBBT -IMEC, Belgium*

Prosper Chemouil, *Orange Labs, France*

Tibor Cinkler, *Budapest University of Technology and Economics, Hungary*

Roberto Clemente, *Telecom Italia, Italy*

Robert Doverspike, *AT&T Labs, USA*

Wayne D. Grover, *TRLabs, University of Alberta, Canada*

Deep Medhi, *University of Missouri-Kansas City, USA*

Ken-ichi Sato, *Nagoya University, Japan*

Dominic Schupke, *Nokia Siemens Networks, Germany*

David Tipper, *University of Pittsburgh, USA*

TPC CHAIRS

Pin-Han Ho, *University of Waterloo, Canada*

János Tapolcai, *Budapest University of Technology and Economics, Hungary*

TECHNICAL CO-SPONSORS

IFIP *(pending)*

HTE: *Hungarian Scientific Association for Infocommunications (Sister Society of IEEE)*

BME: *Budapest University of Technology and Economics*

University of Waterloo

EC FP7 IP COMBO *(pending)*

GENERAL CHAIR

Tibor Cinkler, *Budapest University of Technology and Economics, Hungary*

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
Dept. of Telecommunications
Budapest University of Technology and Economics
2 Magyar Tudosok krt.
Budapest 1117 Hungary
szabo@hit.bme.hu

Our Reviewers in 2012

The quality of a research journal depends largely on its reviewing process and, first of all, on the professional service of its reviewers. It is my pleasure to publish the list of our reviewers in 2012 and would like to express my gratitude to them for their devoted work.

Your Editor-in-Chief

Luigi Atzori,
University of Cagliari, Italy

László Bacsárdi,
BME, Hungary

Attila Börcs,
SZTAKI, Hungary

Iacopo Carreras,
CREATE-NET, Italy

Tibor Cinkler,
BME, Hungary

László Csurgai-Horváth,
BME, Hungary

László Czuni,
University Veszprém, Hungary

Khadija Daoud,
France Telecom – Orange

Franco Davoli,
University of Genova, Italy

Károly Farkas,
BME, Hungary

Zoltán Gál,
University of Debrecen, Hungary

Enrico Gregori,
CNR IIT, Pisa, Italy

András Gulyás,
BME, Hungary

Lajos Hanzo,
University of Southampton, UK

Majd Hawasly,
University Edinburgh, UK

Gábor Horváth,
BME, Hungary

Jukka Huhtamaki,
Tampere University Technology, Finland

Árpád Huszák,
BME, Hungary

Sándor Imre,
BME, Hungary

László T. Kóczy,
Széchenyi University of Győr, Hungary

János Levendovszky,
BME, Hungary

Thomas Magedanz,
Fraunhofer FOKUS, Germany

Shadaydeh Maha,
SZTAKI, Hungary

Oscar Mayora,
CREATE-NET, Italy

Sándor Molnár,
BME, Hungary

Julius Müller,
Fraunhofer FOKUS, Germany

Vilmos Simon,
BME, Hungary

Csaba A. Szabó,
BME, Hungary

János Sztrik,
University of Debrecen, Hungary

István Tétényi,
SZTAKI, Hungary

Jan Turan,
Technical University of Kosice, Slovakia

Adam Wolisz,
Technical University Berlin, Germany

Sandro Zappatore,
University of Genova, Italy

Gergely Zaruba,
University of Texas as Arlington, USA

Honggang Zhang,
Zheijang University, China

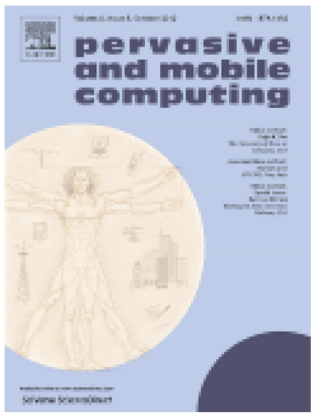
(* BME – Budapest University of Technology and Economics)

Contents
of the Infocommunications Journal 2012 (Volume IV)

IEEE PerCom in Budapest in 2014!

The **IEEE Pervasive Computing and Communication** (PerCom) conference is the worldwide premier scholarly venue in the areas of pervasive computing and communications. Since 2003, the conference has grown significantly in terms of quality and variety of the technical programs – it is recognized as a top tier conference by most universities and organizations across the world.

PerCom provides a high profile, leading edge forum for researchers, engineers, and practitioners to present state-of-the-art research in the respective fields of pervasive computing and communications. The conference features a diverse mixture of presentation forums including core technical sessions, keynote talks, panel discussions from worldwide experts, demonstrations, a PhD forum, and work in progress posters. The conference also hosts a number of workshops that have themselves become well recognized in the community as forums for specialized topics within the field.



The conference also provides formats to honor excellence in the field. The Mark Weiser Best Paper award, sponsored by Elsevier, is given to authors of the PerCom's best paper. In addition, the highest quality papers from the conference are published in a special issue of the Pervasive and Mobile Computing Journal.

The IEEE PerCom Steering Committee has recently decided to accept the joint application of the Budapest University of Technology and Economics (BME), Department of Telecommunications and the Scientific Association for Infocommunications (HTE) to organize PerCom 2014 in Hungary. Thus, after the 2013 edition in San Diego, Budapest will host this prestigious conference in 2014. It will be a 5 day-event, with associated workshops, and will hopefully attract several hundreds of participants. More information can be found on the conference's website:

www.percom.org

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 more than 1300 individual members, the Scientific Association for Infocommunications (in Hungarian: HÍRKÖZLÉSI ÉS INFORMATIKAI 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. HTE is corporate member of International Telecommunications Society (ITS).

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 HUSZTY** • *ghuszty@entel.hu*

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

Managing Director, Deputy Secretary-General: **PÉTER NAGY** • *nagy.peter@hte.hu*

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

Addresses

Office: H-1055 Budapest, V. Kossuth Lajos square 6-8, Room: 422.

Mail Address: 1372 Budapest, Pf. 451., Hungary

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

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