JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY

1/2019

(*Contents Continued on Back Cover*)

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

Circulation: 300 copies Sowa – Druk na życzenie, www.sowadruk.pl, tel. 22 431-81-40

JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY

Preface

We are immensely pleased to introduce you to the latest issue of the *Journal of Telecommunications and Information Technology* quarterly.

The articles presented in this edition cover some important problems experienced in the field of broadly understood telecommunications (from network protocols and services to telecommunications systems and security issues). The current issue of the Journal comprises fourteen papers.

The first group of articles brings up issues related to various problems occurring in wireless communication.

The first paper, titled *ACO-Inspired Energy-Aware Routing Algorithm for Wireless Sensor Networks*, was written by R. Yamamoto, S. Nishibu, T. Yamazaki, Y. Okamura and Y. Tanaka. It proposes a routing algorithm, known as AERO, that was inspired by the concept of ant colony optimization for Wireless Sensor Networks. This algorithm enables to balance traffic loads by utilizing transient optimization behaviors. AREO requires a shorter transmission to send the same amount of data and improves energy efficiency compared with other solutions of this type.

The second paper, titled *Adaptive Load Balancing Ad Hoc Routing Scheme Inspired by True Slime Mold*, by H. Katada, T. Yamazaki and T. Miyoshi, proposes an adaptive ad hoc routing method that is capable of constructing multiple paths based on the bandwidth available for each link, transmission data size and residual battery level of the node. It relies on the physarum solver that is applied to dynamic networks. The authors confirmed that the proposed method may adaptively construct single or multiple paths based on available bandwidth, transmission data size and residual battery level of nodes within a dynamic network topology.

The next paper, titled *Robot Local Network using TQS Protocol for Land-to-Underwater Communications*, by A. Irawan, M. F. Abas and N. Hasan, presents the modeling and analysis of the proposed Tag QoS switching (TQS) protocol for a heterogeneous robot operating in different environments. The proposed TQS protocol was inspired by multiprotocol label switching (MPLS) with good quality of services (QoS) achieved.

The paper titled *Fuzzy Clustering with Multi-Constraint QoS Service Routing in Wireless Sensor Networks*, by J. Agarkhed, V. Kadrolli and S. R. Patil, proposes a fuzzy logic-based, service-differentiated, QoS-aware routing protocol with multipath routing for wireless sensor networks. The proposed solution relies on a modified QoS k-nearest neighborhood technique.

Similarities between human brain and dense wireless networks have become an inspiration for the authors of the article titled *Neuroplasticity and Microglia Functions Applied in Dense Wireless Networks* (Ł. Kułacz and A. Kliks). The proposed concept is based on the idea of wireless neurons. The neurons are stand-alone devices which do not require a central management unit – a feature that enables scalability and easy reconfiguration for dense wireless networks.

The paper titled *Empirical Approach in Topology Control of Sensor Networks for Urban Environment*, by B. Musznicki, presents solutions for controlling topology of wireless sensor networks.

The next paper by T. Miyoshi, Y. Shimomura and O. Fourmaux is titled *A P2P-based Communication Framework for Geo-Location Oriented Networks*. It proposes a novel peer-to-peer communication framework to realize geographical location-oriented networks called GLocON. G-LocON provides geolocation-oriented device-to-device communication, relying solely on current wireless technologies, such as LTE and Wi-Fi, and cooperating with the global positioning system and peer-to-peer overlay networking.

The paper titled *LoCO: Local Cooperative Data Offloading System Based on Location Information*, drawn up by T. Yamazaki, K. Asano, S. Arai, Y. Shimomura and T. Miyoshi, proposes a local cooperative data offloading system (LoCO) that reduces the overall traffic by sharing data via direct communication between neighbors, based on their location information.

Two subsequent papers deal with problems encountered in telecommunication systems. The paper titled *Rectangular Dielectric Resonator Antenna with Single Band Rejection Characteristics*, by M. Debab and Z. Mahdjoub, presents a rectangular dielectric resonator antenna suitable for wideband applications and a band notch of WLAN. The presented results have confirmed the usefulness of the proposed solutions.

In *Product of Three Random Variables and its Application in Relay Telecommunication Systems in the Presence of Multipath Fading*, D. Krstic, P. Nikolic, D. Aleksic, S. Minic, D. Vuckovic and M. Stefanovic consider the product of three random variables. The distribution of the product of independent random variables is very important in many applied problems, as well as in wireless relay telecommunication systems (for example for multiple relay channels).

The next paper, tilted *Enhancement of Ground-to-Aircraft Communication Using Audio Watermarking*, by P. Dymarski, presents the results of research into improving the intelligibility of spoken messages transmitted to aircraft from ground stations. This solution is based on a selective calling system and the audio watermarking technique. It may help improve the comprehension of voice commands transmitted from ground to aircraft using an analog communication link.

In the paper titled *Method for Determining Broadcaster Advised Emergency Wake-up Signal for ISDB-T Digital Television Receivers*, S. Takahashi presents a method for determining a wake-up signal which is used to reduce the rate of false alarms in ISDB-T digital television receivers during their idle phase of operation. The proposed method decreases the number of false alarms, especially for low-mobility users.

The next paper, titled *WannaCry Ransomware: Analysis of Infection, Persistence, Recovery Prevention and Propagation Mechanisms*, by M. Akbanov, V. G. Vassilakis and M. D. Logothetis, presents the results of research concerned with WannaCry Ransomware attacks. Results obtained by the authors may be used for developing relevant detection and defense solutions – both for WannaCry and for other ransomware families that exhibit similar behaviors.

The last paper is titled *Theoretical and Experimental Analysis of Cryptographic Hash Functions*. Its authors – J. Tchórzewski and A. Jakóbik – present a theoretical introduction to the cryptographic hash function theory and a statistical experimental analysis of selected hash functions. Such an analysis facilitates the understanding of the behavior of cryptographic hash functions and may be very helpful in comparing the level of security offered the hashing method selected.

We would like to thank all the authors and reviewers for the effort they have put into preparing this issue of the *Journal of Telecommunications and Information Technology*.

Sławomir Hanczewski
Maciej Piechowiak
Joanna Weissenberg
Guest Editors

https://doi.org/10.26636/jtit.2019.preface1

Paper

ACO-Inspired Energy-Aware Routing Algorithm for Wireless Sensor Networks

Ryo Yamamoto[1,4], Seira Nishibu[2], Taku Yamazaki[3,4], Yasushi Okamura[1], and Yoshiaki Tanaka[2,4]

[1] *Graduate School of Informatics and Engineering, University of Electro-Communications, Chofu-Shi Tokyo, Japan*
[2] *Department of Communications and Computer Engineering, Waseda University, Tokyo, Japan*
[3] *College of Systems Engineering and Science, Shibaura Institute of Technology, Saitama, Japan*
[4] *Global Information and Telecommunication Institute, Waseda University, Tokyo, Japan*

https://doi.org/10.26636/jtit.2019.129718

Abstract—Multi-hop networks, such as WSNs, become an object of increasing attention as an emerging technology which plays an important role for practical IoT applications. These multi-hop networks generally consist of mobile and small terminals with limited resources, which makes them vulnerable to various network status changes. Moreover, the limited nature of terminal resources available, especially in terms of battery capacity, is one of the most important issues to be addressed in order to prolong their operating time. In order to ensure efficient communications in such networks, much research has already been conducted, especially in the field of routing and transmission technologies. However, conventional approaches adopted in the routing field still suffer from the so-called energy hole problem, usually caused by unbalanced communication loads existing due to difficulties in adaptive route management. To address this issue, the present paper proposes a novel routing algorithm that utilizes ACO-inspired routing based on residual energy of terminals. Operational evaluation reveals its potential to ensure balanced energy consumption and to boost network performance.

Keywords—ant colony optimization, load balancing, routing algorithm, sensor networks.

1. Introduction

A wireless sensor network (WSN) generally consists of a number of terminals which have the capability of sensing and communicating. WSN terminals transmit the information collected to a sink, responsible for collecting and processing information, by direct or multi-hop transmission. WSN is thought to be a promising technology for wide-range observation and requires a bunch of sensors to acquire and relay data. WSN terminals are powered by batteries with limited capacity, and powering the network's nodes in a continuous manner is nearly impossible. Moreover, WSNs are intended to operate in the long-term, and smaller batteries are preferred due to manufacturing and deployment costs. In addition, the cost of replacing the batteries significantly increases when terminals are deployed in an environment that cannot be easily accessed by operators, such as deep forests or underwater installations. Therefore, prolonging the lifetime of WSN with limited battery capacity an important issue that needs to be tackled in order extend the network's operating time as much as possible. Therefore, efficient routing and communication technologies are imperative for the achievement of that objective.

A number of routing methods relying on various approaches have been studied with the view of prolonging the lifetime of WSN, such as [1]–[3]. Although all proposals improve efficiency to a certain degree, there is a drawback in the scalability required to increase the physical coverage of the network, because it requires central management for information processing and terminals with specific capability. To address the drawback, autonomous and distributed mechanisms inspired by the behaviors of living organisms, such as insects, are proposed [5]–[8]. They allow to solve various problems with autonomous and distributed optimization procedures, by imitating the behavior of the living organisms. In this paper, we adopt the concept of ant colony optimization (ACO) [5]–[7] to achieve an energy-aware routing mechanism. In the proposed scheme, ACO is utilized not in order to optimize the route, but to dynamically select routes according to the level of the terminals' residual energy, with the transition statuses relied upon for ACO optimization process.

2. Related Work

2.1. Energy-Aware Routing Protocols for WSN

Paper [1] introduces an asymmetric communication approach enabling to save energy. It utilizes the fact that sinks are generally operated by external power supplies. Thus, sinks are capable of conducting longer-range transmissions, compared with terminals powered by batteries. Therefore, terminals with energy constraints adopt the multi-hop communication model, with shorter range communication, to send information to the sink. Then, the sink, as a mains powered device, provides long range communication with the host. In other words, battery operated terminals require less energy compared with the sink. However, it is the placement of the sink that greatly affects its energy saving ability, because energy consumption is determined by the sum of path lengths from the terminals to the sinks.

The routing method introduced in paper [2] utilizes several topologies, depending on network characteristics, to reduce energy consumption. The method adaptively utilizes star-shaped, tree-shaped, chain-shaped and cluster-shaped topologies. In the star-shaped topology, sinks become the center of the star and other terminals use direct transmission to the sinks for reducing the energy required to receive, process and aggregate the data sensed. The tree-shaped topology will be applied to suppress the energy required for transmission by using the multi-hop method. In the chain-shaped topology, the method establishes a single route that reaches every terminal once, and minimizes the route length to improve reliability. In the cluster-shaped topology, the method divides networks into clusters that have 2-hop neighbors at the most, just as conventional clustering in WNS does. Then, the cluster head aggregates the received information and sends it to sinks to suppress the total amount of send and receive data and the transmission distance. However, environment-related changes caused by joining and leaving of terminals or by other factors forces the method to recalculate the optimal topology and the cost increment that is proportional to the network's size becomes an inevitable issue.

Optimized LEACH-C [3] also adopts cluster-based routing that estimates required energy consumption based on the terminals' location and the number of cluster members of a sink. Optimized LEACH-C utilizes the estimated energy consumption to generate an initial solution and uses the simulated annealing to generate heuristic solutions. Then, the solution is notified to each terminal and clusters will be assigned to terminals entirely. However, in optimized LEACH-C, sinks must play the role of collecting information, performing clustering calculations and notifying the results, which increases the operational costs.

2.2. Ant Colony Optimization

The issues described in Subsection 2.1 may be solved by network-wide optimization which is accomplished by autonomous and distributed state prehension and a decision made by an individual terminal, i.e. by the so-called divide-and-conquer method. The swarm intelligence strategy may serve as an example of such an approach, as it is inspired by the group behavior of insects. Their simple individual behaviors optimize objectives entirely. There are methods that apply the optimization mechanism to manage the behavior of terminals acting as elements of swarms, such as [4].

Ant colony optimization (ACO), inspired by the feeding behaviors of ants, is proposed as one particular application [5]–[7]. ACO generally utilizes agents, called "ants", that secrete "pheromone" to the traveled route, serving as an evaluation value of the route, for adaptive and continuous route updating. Therefore, application of ACO in such an environment as WSN, where the communications conditions change over short periods of time and mutual state prehension by the terminals is difficult, allows to achieve effective performance.

ACO has an ability to discover the shortest route without an effort of centralized management by utilizing the behaviors of ants and the secretion of pheromones, as described previously. Thus, ACO is applied in various combinatorial issues, such as the traveling salesman problem (TSP). A feeding ant detects pheromones on the ground, follows them towards the food source and then returns to the nest with the food, secreting pheromones. As the secreted pheromones volatilize at a constant pace, more pheromones are present along shorter, rather than longer routes. A route with more pheromones attracts more ants and pheromone secretion in regions adjacent to the shortest route becomes active, i.e. ants tend to select the shortest route, as the time passes, as shown in Fig. 1.

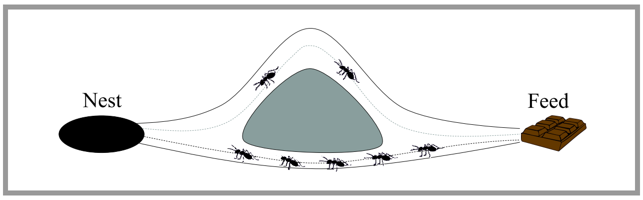


Fig. 1. The principle of ACO routing.

Papers [6], [7] proposed a basic ACO model, called the ant system (AS). Here, we will explain AS with TSP, which is applied, in particular, to combinatorial problems. In the specific application, each ant is treated as m agents and placed in n cities, and creates a route based on the rule that each agent visits each city only once and decides the next city to be visited based on the pheromone level. Equation (1) calculates the probability that agent k in city i on cycle t travels to city j in the next cycle:

$$p_{ij}^{k}(t) = \begin{cases} \dfrac{[\tau_{ij}(t)]^{\alpha}\,[\eta_{ij}]^{\beta}}{\sum\limits_{s \in J_k(i)} [\tau_{is}(t)]^{\alpha}\,[\eta_{is}]^{\beta}} & j \in J_k(i) \\ 0 & \text{otherwise} \end{cases}, \qquad (1)$$

where $\tau_{ij}(t)$ represents the pheromone level between city i and j on cycle t, η_{ij} represents the invert of route length between city i and j, $J_k(i)$ represents the set of visiting cities of agent k in city i, and α, β are the constant.

After the agent finishes the trip upon visiting each city and after the route has been created, AS calculates the pheromone level to be secreted along the traveled route with the following Eq. (2):

$$\Delta\tau_{ij}^{k} = \begin{cases} \dfrac{1}{C_k} & (i,j) \in T_k \\ 0 & \text{otherwise} \end{cases}, \qquad (2)$$

where τ_{ij}^{k} represents the pheromone level to be secreted between city i and j and C_k represents the length of route T_k between city i and city j that agent k created. Then, AS applies the calculated pheromone level to the route and update the residual pheromone level using Eq. (3):

$$\tau_{ij}(t+1) = \rho\,\tau_{ij}(t) + \sum_{k=1}^{m} \Delta\tau_{ij}^{k}\ , \tag{3}$$

where ρ represents the volatile coefficient, i.e. AS volatilizes a certain pheromone level from the remaining pheromones and adds the pheromones secreted by agents. AS continuously repeats this procedure until it discovers the optimal solution.

AS enables ACO-based routing, by relying on simple procedures performed by individual terminals, to find the optimal route without using centralized network management. In addition, as the data travels along the optimal route, reliability improves and energy consumption per packet becomes lower. Although the mechanism identifies and utilizes a route that is most efficient in terms of network performance, concentration of traffic along specific routes may cause an early drop out of terminals due to the exhaustion of batteries.

3. ACO-Inspired Energy-Aware Routing

Here we propose an ACO-inspired energy-aware routing algorithm, named AERO, based on the residual energy of terminals and relied upon for adaptive and dynamic routing. The significant characteristic of AERO is that the agent ant behavior tries not to find the optimal solution, but strives to identify semi-optimal solutions. This prevents the routes with a sufficient pheromone level from being utilized on a continuous manner, until the terminals along the route exhaust their batteries, that is until AERO positively utilizes the transient state of ACO to improve route diversity.

AERO introduces three types of ant imitating control packets to apply ACO while routing, namely forward ant (F-ANT), backward ant (B-ANT) and data ant (D-ANT). In addition, AERO does not secrete pheromones into links between terminals, as the conventional ACO does, but into terminals. The secreted and residual pheromone levels are notified to neighboring terminals with periodical hello message exchanges, just as in the case of conventional routing. A brief description of the routing procedure is presented below.

In AERO, a source terminal first sends F-ANTs towards the desired destination in the same way as conventional routing protocols do, as shown in Fig. 2. The F-ANTs sent by the source terminal travel along various routes and F-ANTs store the terminal ID and the residual energy of each intermediate terminal during the travel. The destination terminal that receives the F-ANTs waits for other F-ANTs, for a predetermined period of time, to collect information about multiple routes.

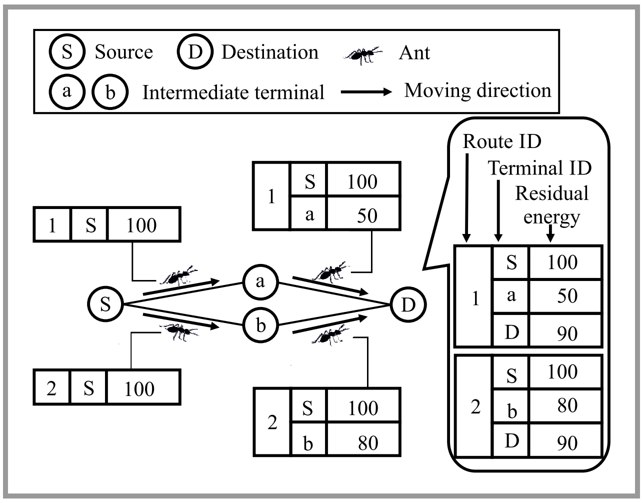


Fig. 2. Forward ant.

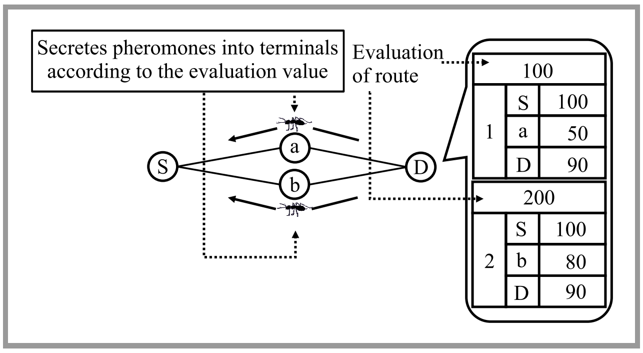

Secretes pheromones into terminals according to the evaluation value
Evaluation of route
100
1 S 100 a 50 D 90
200
2 S 100 b 80 D 90
S a b D

Fig. 3. Backward ant.

After the predetermined waiting time elapses, the destination terminal that received multiple F-ANTs evaluates each route using the information stored in the F-ANTs. It needs to be noted that the detailed evaluation procedure will be explained later. The destination terminal generates B-ANTs that contain information about the route and its evaluation value after the evaluation procedure is completed. Then, B-ANTs start their travel by tracing back along the route that F-ANTs traveled, and B-ANTs secrete pheromones to the intermediate terminals along the route during the travel (Fig. 3). This is recursively performed until the B-ANTs reach the source terminal. Note that the source terminal also waits for other B-ANTs, over a predetermined period of time, to receive multiple B-ANTs, just as it was the case with F-ANTs.

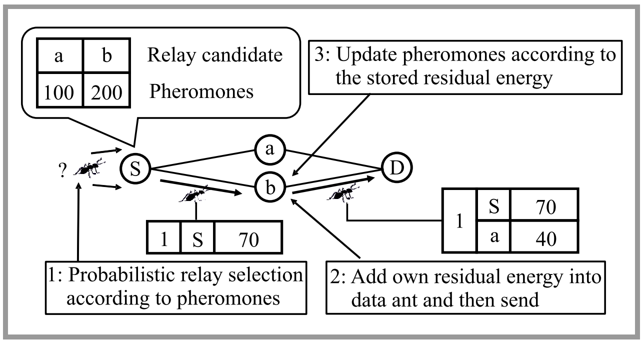

a b Relay candidate
100 200 Pheromones
3: Update pheromones according to the stored residual energy
?
S a b D
1 S 70
1 S 70 a 40
1: Probabilistic relay selection according to pheromones
2: Add own residual energy into data ant and then send

Fig. 4. Data ant scheme.

After the arrival of B-ANTs at the source terminal, it starts the forwarding procedure for data encapsulated by D-ANT (Fig. 4). Senders of D-ANT, namely source and intermediate terminals, select the next hop terminal probabilistically, according to the pheromone level at the candidate receivers. Once the sender determines the receiver, it records its own residual energy to the D-ANT, and the D-ANT travels to the receiver. The receiver selection procedure will be explained in detailed later in this section. The receiver that the D-ANT reaches then updates own pheromones according to the information stored in the D-ANT. By repeating the above scheme recursively, AERO updates the pheromone levels on intermediate terminals and the data encapsulated by the D-ANTs reaches the destination node.

3.1. Route Evaluation and Pheromone Update

The pheromone level on each terminal is calculated at the destination terminal by means of two evaluation values, with the use of the collected route information and the terminal information stored in D-ANT.

We will describe the evaluation values as $H_{A,i}$ and $H_{B,i}$. It needs to be noted that the pheromone level in AERO will always be positive, and that AERO assigns upper and lower limits to that value.

Evaluations at destination terminals. The destination terminals calculate the evaluation values for each route using the information obtained by F-ANTs. In this procedure, AERO first calculates the average residual energy $E_{sd,i}$ of each terminal along route i whose source and destination terminals are s and d, by:

$$E_{sd,i} = \frac{\sum\limits_{j \in n_{sd,i}} e_{ij}}{\left|n_{sd,i}\right|} , \tag{4}$$

where e_{ij} represents the residual energy of terminal j along route i, $n_{sd,i}$ represents the set of terminals along route i. Then, the destination terminal calculates the average residual energy of complete routes using the result of Eq. (4) and:

$$E_{sd} = \frac{\sum\limits_{i \in r_{sd}} E_{sd,i}}{\left|r_{sd}\right|} , \tag{5}$$

where r_{sd} represents the route set obtained by F-ANTs. Next, the destination terminal calculates the evaluation value $H_{A,i}$ as:

$$H_{A,i} = (1-\beta)\frac{E_{sd,i}}{E_{sd,\max}} + \beta\left(\frac{\sum\limits_{j \in n_{sd}, e_{ij} \le E_{sd}} (e_{ij} - E_{sd})}{\left|e_{i,\text{low}}\right| E_{sd}} + 1\right) , \tag{6}$$

where $E_{sd,\max}$ represents the maximum average residual energy along the route set r_{sd}, $e_{i,\text{low}}$ represents the number of terminals along route i whose residual energy is lower than E_{sd}, and β is a constant.

The first member of Eq. (6) becomes closer to 1 when the residual energy of terminals composing route i is high. The second member of Eq. (6) gets closer to 1 when the variance between the residual energy levels of terminals along route i is low. The evaluation value $H_{A,i}$ will be stored in B-ANTs, and the intermediate terminals that the B-ANTs travel along update their pheromones by adding the evaluation value to the current pheromone level.

In addition to the above, AERO takes hop counts into account to calculate the overall evaluation H_i. AERO evaluates the hop count of each route and calculates $H_{B,i}$ as:

$$H_{B,i} = \frac{h_i - (1+\alpha)h_{sd}}{(1+\alpha)h_{sd}} , \tag{7}$$

where h_i represents the hop-count of route i, h_{sd} represents the average hop-count of all routes from source s to destination d, and α represents the acceptable route length increment ratio. With $H_{A,i}$ and $H_{B,i}$, AERO calculates the overall evaluation using the weight parameter γ as:

$$H_i = (1-\gamma)H_{A,i} + \gamma H_{B,i} \ . \tag{8}$$

Pheromone update with data ant. D-ANTs record the residual energy of the terminals along the route and intermediate terminals update their pheromones using the evaluation value calculated with the use of the information stored. The evaluation value for D-ANTs $H_{C,j}$ for intermediate terminal j will be calculated by:

$$H_{C,j} = \frac{e_j - E_{sj}}{E_{sj}} , \tag{9}$$

where E_{sj} represents the average residual energy of intermediate terminals after the source terminal s, and e_j represents the residual energy of terminal j at which the D-ANT is currently staying. The evaluation value $H_{C,j}$ becomes positive when the residual energy of the current terminals is higher than the average residual energy, and becomes negative when the latter value is lower. Afterward, the terminal adds $H_{C,j}$ to its own pheromone level, in order to increase or decrease the pheromone level:

$$P_i(t+\Delta t) = (1-\rho)P_i + H_{C,j} , \tag{10}$$

$$\rho = \theta\Delta t , \tag{11}$$

where Δt represents the time gap between the current time and the last update time, ρ represents volatilization rate, and θ represents a fixed parameter to determine the rate ρ.

3.2. Route Selection

In AERO, route selection is done by the probabilistic way based on the pheromone level of each terminal. Each terminal first confirms the set of candidate intermediate terminals for sending data towards the destination, before D-ANTs travel to other nodes. If there is only one candidate in the set, D-ANTs just start their travel towards the terminal. If there are multiple candidates, sender m calculates the probability of D-ANTs' travel towards the next intermediate terminal n:

$$Q_{mn} = \frac{P_n}{\sum\limits_{i \in N_m^d} p_i}, \tag{12}$$

where Q_{mn} represents the probability that terminal m selects terminal n as the next hop, P_n represents the pheromone level in n, and N_m^d represents the set of candidate intermediate terminals for F-ANTs, leading towards destination d from m. By relying on the probabilistic intermediate terminal selection procedure described above, AERO assigns a higher priority to the node with a higher pheromone level and data encapsulated by D-ANTs travel towards the destination terminal.

4. Performance Evaluation

4.1. Simulation Setup

Computer simulations have been conducted to evaluate the effectiveness of AERO compared to conventional routings, using the QualNet [9] network simulator. In the simulations, we adopted AODV [10], optimized LEACH-C [3], and AS [6], [7] for the routing to be compared. Two scenarios were used to evaluate the performance from the viewpoint of communication qualities and network lifetime. The first evaluates network performance by changing terminal densities that greatly affect the routing results. The second evaluates network lifetime by observing the number of active terminals over time. The common parameters for the simulations are shown in Table 1.

Table 1
Simulation parameters

Parameter	Value
Routing methods	AODV, optimized LEACH-C, AS
Simulation duration	1000 s
Simulation area	1000 × 1000 m
The number of terminals	100–400
The number of sinks	2–10
Wireless medium	IEEE 802.11b
Bandwidth	11 Mbps
Communication radius	150 m
Terminal placement	Random
The number of sessions	50 sessions
Source terminals	Randomly chosen
Packet generation interval	100 ms
Packet size	1000 bytes
Battery capacity	18,000 mAs
Power consumption for sending	840 mAs
Power consumption for receiving	800 mAs

In the simulations, terminals are randomly placed in the square area of 1000×1000 m, and communicate with each other using IEEE 802.11b with the radius of 150 m at the most. Source terminals and the number of packets to be transmitted are randomly chosen, and every packet with the size of 1000 bytes is transmitted every 100 ms. In this paper, we have conducted two simulations by changing terminal density and sink density. The number of terminals is changed from 100 to 400 and the number of sinks from 2 to 10 with 400 nodes.

4.2. Network Performance Evaluation

We evaluate the impact that terminal or sink density has on communication performance by relying on successful delivery rate and end-to-end delay. The successful delivery rate is calculated by dividing the number of received packets by the number of packets generated in terminals. The end-to-end delay indicates the time gap between the initiation time of packet transmission and the time that the destination sink receives the packet.

4.3. Network Lifetime Evaluation

In this simulation, we evaluate the number of active terminals every 25 s to show the efficiency of each routing method. We defined the active terminal as the terminal with the battery level of 40% of the initial capacity. We firstly conducted simulations with 100 and 200 terminals to evaluate the performance in an environment that is tough for the routing methods since the available route diversity is limited to a certain degree. In addition to the aforementioned simulations, we conducted simulations using 6 or 10 sinks with 400 terminals. It is obvious that with the higher number of singles, the path diversity increases and balances traffic load and energy use. However, the improvement in traffic load performance and energy consumption, generally derives from how the routing protocols select or manage routes. Therefore, the simulations reveal the balancing performance from a different point of view.

4.4. Simulation Results

Figures 5–8 show the impact of terminal density on communication performance. In the results, we exclude abnormal outcomes caused by unclosed sessions. Moreover, the values of top and bottom 5%, such as the outlier in the calculation of end-to-end delay, were excluded as well.

Figures 5–7 show the successful delivery rate results. The result indicates that the proposed method could achieve a successful delivery rate of nearly 90%, regardless of terminal and sink density. This is due mainly to the adaptive and dynamic route management of AERO, which effectively suppresses unnecessary route reestablishment by avoiding energy exhaustion of terminals caused by the exhaustion of batteries. The conventional AODV decreases its

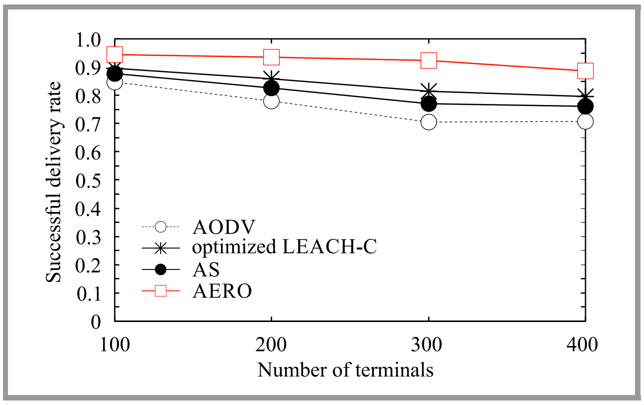


Fig. 5. Simulation of successful delivery rate versus number of terminals.

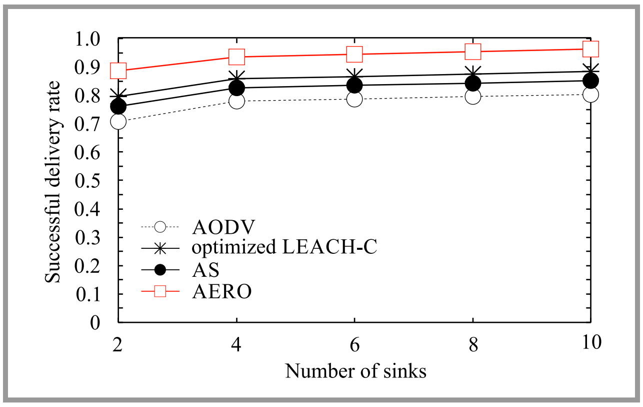


Fig. 7. Simulation of successful delivery rate as a function of number of sinks.

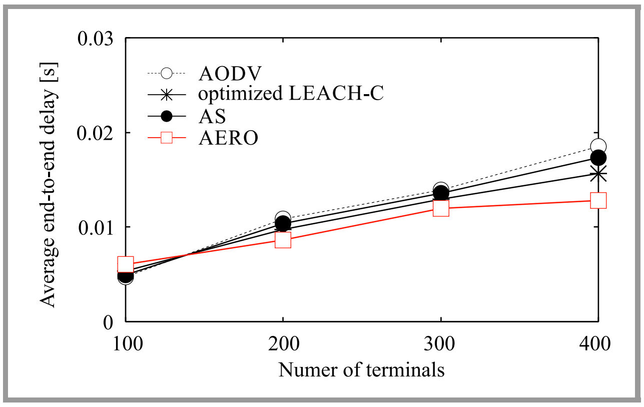


Fig. 6. Average end-to-end delay versus number of terminals.

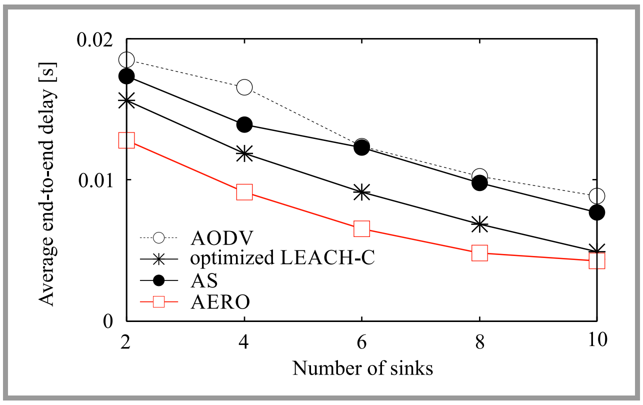


Fig. 8. Average end-to-end delay as a function of number of sinks.

reliability since the routing procedure is basically aimed to establish a single end-to-end route from a source terminal to the destination terminal based on a route length. Moreover, the route length only takes hop-counts into a consideration, and other parameters such as residual energy and reliability are not the metrics for evaluating route quality. Thus, the route established by AODV could not achieve better route quality except for route length. Optimized LEACH-C could achieve better routing performance due to its complex and centralized comprehensive route management, since it can comprehend the states of the entire network and is capable of deriving entirely optimal solutions. Pure AS could achieve a certain degree of improvement compared to AODV, since AS can take other metrics into account, such as pheromone level. However, improvement is limited because pure AS ceases optimization once the optimal solutions are found, and further optimization will be suspended until another route request comes in.

In addition, a common feature could be observed. Namely, the success rate of routing protocols gradually improves as the number of sinks increases. The reason of this is obvious: routes established within the network were autonomously distributed, since the overlapping link usage is autonomously eliminated to a certain degree. However, AERO could achieve a higher rate since the aforementioned characteristics were capable of increasing the base performance of AERO to the higher degree than in the remaining cases.

Figures 6 and 8 show the results of end-to-end delay for each of the routing methods. The results show that each protocol gradually increases the delay as the terminal density increases, whereas the delay is decreased as the sink density grows. The reason for the delay increase mainly derives from the increase in overall traffic within the networks, which will be a cause of a higher queuing delay and interference in transmission to other terminals. Although the delay increase is inevitable, AERO could suppress this type of degradation by means of its adaptive route management and could decrease the delay compared to the other methods. In other words, the probabilistic intermediate terminal selection by D-ANTs could efficiently select the intermediate terminals with less traffic load. On the other hand, other single route-based routing methods degrade the performance compared with AERO, since their routing procedures only show an advantage in terms of route establishment. The decrease of delay, observed as the sink density increases, can also be explained with the same characteristic as described in the explanation of the success rate. In other words, the increase in sink density autonomously balances traffic load without a systematic procedure. In addition to that, we could observe that the base performance is also affected as the density increases.

4.5. Network Lifetime Evaluation

Figures 9–12 show the transition of the active terminal ratio versus elapsed time. The result shows that AERO could reasonably reduce the number of inactive terminals compared with other routing methods. Moreover, the decrease observed in AERO seems to be linear, whereas the decrease typical of other methods seems to be an inversely proportional or exponential. The main reason for the difference can be explained by the routing strategy, as AERO relies on the principle of dynamic and adaptive intermediate selection, whereas other approaches adopt the one-time optimization principle. Another characteristic trend may be identified as the simulation time elapses, namely the rate at which the number decreases is more gentle in the case of conventional routing methods. This can be explained by the manner in which intermediate terminals are selected by the individual methods, since they attempt to utilize the optimal terminals for end-to-end routes and such devices must transmit more packets than others. Thus, the optimal terminals exhaust their batteries and become inactive sooner than other non-optimal nodes. After the rapid exhaustion phase, the methods must select the rest of the terminals as intermediate devices and the selection procedure may autonomously balance traffic loads.

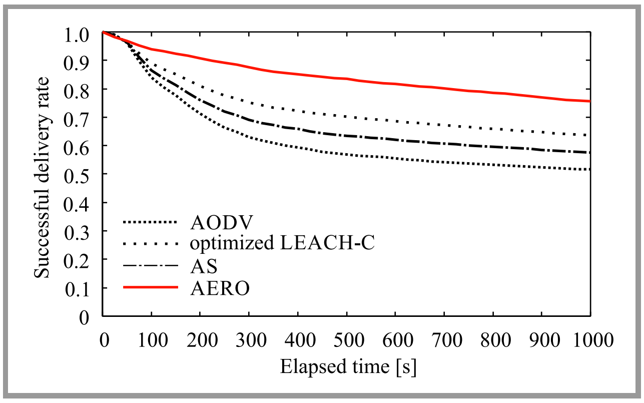


Fig. 9. Active terminal ratio (100 terminals) vs. time.

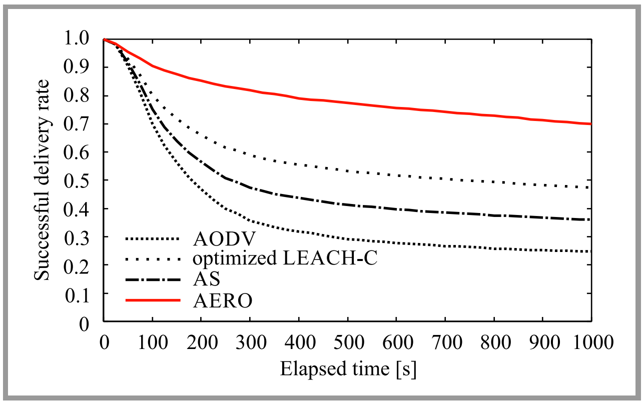


Fig. 10. Simulation results on active terminal ratio (200 terminals) vs. time.

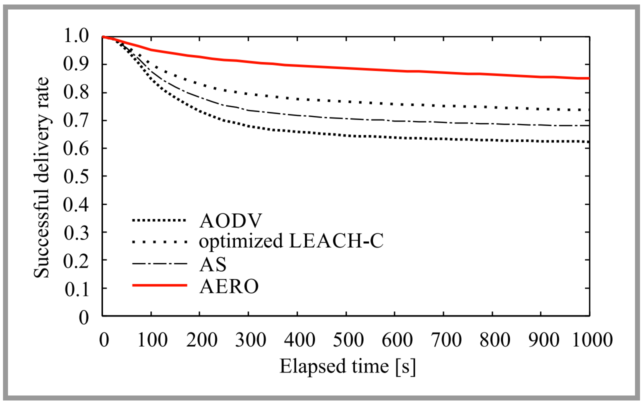


Fig. 11. Active terminal ratio value (6 sinks) vs. simulation time.

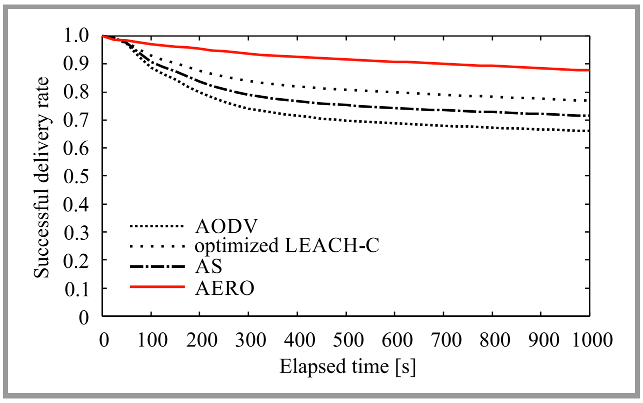


Fig. 12. Results on active terminal ratio (10 sinks) vs. simulation time.

4.6. Summary of the Simulations

Through the simulations conducted above, we confirmed that the proposed AERO approach may extend the lifetime of a network while maintaining its reasonable performance. The major contribution of the proposed solution derived primarily from its adaptive and dynamic route and intermediate terminal selection principle, which utilizes the transient state of ant-colony optimization. Moreover, the unique characteristic consisting in the fact that AERO secretes pheromones not to links, but to terminals, enables adaptive and dynamic intermediate terminal selection.

5. Conclusions

This paper proposes an ACO-inspired routing strategy, known as AERO, for WSNs, enabling to balance traffic loads by utilizing transient behaviors for optimization. Performance evaluation reveals that the AERO approach proposed may achieve improved routing efficiency compared with other existing routing methods. In other words, AERO requires less transmission effort to send the same amount of data and improves energy efficiency.

Although the improvement achieved by AERO contributes to prolonging the lifetime of WSNs, there is still room for further improvement, since AERO currently does not take into account terminal statuses, such as their awake and sleep modes. Moreover, such issues as refining the procedure relied upon to calculate the evaluation values, as well as assessment of performance with the use of realistic models may also be addressed in the future.

Acknowledgments

This work was supported by JSPS KAKENHI grant number 15K15978 and JP17K12680.

References

[1] J. Neander, E. Hansen, M. Nolin, and M. Bjorkman, "Asymmetric multihop communication in large sensor networks", in *Proc. 1st Int. Symp. Wireless Pervasive Comput. ISWPC 2006*, Phuket, Thailand, 2006 (doi: 10.1109/ISWPC.2006.1613561).

[2] H. Taka, H. Uehara, and T. Ohira, "Energy-efficiency of sensor networks in terms of network topology", *IEICE Trans. Commun.*, vol. J96-B, no. 7, pp. 680–689, 2013 [in Japanese].

[3] S. Shi, X. Liu, and X. Gu, "An energy-efficiency optimized LEACH-C for wireless sensor networks", in *Proc. 7th Int. Conf. on Commun. and Network. in China ChinaCom 2012*, Kun Ming, China, 2012, pp. 487–492 (doi: 10.1109/ChinaCom.2012.6417532).

[4] K. Bennani and D. E. Ghanami, "Particle swarm optimization based clustering in wireless sensor networks: The effectiveness of distance altering", in *Proc. IEEE Int. Conf. on Complex Syst. ICCS 2012*, Agadir, Morocco, 2012 (doi: 10.1109/ICoCS.2012.6458564).

[5] S. Tsutsui, "ACO: Ant colony optimization", *Syst., Control and Inform.*, vol. 52, no. 10, pp. 390–398, 2008 (doi: 10.11509/isciesci.52.10_390) [in Japanese].

[6] M. Dorigo, V. Maniezzo, and A. Colorni, "Ant system: Optimization by a colony of cooperating agents", *IEEE Trans. on Syst., Man, and Cybernet.*, Part B (Cybernetics), vol. 26, no. 1, pp. 29–41, 1996 (doi: 10.1109/3477.484436).

[7] M. Dorigo and L. M. Gambardella, "Ant colony system: A cooperative learning approach to the traveling salesman problem", *IEEE Trans. on Evolut. Computat.*, vol. 1, no. 1, pp. 53–66, 1997 (doi: 10.1109/4235.585892).

[8] N. Wakamiya and M. Murata, "Biologically-inspired information network technologies", *IEICE Trans. on Commun.*, vol. J89-B, no. 3, pp. 316–323, 2006 [in Japanese].

[9] "QualNet Network Simulator Software", Scalable Network Technologies Inc., Aug. 2017 [Online]. Available: http://web.scalable-networks.com/content/qualnet/

[10] C. E. Perkins and E. M. Royer, "Ad-hoc on-demand distance vector routing", in *Proc. 2nd IEEE Worksh. on Mobile Comput. Syst. and Appl. WMCSA 1999*, New Orleans, LA, USA, 1999, pp. 90–100 (doi: 10.1109/MCSA.1999.749281).

Ryo Yamamoto received his B.Eng. and M.Eng. degrees in Electronic Information Systems from Shibaura Institute of Technology, Tokyo, Japan, in 2007 and 2009. He received D.Sc. in global telecommunication studies from Waseda University, Tokyo, Japan, in 2013. He was a research associate at Graduate School of Global Information and Telecommunication Studies, Waseda University, from 2010 to 2014, and has been engaged in researching in wireless communication networks. He is presently an Assistant Professor at Graduate School of Informatics and Engineering, The University of Electro-Communications. He received the IEICE young researcher's award in 2010, the IEICE Network System Research Award in 2014, the CANDAR/ASON Best Paper Award in 2014, IEICE Communications Society Distinguished Contributions Award in 2017, IEICE Information and Communication Management Distinguished Contributions Award in 2014 and 2017. His current research interests are ad hoc networks, sensor networks, IoT/M2M networks, and network protocols for the networks. He is a member of IEICE and IEEE.

E-mail: ryo-yamamoto@uec.ac.jp
Graduate School of Informatics and Engineering
The University of Electro-Communications
1-5-1 Chofugaoka
Chofu-Shi Tokyo, 182-8585 Japan

Global Information and Telecommunication Institute
Waseda University
3-4-1 Okubo
Shinjuku-ku, Tokyo, 169-8555 Japan

Seira Nishibu received the B.Eng. degree in Communication and Computer Engineering from Waseda University, Tokyo, Japan, in 2018. She is presently an employee of The Kansai Electric Power Co., Japan.

Department of Communications and Computer Engineering
Waseda University
3-4-1 Okubo
Shinjuku-ku, Tokyo, 169-8555 Japan

Taku Yamazaki received the B.E. and M.Sc. degrees in Electronic Information Systems from Shibaura Institute of Technology, Tokyo, Japan, in 2012 and 2014, respectively. He received the D.E. degree in Computer Science and Communications Engineering from Waseda University, Tokyo, Japan, in 2017. He was a research associate at Department of Communications and Computer Engineering, Waseda University, from 2015 to 2018. He is presently an Assistant Professor at Department of Electronic Information Systems, Shibaura Institute of Technology, Saitama, Japan. He is also presently an Adjunct Researcher at Global Information and Telecommunication Institute, Waseda University, Tokyo, Japan. His research in-

terests are ad hoc networks, sensor networks, and Internet of Things.
E-mail: taku@shibaura-it.ac.jp
College of Systems Engineering and Science
Shibaura Institute of Technology
307 Fukasaku, Minuma-ku
Saitama-shi, Saitama, Japan

Global Information and Telecommunication Institute
Waseda University
3-4-1 Okubo
Shinjuku-ku, Tokyo, 169-8555 Japan

Yasushi Okamura received the B.Eng. degree in Informatics and Engineering from The University of Electro-Communications, Tokyo, Japan, in 2018. He is presently a master course student at Graduate School of Informatics and Engineering, the University of Electro-Communications. His research interests are ad hoc networks, sensor networks, and Internet of Things.
E-mail: okamura@net.lab.uec.ac.jp
Graduate School of Informatics and Engineering
The University of Electro-Communications
1-5-1 Chofugaoka
Chofu-Shi Tokyo, 182-8585 Japan

Yoshiaki Tanaka received the B.Eng., M.Eng., and D.E. degrees in Electrical Engineering from the University of Tokyo, Tokyo, Japan, in 1974, 1976, and 1979, respectively. He became a staff at Department of Electrical Engineering, the University of Tokyo, in 1979, and has been engaged in teaching and researching in the fields of telecommunication networks, switching systems, and network security. He is presently a professor at Department of Communications and Computer Engineering, Waseda University. He received the IEICE Achievement Awards in 1980, the Okawa Publication Prize in 1994, the Commendation by Minister for Internal Affairs and Communications in 2009, the IEICE Distinguished Achievement and Contributions Awards in 2013. He is an Honorary Member of IEICE.
E-mail: ytanaka@waseda.jp
Department of Communications and Computer Engineering
Waseda University
3-4-1 Okubo Shinjuku-ku, Tokyo, 169-8555 Japan

Global Information and Telecommunication Institute
Waseda University
3-4-1 Okubo
Shinjuku-ku, Tokyo, 169-8555 Japan

Paper

Adaptive Load Balancing Ad Hoc Routing Scheme Inspired by True Slime Mold

Hiroshi Katada, Taku Yamazaki, and Takumi Miyoshi

Shibaura Institute of Technology, Saitama, Japan

https://doi.org/10.26636/jtit.2019.129518

Abstract—Engineering neo-biomimetics, i.e. imitation models based on body structures and behavior of living organisms, relied upon to solve complex problems, have been studied in various fields. In distributed networks, such as ad-hoc networks and wireless sensor networks, the behavior of a variety of true slime molds which are capable of constructing multipath flow networks based on the amount of body, has been studied. Ad hoc networks only consist of mobile terminals (nodes) that can relay packets along an established route. However, link relations and the available bandwidth of the nodes change dynamically due to the mobility of nodes. In addition, the speed of communication between nodes also varies due to node positions and their communication-related quality. Thus, practical use of ad-hoc networks still remains an issue, because it is difficult to establish stable routes under such environments. This study aims to propose an adaptive load balancing routing technique that adaptively diversifies the transmission paths based on the available bandwidth, residual battery life, and the data transmission volume, by applying a mathematical model of slime mold routing, known as the physarum solver. We confirm the effectiveness of its adaptive behavior in dynamic environments using computer simulations.

Keywords—adaptive ad hoc routing, ad hoc network, engineering neo biomimetic, physarum solver, true slime mold.

1. Introduction

Life forms change and optimize their structures and behavioral patterns in the course of evolution. Recently, biomimetic technologies used to design artifacts inspired by specific abilities and structures have been studied [1]. Primary examples of the application of biomimetic technologies include swimsuits inspired by the skin structure of sharks [2], and nylon fiber inspired by the fiber structure of cotton [3]. In the field of network research, various bio-inspired mechanisms, such as multiple route optimization inspired by the feeding behavior of physarum [4], shortest-path route optimization inspired by the feeding behavior of ants [5], and a synchronization mechanism inspired by the synchronous behavior of fireflies, have also been studied [6]. The features of *Physarum Polycephalum*, which is a variety of physarum, are applied to design a routing protocol. Please note that *Physarum Polycephalum* is simply referred to as physarum below. As far as the nature of physarum's feeding process is concerned, it has been confirmed that physarum creates a tube for nutritional transport, utilizes its own body and connects foods using the tube when it finds multiple baits at different places. It also has been confirmed that the number of constructed paths varies depending on the amount of liquid that is the constituent of physarum. Additionally, physarum has a negative phototaxis, which means that physarum can limit the area over which to spread its body [7]. Therefore, physarum may optimize trade-offs between efficiency and stability of nutritional transport paths, which means that is characteristics may be applied to the selection of a relay node in wireless multi-hop networks. Those characteristics may be applied, for instance, in an ad hoc network that consists of mobile wireless terminals (nodes) only, without relying on a base station [8]. The ad hoc network finds exceptional use in an area where a base station cannot be placed or has been destroyed due to a disaster. In ad hoc networks, nodes can communicate with each other by relaying packets. Such multi-hop communication may be performed when relay nodes exist, even if they are located outside the communication range. However, due to such factors as node mobility, low battery states, radio interference, dynamic topology changes, and network stability, degradations may occur. One solution includes the use of a multipath routing protocol that can alleviate the impact of the dynamic changes by simultaneously using multiple paths.

In addition, in recent wireless communication media, such as IEEE 802.11 [9], [10], offer a functionality that may select the appropriate transmission speed between nodes [11]. Some research concerning the rate adaptation algorithm has been conducted as well [12]–[14]. Especially, multiple input and multiple output (MIMO) systems [15] which drastically improve the transmission speed have been proposed by using multiple transmit and receive antennas. As a result, the speed of communication between nodes varies widely due to the variation in communication quality caused by node mobility, radio interference and other factors experienced in real life environments. Hence, the effect of varying communication speed is approached adaptively as well.

In this study, we propose a multipath ad hoc routing method by applying a mathematical model inspired by the path finding ability typical of physarum, known as the physarum solver (PS) [16]. By applying PS, the proposed method constructs multiple paths and adaptively allocates bandwidth to such paths based on the data transmission volume, bandwidth available within each link and the residual battery level of each node.

2. Related Work

Physarum consists of a stretchable tube and viscous liquid flowing through it. The tube is a path connecting multiple baits, and becomes thinner or thicker in response to flow rate fluctuations. In addition, the tube has an upper limit to its thickness. Figure 1 shows an example of a route constructed by physarum in a maze. First, physarum spreads its tube over the entire maze. Next, it selects paths where food has been located. Then, the tube gradually becomes thinner as the liquid flow rate decreases, when the path becomes relatively longer than the remaining paths. In addition, since no flow takes place in a blind tube, the tube degenerates relatively quickly. Therefore, physarum prioritizes shorter and continuous paths, and the number of remaining paths varies based on the total amount of liquid constituting physarum. In addition to the above properties, physarum moves away from sources of light (negative phototaxis), and therefore the tube becomes thinner along the segment exposed to light. If the path is disrupted by an external factor, physarum takes a detour path to reallocate the flow of the disrupted path to other paths to avoid path disruption of connectivity between the baits. Therefore, physarum may adjust the efficiency and stability of nutritional transport paths.

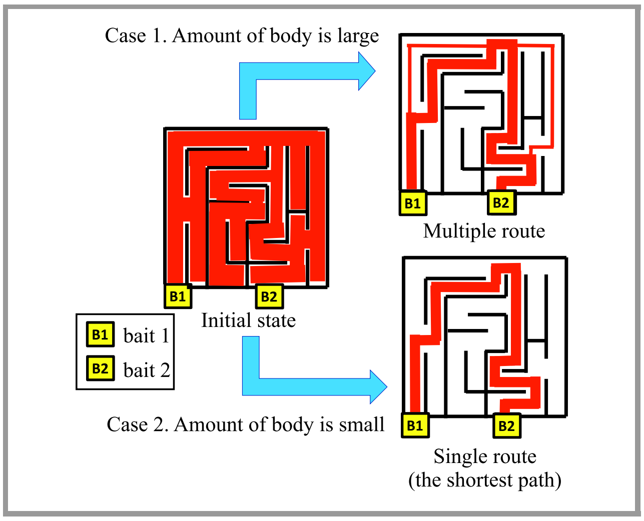


Fig. 1. Behavior of slime mold in the maze.

An experiment [17], in which the behavior of physarum on a railway network was observed, has been conducted. In the experiment, an agar medium imitating the railway network in the Kanto region was prepared, and bait was placed at positions that correspond to major cities in the region. Then, physarum was placed in the position that representing Tokyo, and the behavior of physarum was observed. In the experiment, the light quantity is adjusted according to the elevation and position of the river to imitate the topography of the Kanto region. In the experimental results, physarum spreads in approximately half a day, and it becomes clear that the paths similar to the current railway network are constructed between baits. This result has revealed that the efficiency and stability of transport paths designed by humans and physarum are similar.

2.1. Physarum Solver

Physarum solver (PS) [16] is a mathematical model which focuses on the feeding behavior of physarum when physarum constructs a route between baits. PS calculates the flow rate of each tube based on the total amount of liquid in physarum, as well as on the length, thickness and pressure loss of each tube. In the initial state, PS calculates the pressure loss in each tube based on the total amount of liquid, length and thickness of the tube. Thereafter, the flow rate of each tube is temporarily determined in accordance with the pressure loss of the tube, and then the flow rate of the tube changes as the thickness varies. By iterating such a process, the flow rate and thickness of the tube converge to an appropriate value, and finally the tubes that remain are based on the total amount of liquid. Furthermore, by setting the parameters of phototaxis, it is possible to control the thickness of a tube independently of the flow rate.

The functioning principle of PS, with baits placed at two locations, is shown below. Here, PS deals with tubes that have branch points i and j at both ends.

1. PS calculates the pressure loss of the tube between i and j based on the flow rate of the tube, as calculated by Eq. (1), and based on the flow conservation law of the tube, calculated by Eq. (2):

$$Q_{ij}(t) = \frac{D_{ij}(t)}{L_{ij}}\left[p_i(t) - p_j(t)\right] , \tag{1}$$

$$\sum_i Q_{ij} = \begin{cases} -Q_{\text{all}}\,, & j = \text{bait } 1 \\ Q_{\text{all}}\,, & j = \text{bait } 2 \\ 0 \quad , & \text{otherwise} \end{cases} , \tag{2}$$

   where Q_{all} is the total amount of liquid; $D_{ij}(t)$ and L_{ij} are the thickness and length of the tube, respectively; $p_i(t) - p_j(t)$ is the pressure loss of the tube.

2. By substituting the derived pressure loss into Eq. (1), the flow rate of each tube $Q_{ij}(t)$ is temporarily determined, and thereby a shorter and thicker tube allows a larger flow rate. However, due to pressure loss in the blind tube, the flow rate also becomes low.

3. By substituting the flow rate of each tube in Eq. (3), PS updates the tube thickness:

$$D_{ij}(t+\delta t) = D_{ij}(t) + \delta t\left\{f(|Q_{ij}(t)|) - aD_{ij}(t)\right\}, \tag{3}$$

Here,

$$f(|Q_{ij}(t)|) = \frac{|Q_{ij}(t)|^{\mu}}{1+|Q_{ij}(t)|^{\mu}}, \quad \mu > 1 \ .$$

a is a parameter expressing the extent of phototaxis and it controls the degeneration speed of the tube. When the flow rate is high or low, variations in the thickness of the tube decrease since it changes based on a sigmoid curve. μ is a gradient of the sigmoid function. It is the convergence speed of the thickness of the tube.

4. PS re-enters the updated tube thickness into Eq. (1).

By iterating the above mentioned procedure, PS converges the route between baits.

As described above, PS operates by treating the maze as a flow network. Therefore, it is possible to use PS as a routing solution in computer and transportation networks.

Car navigation has been proposed as one application of PS [18]. In this method, the system determines a route along an interstate highway between Seattle and Houston, USA, based on PS. PS derives a single route with the shortest mileage, when no trouble in the transportation network is encountered. When routes are congested, PS obtains the route with the shortest time of travel, by changing the value of a according to the traffic volume. Moreover, if an accident takes place in the middle of the route, PS identifies the optimum detour route to avoid the section of the road where the accident has taken place.

2.2. Physarum-based Routing Scheme

A Physarum-based routing scheme (P-bRS) has been proposed that applies PS to routing in wireless sensor networks (WSN) [19]. The network model of P-bRS assumes that a multi-hop WSN consists of static sensor nodes and a single mobile sink node, which are uniformly arranged in a two-dimensional space. The network model also assumes that each sensor node may obtain information about the position of all sensor nodes and their residual battery levels. The sink node broadcasts its current position periodically to the surrounding sensor nodes while moving along a specific route.

In P-bRS, each parameter of PS is redefined to apply PS to WSN. Equation (1) and Eq. (3) are changed to Eq. (4) and Eq. (6), respectively:

$$\begin{aligned} Q_{ij}(t) &= \frac{D_{ij}(t)}{L_{ij}}\left[p_i(t) - p_j(t)\right] \\ &= \frac{k\mathrm{ER}_j(t) + (1-k)\cos\theta_{jid}}{L_{ij}} = \frac{P_{ij}(t)}{L_{ij}} \ , \end{aligned} \tag{4}$$

$$\theta_{jid} = \arccos\frac{L_{ij}^2 + L_{id}^2 - L_{jd}^2}{2L_{ij}L_{id}}, \tag{5}$$

$$P_{ij}(t+\delta t) = P_{ij}(t) + \delta t\{(Q_{ij}(t))^{\mu} - P_{ij}(t)\} \ , \tag{6}$$

where: $Q_{ij}(t)$ is the virtual data packet size between node i and j, L_{ij} is the Euclidean distance between node i and j, ER_j is the residual battery level of node j, θ_{jid} is the angle of deviation that is derived from the cosine formula, and its range is $\left[-\frac{\pi}{2} \leq \theta_{ijd} \leq \frac{\pi}{2}\right]$. Figure 2 shows an example of next-hop selection in P-bRS. As shown in the figure, nodes closer to the sink node have a smaller angle of deflection. $D_{ij}(t)$, which is defined as the link quality, is omitted in Eq. (4) because the model assumes that $D_{ij}(t)$ is always constant. Coefficient k is used to adjust the weight of the residual battery level and the angle of deviation. Therefore, P-bRS constructs a route based on both the residual battery level and the angle of deviation to replace the pressure loss in PS.

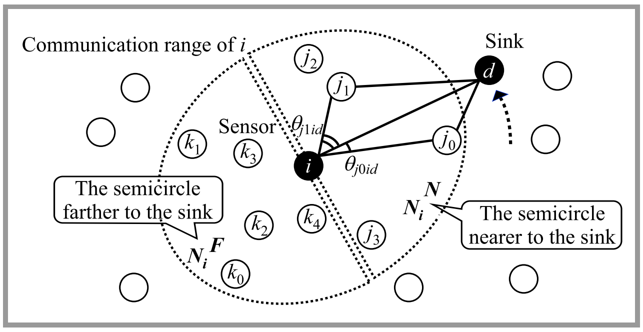


Fig. 2. Example of topology in P-bRS.

The operating principle of P-bRS in the scenario in which data is transmitted from sensor node i to sink node d is shown below. When a data transmission request is placed, sensor node i divides its communication range into semicircle N_i^N, which is closer to the sink node, and semicircle N_i^F, which is further from the sink node, based on the position information received from the sink node. Thereafter, node i calculates $P_{ij}(t)$ for node j that belongs to N_i^N using Eq. (4)–Eq. (6). Here, node i gives order j_x, which is arranged by $P_{ij}(t)$ in a descending order. Node i selects the node j_0 as a relay node and transmits a control packet to the node j_0. Node j_0 transmits an acknowledgement packet (ACK) to node i after receiving the control packet. Then, if the node j_0 satisfies $N_{j_0}^N = \phi$, it does not transmit ACK. If node i has not received ACK from node j_0 after a certain period, node i selects the relay node j_x of the smallest x that satisfies $|\theta_{j_0id} - \theta_{j_xid}| \geq \frac{\pi}{2}$ and $N_{j_x}^N \neq \phi$, and transmits a control packet. If node j_x which satisfies these conditions does not exist, node i calculates $P_{ik}(t)$ for node k, which belongs to N_i^F. Here, node i gives an order k_y to the node k, which is arranged by $P_{ij}(t)$ in an ascending order. Node i selects relay node k_y, as in case of selection of node j_x, and retransmits the control packet to k_y. Therefore, P-bRS constructs a route to avoid nodes in low density areas, thus preventing an increase in the transmission delay. Node i initiates data transmission after receiving ACK from the node j_x or k_y to finish its routing process. By iterating the process described above until the data reaches the sink node, P-bRS constructs a route based on the Euclidean distance to the sink node, angle of deviation, and residual battery level, while avoiding the low node density area.

The results of computer simulation evaluations have confirmed that P-bRS improves the efficiency of the battery use and transmission delay of the nodes compared to the previous method. However, P-bRS does not consider bandwidth use, as derived from the communication performed by other source nodes. Therefore, data collection may be difficult when congestion occurs in the area closest to the sink node in a scenario with considerable communication volume.

3. Adaptive Load Balancing Routing Inspired by True Slime Mold

Now we propose an adaptive load balancing routing mechanism inspired by true slime mold model for ad hoc networks, which is capable of constructing multiple paths based on data transmission volume, available bandwidth of each link and the residual battery level of each node, by applying PS to a dynamic network.

Figure 3 presents an overview of PS in the proposed method. Here, the volume of data transmitted between end-to-end nodes, data transfer rate on each link and the transmission delay on each link in the figure correspond to the total amount of liquid, the flow rate of each tube and the length of each tube in physarum, respectively. Additionally, the bandwidth utilization rate is the rate occupied by the current bandwidth used within the maximum bandwidth of a link. The bandwidth occupancy rate is the part of the bandwidth occupied by the data transmission volume requested by the source node, within the maximum bandwidth of the link. The bandwidth occupancy rate corresponds to the thickness of the tube in physarum. Each node periodically calculates the bandwidth occupancy rate and the transmission delay time of all the links from the bandwidth utilization rate and the maximum bandwidth.

The operating principle of the proposed method is shown below.

1. When a data transmission request occurs, the source node calculates the data packet transfer time $p_{ij}(t)$ by Eqs. (7)–(8) from the transmission data size Q_{all}, bandwidth occupancy rate $D_{ij}(t)$, and the transmission delay $L_{ij}(t)$ between nodes i and j:

$$Q_{ij}(t) = \frac{D_{ij}(t)}{L_{ij}(t)} p_{ij}(t) \ , \tag{7}$$

$$\sum_i Q_{ij} = \begin{cases} -Q_{\text{all}} \, , & j = \text{source} \\ Q_{\text{all}} \, , & j = \text{destination} \\ 0 \quad , & \text{otherwise} \end{cases} . \tag{8}$$

2. The source node calculates the transfer data size $Q_{ij}(t)$ on the link i–j from the data packet transfer time by Eq. (7). Consequently, larger data transfer volumes are allocated to the link that has a larger bandwidth occupancy rate and a smaller transmission delay.

3. The bandwidth occupancy rate is updated by:

$$D_{ij}(t+\delta t) = D_{ij}(t) + \delta t \{ f(|Q_{ij}(t)|) - a D_{ij}(t) \} \ . \tag{9}$$

Here,

$$f(|Q_{ij}(t)|) = \frac{|Q_{ij}(t)|^{\mu}}{1+|Q_{ij}(t)|^{\mu}}, \quad \mu > 1 \ . \tag{10}$$

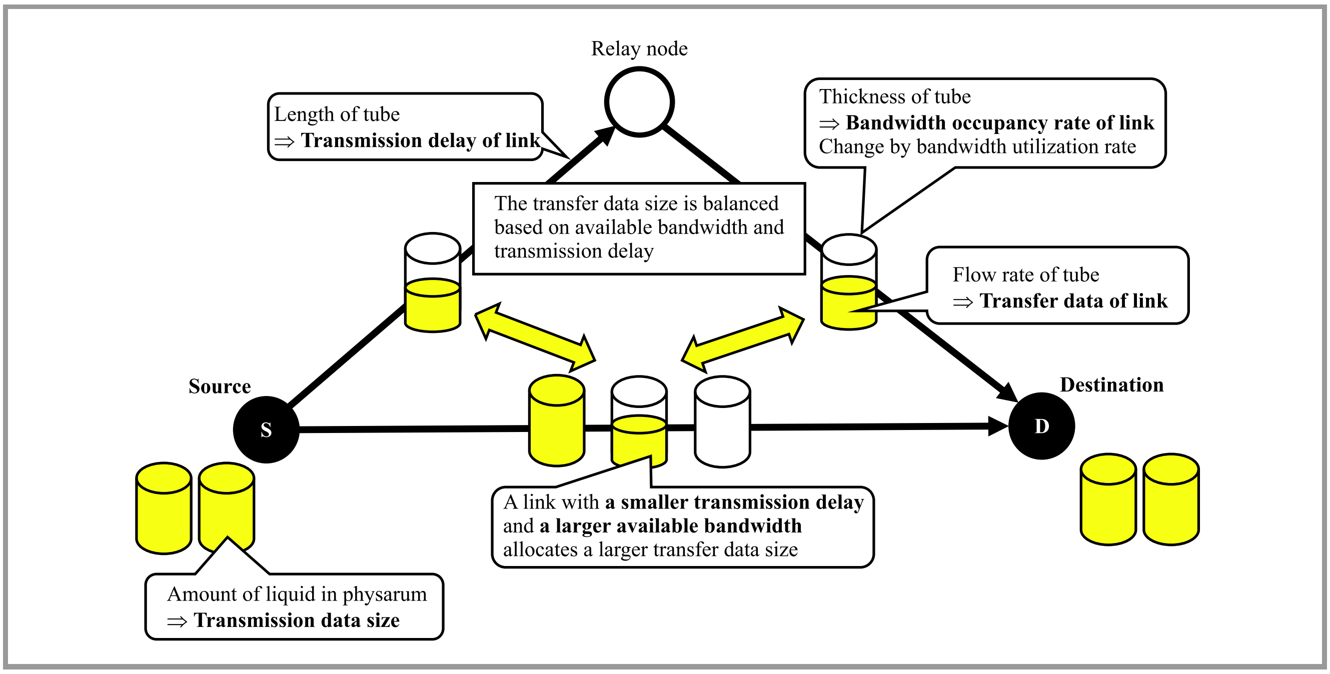

Relay node
Length of tube
⇒ **Transmission delay of link**
Thickness of tube
⇒ **Bandwidth occupancy rate of link**
Change by bandwidth utilization rate
The transfer data size is balanced based on available bandwidth and transmission delay
Flow rate of tube
⇒ **Transfer data of link**
Source
S
Destination
D
A link with **a smaller transmission delay** and **a larger available bandwidth** allocates a larger transfer data size
Amount of liquid in physarum
⇒ **Transmission data size**

Fig. 3. Overview of PS in the proposed method.

$\delta t\{f(|Q_{ij}(t)|) - aD_{ij}(t)\}$ in Eq. (9) represents the variation in the bandwidth occupancy rate after δt. The bandwidth occupancy rate asymptotically converges to specific values because the variation becomes small when using a sigmoid function, even if the transfer data size is large. Therefore, as the data transmission volume increases, the number of links with saturated bandwidth occupancy rates increases as well. The damping coefficient a in Eq. (9) is a parameter for changing the bandwidth occupancy rate independently of the data transfer volume. It is normally set to 1. It enables a path to be constructed that prioritizes a relay node with a large amount of residual battery level, by changing the damping coefficient according to the residual battery life of the node. Thereby, it may improve stability of the route to avoid route disruption due to the loss of battery charge.

The source node iterates the above calculation until the data transfer volume on each link converges. After convergence, the source node begins data packet transmission based on the data transfer volume of each link, while distributing the traffic among the links. Therefore, the proposed method may transmit data along a constructed route based on the data transmission volume and available bandwidth.

4. Performance Evaluation

4.1. Simulation Environment

Simulations evaluate the behavior and performance of the proposed method using software relying on the C++ programming language. First, performed a simulation based on the topology shown in Fig. 4. In simulations 1 to 3, the maximum bandwidth of each link was set to 11 Mbps on the assumption that the nodes used IEEE 802.11b [9] as the wireless communication medium, and the initial bandwidth utilization rate was set to 0%. Meanwhile, in simulations 4.1 and 4.2, the maximum bandwidth of each link was set to several fixed values based on the rates supported by IEEE 802.11ac [10] and the initial bandwidth utilization rate was set to 0%. We observed and evaluated the transition of the allocated data transfer volume on each path, and used it as an evaluation index, under several different conditions.

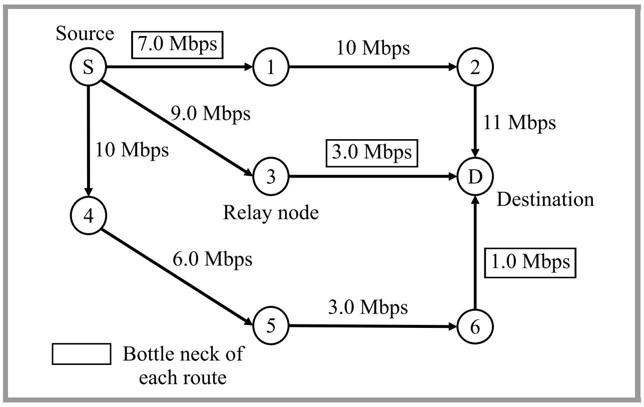


Fig. 4. Simulation topology utilizing 802.11b and the available bandwidth when the iteration count is 2000 (default is 11 Mbps).

In this paper, we completed the following simulations under different conditions.

In simulation 1, we evaluate the effect of varying available bandwidth using the topology as shown in Figs. 4 and 5. We assume that the available bandwidth of each link changes due to radio interference in the simulation. Therefore, the maximum bandwidth of each link and bandwidth utilization rate of each link varies within the range of 1 and 11 Mbps and 1% to 100%, respectively, when the number of iterations of calculations reaches 2000 and 4000. Figures 4 and 5 show the simulation topology and available bandwidth on each link when the number of calculation iterations reaches 2000 and 4000.

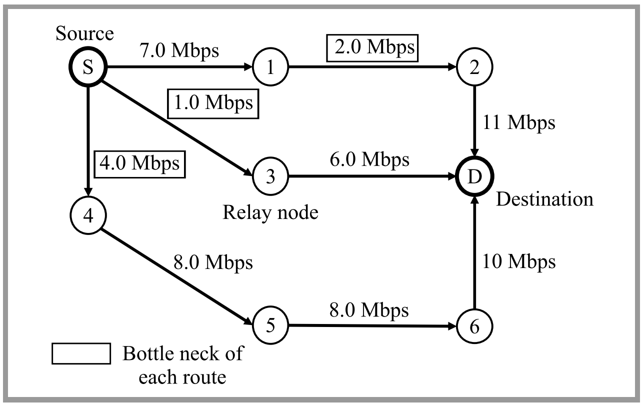


Fig. 5. Simulation topology utilizing 802.11b and the available bandwidth when the iteration count is 4000.

In simulation 2, we evaluate the effect of varying data transfer volume using the same topology of simulation 1 that is shown in Fig. 4. We assume the data transmission volume is changed before determining a route. Therefore, the data transfer volume is set to 5 MB at the initial time. Then, when the number of calculation iterations reaches 2000, the data transfer volume is changed to 10 MB, and when the number of iterations reached 4000, it changed to 5 MB.

In simulation 3, we evaluate the effect of varying the residual battery level of the node using the topology as shown in Fig. 4. In the topology of Fig. 4, node 3 transmits data packets with a higher frequency in comparison with other nodes since it is located along the shortest path. Namely, the battery consumption of the node also increases in comparison with other nodes due to the reason referred to above. Assuming that node 3 exceeds its battery power available when the number of iterations reaches 2000 and 4000, node 3 adds one to its damping coefficient. Additionally, if the bandwidth occupancy rate of the link decreases as the residual battery level of node 3 becomes lower, it is expected that the data transfer volume and the battery consumption of the nodes along the second shortest path S-1-2-D increase. Thus, node 1 adds one to its damping coefficient every 1000 iterations after the number of iterations has reached 6000.

In simulation 4.1, we evaluate the effect of varying available bandwidth widely using the topology as shown in Figs. 6 and 7. Therefore, the simulation assumes a fluctuation of bandwidth when using IEEE 802.11ac as the wireless communication medium. Here, we assume IEEE 802.11ac that adopted OFDM-MIMO as the primary modulation scheme, 64 quadrature amplitude modulation 5/6 as the secondary modulation scheme, and the guard interval is 800 ns. The maximum number of enable streams of each node is 8. Under the above conditions, the theoretical values of the communication speed are 292.5, 585, 877.5, and 1170 Mbps when the number of antennas is 1, 2, 3, and 4, respectively. Additionally, we suppose that the bandwidth occupancy rate on each link is 100% in the simulation. Hence, we change the transmission speed stepwise based on the rate candidates when the number of calculation iterations reaches 2000 and 4000. Figures 6 and 7 show the simulation topology and available bandwidth on each link when the number of calculation iterations reaches 2000 and 4000.

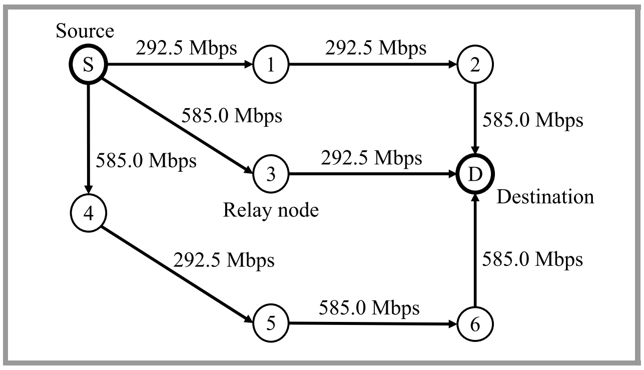


Fig. 6. Simulation topology utilizing 802.11ac and the available bandwidth when the iteration count is 2000 (default is 292.5 Mbps).

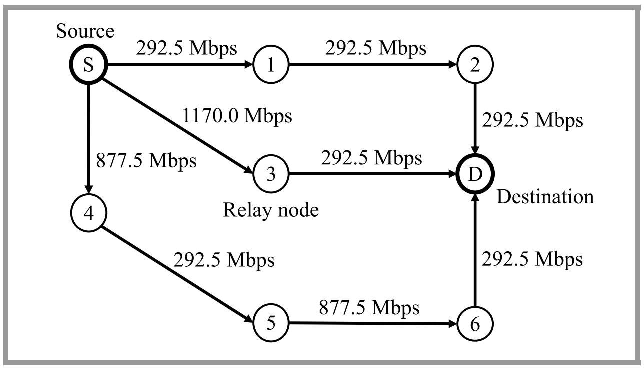


Fig. 7. Simulation topology utilizing 802.11ac and the available bandwidth when the iteration count is 4000.

In simulation 4.2, we evaluate the effect of varying available bandwidth using another topology as shown in Figs. 8 and 9. Unlike in the above simulations, in the topologies of the simulation shown in Figs. 8 and 9 relay nodes have three or more links. Therefore, we observe the changes in the convergence behavior of the proposed method, since the network topology changes.

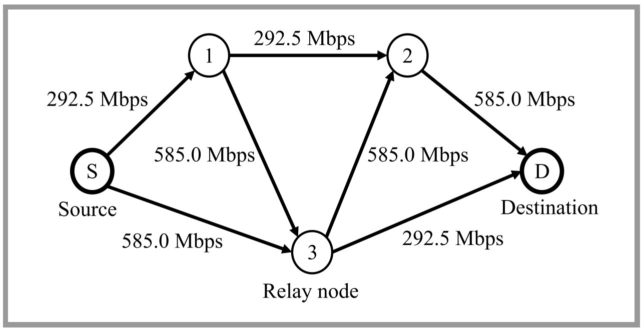


Fig. 8. Simulation topology and the available bandwidth when the iteration count is 2000 (default is 292.5 Mbps).

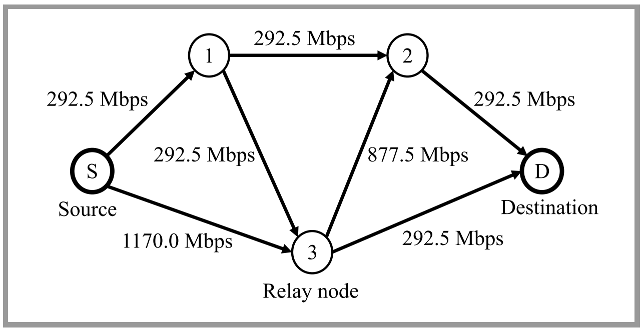


Fig. 9. Simulation topology and the available bandwidth when the iteration count is 4000.

4.2. Simulation Results

The results of the four simulations described in 4 are shown in Figs. 10 to 12, Fig. 13, Figs. 14 and 15, and Figs. 16 and 17, respectively. The simulations sufficiently iterate the calculation to observe the effect of convergence of the proposed method. Note that the number of iterations may be shortened by aborting the calculation after a certain number of iterations, when the data transfer volume has been sufficiently converged in a practical situation.

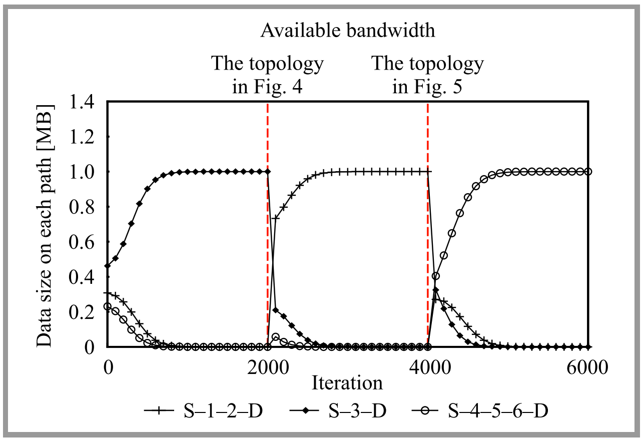


Fig. 10. Simulation 1: the transfer data size when the data transmission volume is fixed at 1 MB.

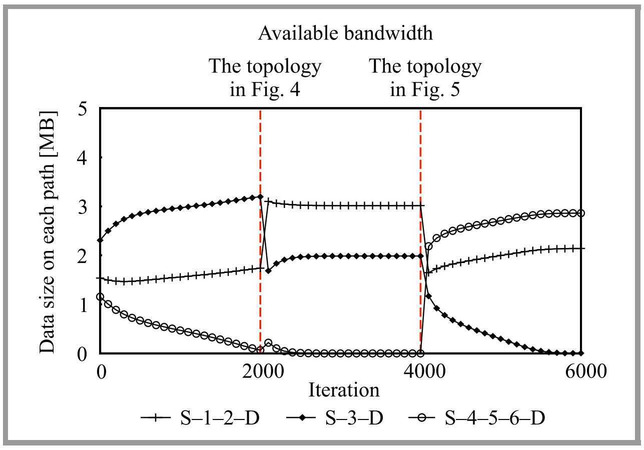


Fig. 11. Simulation 1: the transfer data size when the data transmission volume is fixed at 5 MB.

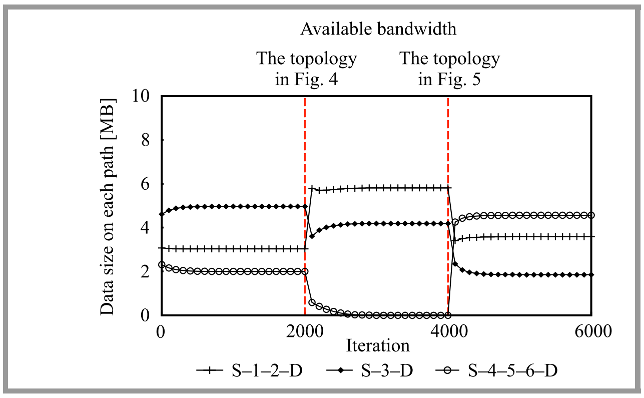


Fig. 12. Simulation 1: the transfer data size when the data transmission volume is fixed at 10 MB.

Simulation 1. Figures 10 to 12 show that the data transfer volume of each path increases as the available bandwidth of each link in the path increases. Furthermore, if multiple paths exist, the proposed method balances the required time to transfer data among the paths by calculating their appropriate data transfer volumes based on available bandwidth. In particular, the largest data transfer volume among the paths is allocated to path S-3-D with the minimum hop count at the initial state of the simulation. This is because the data transfer volume is preferentially allocated to the path with the minimum transmission delay since the available bandwidth of all links is uniform in the simulation.

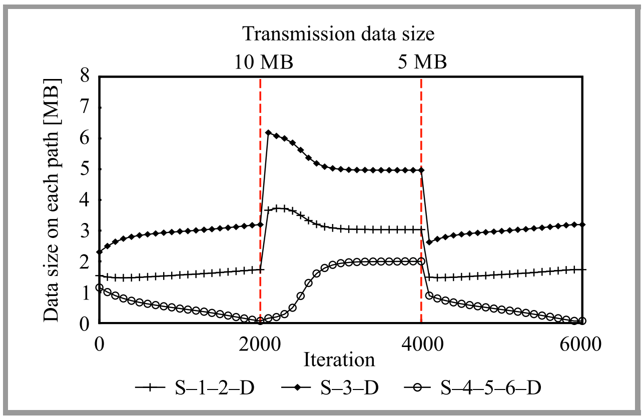


Fig. 13. Simulation 2: data transfer volume when varying data transmission volume.

Simulation 2. Figure 13 shows that the number of paths increased when the data transfer volume increased. This is because path S-4-5-6-D is constructed as the third path, since the bandwidth occupancy rate of the other two paths is saturated owing to the increase in data size. Additionally, after the data transfer volume is reduced to 5 MB, the allocated data size is also decreased and then path S-4-5-6-D disappears. Since the bandwidth occupancy rate of the other two paths gives a bandwidth margin due to a decrease in the data transfer volume.

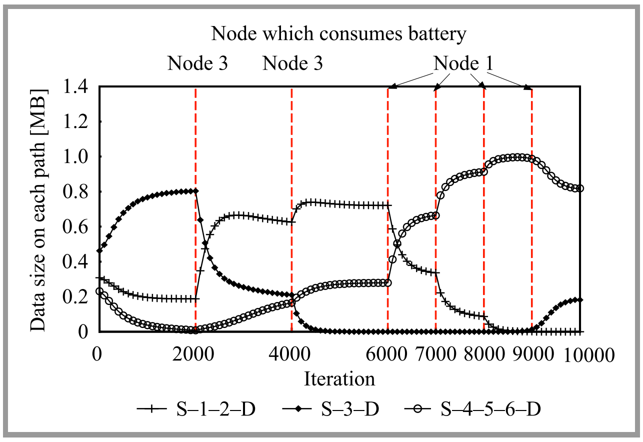


Fig. 14. Simulation 3: data transfer volume when varying residual battery level (data transmission volume: 1 MB).

Simulation 3. Figure 14 shows that the detour path S-1-2-D is used since the data transfer volume on path S-3-D via node 3 decreases. In addition, the data transfer volume on path S-4-5-6-D also increases. After, the data transfer data volume on path S-1-2-D decreases because node 1 consumes its battery, when the number of iterations is 6000. Finally, path S-4-5-6-D, that has a large amount of residual battery of the nodes, is used. Figure 15, shows that paths

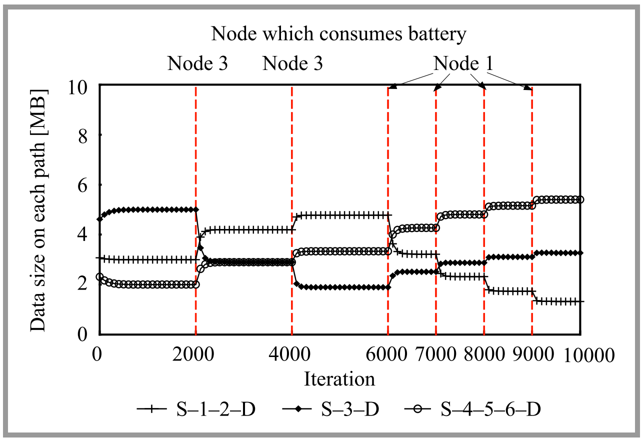


Fig. 15. Simulation 3: the transfer data size when varying residual battery (data transmission volume: 10 MB).

S-1-2-D and S-3-D do not disappear, unlike the result in Fig. 14. Since the bandwidth occupancy rate of each path is saturated due to the assigned large data size, the variation of the data transfer volume becomes very small. These results indicate that PS can select routes with a low risk of route disruption due to battery loss.

Simulation 4.1. Figure 16 shows that the path convergence is faster than in simulations 1 to 3 since the variation of the bandwidth occupancy rate reaches its peak at the first calculation step by the sigmoid function due to the large data transfer volume.

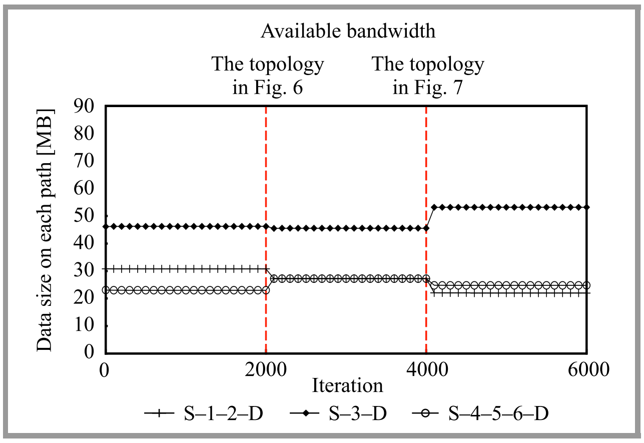


Fig. 16. Simulation 4.1: data transfer volume when the data transmission volume is fixed at 100 MB.

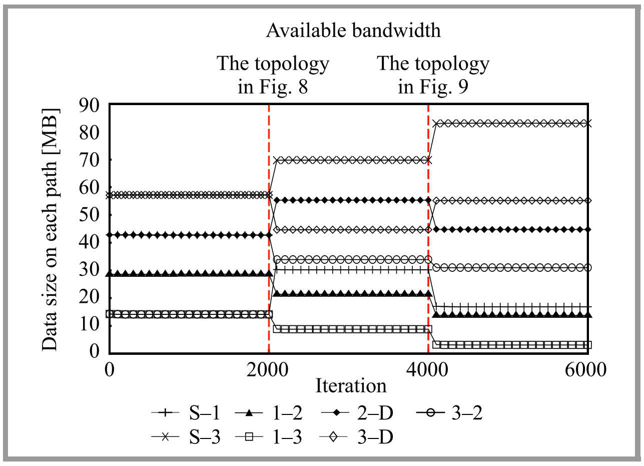


Fig. 17. Simulation 4.2: data transfer volume when the data transmission volume is fixed at 100 MB.

Simulation 4.2. Figure 17 shows that the proposed method allocates the data transfer volume based on the available bandwidth of each link while continuing to balance the sent and received data volumes. Additionally, even if the data transfer volume is the same as in simulation 4.1, link 1–3 appears, to which the data transfer volume is scarcely allocated, when the number of iterations is 6000. This is because the number of available paths increased due to the topological change. Hence, these results indicate that the proposed method realizes the allocation of data transfer volume based on the bandwidth available along each path even when a relay node on which the confluence of flows occurs exists in the topology.

5. Conclusions

In this paper, we proposed an adaptive ad hoc routing method that can construct multiple paths based on the available bandwidth of each link, data transfer volume and residual battery level of the node by applying PS to the dynamic networks. As a result of our simulations, we confirmed that the proposed method is capable of adaptively constructing single or multiple paths based on the available bandwidth, data transfer volume and residual battery level of nodes, in the dynamic network topology. Our data suggested that PS improves availability of ad hoc network. In the present study, we found that the path convergence time and the number of available paths change according to the number of nodes and links in the topology and the parameters of each link. In the future, the increase in transmission delay due to the increasing of hop counts will be investigated. Moreover, we will define and evaluate the load balancing rate and will propose the parameters needed to optimize the number of paths in various topologies. Additionally, we will conduct a detailed performance evaluation of the proposed method by extension of the existing routing protocol through a network simulator. Furthermore, the proposed method may be extended to be used in a table-driven routing protocol, because the operation of the proposed method requires such network-related information as the maximum bandwidth of the link.

Acknowledgments

This work was supported by JSPS KAKENHI under Grant Number 17K12680.

References

[1] M. Shimomura, "The new trends in next generation biomimetics material technology: Learning from biodiversity", *Sci. Technol. Trends. Q. Rev.*, vol. 37, pp. 53–75, 2010 (doi: 11035/2843).

[2] D. Chen, Y. Liu, H. Chen, and D. Zhang, "Bio-inspired drag reduction surface from sharkskin", *Biosurf. and Biotribol.*, vol. 4, no. 2, pp. 39–45, 2018 (doi: 10.1049/bsbt.2018.0006).

[3] S. Das, M. Bhowmick, S. K. Chattopadhyay, and S. Basak, "Application of biomimicry in textiles", *Curr. Sci.*, vol. 109, no. 5, pp. 893–901, 2015 (doi: 10.18520/v109/i5/893-901).

[4] T. Nakagaki, H. Yamada, and A. Tóth, "Path finding by true morphogenesis in an amoeboid organism", *Biophys. Chemistry*, vol. 92, no. 1, pp. 47–52, 2001 (doi: 10.1016/S0301-4622(01)00179-X).

[5] G.-D. Caro, F. Ducatelle, and L.-M. Gambardella, "AntHocNet: an adaptive nature-inspired algorithm for routing in mobile ad hoc networks", *Eur. Trans. on Telecommun.*, vol. 16, pp. 443–455, 2005, (doi: 10.1002/ett.1062).

[6] R. Leidenfrost and W. Elmenreich, "Firefly clock synchronization in an 802.15.4 wireless network", *EURASIP J. on Embed. Syst.*, vol. 2009, article no. 7, 2009 (doi: 10.1155/2009/186406).

[7] M. Hato, T. Ueda, K. Kurihara, and Y. Kobatake, "Phototaxis in true slime mold physarum polycephalum", *Cell Struct. & Funct.*, vol. 1, no. 3, pp. 269–278, 1976 (doi: 10.1247/csf.1.269).

[8] S. Corson and J. Macker, "Mobile ad hoc networking (MANET): routing protocol performance issues and evaluation considerations", RFC 2501, IETF, 1999 (doi: 10.17487/RFC2501).

[9] "IEEE standard 802.11. Wireless LAN medium access control (MAC) and physical layer (PHY) specifications", IEEE Std 802.11, June 1999.

[10] "IEEE Standard for Information technology – Telecommunications and information exchange between systems Local and metropolitan area networks – Specific requirements – Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications – Amendment 4: Enhancements for very high throughput for operation in bands below 6 GHz", IEEE Std 802.11ac-2013, Dec. 2013.

[11] A. Kamerman and L. Monteban, "WaveLAN-II: a high-performance wireless LAN for the unlicensed band", *Bell Labs Tech. J.*, vol. 2, no. 3, pp. 118–133, 2002 (doi: 10.1002/bltj.2069).

[12] R. Karmakar, S. Chattopadhyay, and S. Chakraborty, "Impact of IEEE 802.11n/ac PHY/MAC high throughput enhancements on transport and application protocols – A Survey", *IEEE Commun. Surveys & Tutor.*, vol. 19, no. 4, pp. 2050–2091 (doi: 10.1109/COMST.2017.2745052).

[13] G. Holland, N. Vaidya, and P. Bahl, "A rate-adaptive MAC protocol for multi-hop wireless networks", in *Proc. 7th Ann. Int. Conf. on Mob. Comput. & Netw. MobiCom'01*, Rome, Italy, 2001, pp. 236–251 (doi: 10.1145/381677.381700).

[14] M. Lacage, H. Manshaei, and T. Turletti, "IEEE 802.11 rate adaptation: a practical approach", in *Proc. 7th ACM Int. Symp. on Model., Anal., and Simul. of Wirel. and Mob. Syst. MSWiM 2004*, Venice, Italy, 2004, pp. 126–134 (doi: 10.1145/1023663.1023687).

[15] X, Wang and C. Li, "A topology-independent broadcasting protocol in ad hoc networks with MIMO links", in *Proc. Int. Conf. on Commun. Mob. Comput. CMC 2010*, Shenzhen, China, 2010, vol. 3, pp. 219–223 (doi: 10.1109/CMC.2010.291).

[16] A. Tero, R. Kobayashi, and T. Nakagaki, "A mathematical model for adaptive transport network in path finding by true slime mold", *J. of Theor. Biology*, vol. 244, no. 4, pp. 553–564 (doi: 10.1016/j.jtbi.2006.07.015).

[17] A. Tero *et al.*, "Rules for biologically inspired adaptive network design", *Science*, vol. 327, no. 5963, pp. 439–442, 2010 (doi: 10.1126/science.1177894).

[18] A. Tero, R. Kobayashi, and T. Nakagaki, "Physarum solver: a biologically inspired method of road-network navigation", *Physica A: Statis. Mechan. and its Appl.*, vol. 363, no. 1, pp. 115–119, 2006 (doi: 10.1016/j.physa.2006.01.053).

[19] M. Zhang, W. Wei, R. Zheng, and Q. Wu, "P-bRS: a physarum-based routing scheme for wireless sensor networks", *The Scientific World J.*, vol. 2014, Article ID 531032, 2014 (doi: 10.1155/2014/531032).

Hiroshi Katada received his B.E. degree in Electronic Information Systems from Shibaura Institute of Technology, Tokyo, Japan, in 2018. He is presently a master's course student at the Graduate School of Engineering and Science, Shibaura Institute of Technology, Saitama, Japan. His research interests include mobile ad hoc networks and engineering neo-biomimetics.

E-mail: mf18024@shibaura-it.ac.jp
Graduate School of Systems Engineering and Science
Shibaura Institute of Technology
Saitama, Japan

Takumi Miyoshi received his B.E., M.E., and Ph.D. degrees in Electronic Engineering from the University of Tokyo, Tokyo, Japan, in 1994, 1996, and 1999, respectively. He started his career as a research associate at Global Information and the Telecommunication Institute, Waseda University, where he worked from 1999 to 2001. He is presently a Professor at the Department of Electronic Information Systems, College of Systems Engineering and Science, Shibaura Institute of Technology, Saitama, Japan. He was a visiting scholar in Laboratoire d'Informatique de Paris 6 (LIP6), Sorbonne Université, Paris, France, from 2010 to 2011. His research interests include content delivery networks, overlay networks, as well as mobile ad hoc and sensor networks.

E-mail: miyoshi@shibaura-it.ac.jp
College of Systems Engineering and Science
Shibaura Institute of Technology
Saitama, Japan

Taku Yamazaki – for biography, see this issue, p. 12.

Paper

Robot Local Network Using TQS Protocol for Land-to-Underwater Communications

Addie Irawan, Mohammad Fadhil Abas, and Nurulfadzilah Hasan

Faculty of Electrical and Electronics Engineering, Universiti Malaysia Pahang, Pekan, Malaysia

https://doi.org/10.26636/jtit.2019.125818

Abstract—This paper presents a model and an analysis of the Tag QoS switching (TQS) protocol proposed for heterogeneous robots operating in different environments. Collaborative control is topic that is widely discussed in multirobot task allocation (MRTA) – an area which includes establishing network communication between each of the connected robots. Therefore, this research focuses on classifying, prioritizing and analyzing performance of the robot local network (RLN) model which comprises a point-to-point topology network between robot peers (nodes) in the air, on land, and under water. The proposed TQS protocol was inspired by multiprotocol label switching (MPLS), achieving a quality of service (QoS) where swapping and labeling operations involving the data packet header were applied. The OMNET++ discrete event simulator was used to analyze the percentage of losses, average access delay, and throughput of the transmitted data in different classes of service (CoS), in a line of transmission between underwater and land environments. The results show that inferior data transmission performance has the lowest priority with low bitrates and extremely high data packet loss rates when the network traffic was busy. On the other hand, simulation results for the highest CoS data forwarding show that its performance was not affected by different data transmission rates characterizing different mediums and environments.

Keywords—class of service, land-to-underwater communications, robot local network, tag switching.

1. Introduction

Heterogeneous robot communication or multirobot networking is conducted mainly via the Internet. There are many issues related to establishing robot local networks (RLN) for collaboration purposes, and they need to be resolved in order to achieve specific objectives, with a particular emphasis placed on dependability, safety and security of the system. To enable collaboration between robots, two elements need to be defined, i.e. the communication protocol and the collaborative procedures. Undoubtedly, an efficient protocol is essential for effective decentralization of the distribution of data between the robots.

Researchers have taken several approaches to improve communication between heterogeneous robots. Some of them rely on the communication protocol, while others use the intelligent control technique to improve the performance of the system. There is also a hybrid approach which combines both these techniques. Stamatescu *et al.* described the communication protocol by applying the cognitive radio (CR) scheme, i.e. by exploiting the time, frequency and spatial stream of the wireless environment. According to the testing results, they claimed that the communication reliability at each hierarchical level increased [1]. Another approach consists in applying a formal taxonomy to the allocation of tasks to a mobile robot, as proposed by Gerkey and Mataric [2]. This method is further improved by Korsah *et al.* with their proposed iTax, a taxonomy addressing interrelated utilities and constraints via a combination of optimization methods and operation research. This method is based on the recognition that the key distinguishing factor between different types of multirobot task allocation (MRTA) problems is the degree of interdependence of agent-task utilities [3].

Adaptation of intelligent systems is also becoming an approach that is favored in multirobot collaborative control and communication. Zhang *et al.* applied an adaptive fuzzy logic in tackling MRTA reliance on the intuitionistic fuzzy set theory [4]. Similarly, Cheng *et al.* proposed a linear temporal logic (LTL) which optimizes path planning by using the formation control feedback mechanism [5]. Power is another constraint that needs to be considered while establishing a collaborative robotic system. Moreover, there is a tradeoff between the number of nodes that can be deployed in the mission and the density of information that can be exchanged. Bano *et al.* in [6] explored these constraints and proposed a random waypoint mobility model in a mobile ad-hoc network (MANET). This system was tested with a robot group comprising 5–6 collaborating robotic nodes and the results showed that it was better than the Manhattan mobility model. MANET is also implemented by Kulla *et al.* for a real-time emergency scenario of moving the multirobot (nodes) indoor [7]. Bandwidth sharing is also possible for an MRTA communication system using resource controller (RC) and aggregate resource controller (ARC) management techniques [8].

Classification and aggregation also offers a vast potential that may be explored in robot-to-robot or human-to-robot communication, and mainly in enhancing reliability performance and in ensuring priority for different types of data exchanged over the Internet. Automation and robotics place a greater emphasis on the physical layers, with real-time data and control area network (CAN) serving as the primary platform. A virtual private network is available in the Internet protocol (IP) version, such as the virtual private network (VPN) used for long distance and indoor remotely controlled mobile robots [9], [10]. However, the problem is still an issue in the case of communication between mobile robots (swarm scenario).

The present research proposes the application of tag switching inspired by multiprotocol label switching (MPLS) [11], as a data carrying technique for an RLN operating in a different environment. A modular network testbed in C++ (OMNeT++) [12], i.e. a discrete event simulator, was used to develop a logical RLN model to perform a case study focusing on multirobot communication in a different environment. The nodes in the RLN were programmed with the Tag QoS switching (TQS) protocol proposed, and were considered to be robots (moving nodes). Data forwarding performance, expressed as the percentage of packets lost, average edge-to-edge access delays, as well as throughput, was verified by comparing one line of transmission between the edges of robots in a different environment.

2. Modeling of a Local Multirobot Network with Tag Switching Protocol

2.1. Case Study Involving Land-to-underwater Communication

For the case of multirobot communication involving different environments, e.g. in the land-to-underwater scenario, the robots may be flying in the air, may be submerged under water or may be placed on the ground, as shown in Fig. 1. Underwater communications have limitations in terms of distance and bandwidth. For example, optical wave transmission requires high precision in pointing narrow laser beams and is affected by scattering, although is resistant to high attenuations [13]. On the other hand, electromagnetic waves are also limited to short distance with the highest frequency at about 2.4 GHz for 250 Kbps, according to the IEEE 802.15.4 standards [14]. Alternatively, data signal may be also propagated in conductive salty water by using radio frequency (RF), but only at extra-low frequencies (30–300 Hz) that may require large antennas and high transmission power [15]. Several attempts have been made to enhance the speed and throughput of transmission, such as a routing technique relying on the surface of the water for underwater communications [16], as well as using the water surface relay to increase the overall transmission speed [17].

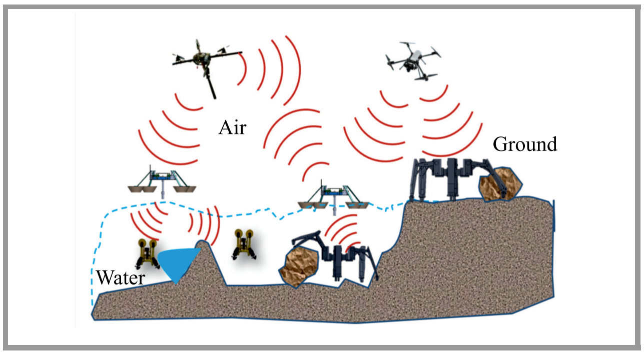


Fig. 1. Example of a heterogeneous RLN topology for land-to-underwater communications.

Therefore, this study investigates and analyzes data forwarding and switching/routing performance for the land-to-underwater communication scenario, using the proposed TQS protocol. It emphasizes the RLN topology in which robots are considered to be dynamic nodes of the switches, or routers for data transceivers. This study neglects all salt- and tide-related factors, as well as noise present in both mediums. The analysis focuses on logical data transmission implications and on the dynamic changes in data transmission rates experienced when the packet of data enters the water using the proposed protocol. Moreover, the focus is placed mainly on data the forwarding period and on traffic management. Some of the switching/router models were programmed to operate at low data transmission rates (10–100 Kbps) to represent nodes in the underwater environment, while other were programmed for speeds of up to 1 Gbps to represent nodes on land/in the air.

2.2. Tag-QoS Switching Protocol

The TQS protocol is suited for a network topology with dynamic nodes, such as RLN, as it is inspired by MPLS. MPLS was released by Cisco System in 1998 and started to gain popularity in IP deployment for wide area networks (WAN) and metropolitan area networks (MAN) [18], [19] in 2000. MPLS allows tunnel routing known as label switched paths (LSP), where a tunnel is characterized by a path in the network and by a reserved bandwidth [20], [21]. This protocol belongs to layer 2.5 in the open systems interconnection (OSI) model [22]. It improves both layers 2 and 3 by providing fast switching and reliable routing. Moreover, this protocol is bonded to an IP network as an extra header that involves the Internet service provider (ISP) area [23], [24]. However, the dynamic tagging and stacking methods used in MPLS have the potential to be deployed in a small-scale local network and data communication scheme, such as RLN shown in Fig. 1. The label swapping concept in MPLS [25] enables dynamic establishment of tunnels that depends on traffic demand. The tunnels are opened based on aggregation, classification and prioritization of communication between peer robots. MPLS also enables network virtualization through the labeling or tagging method, in order to create virtual and physical layers [26] that leverage the implementation of energy-aware traffic engineering [27].

The TQS protocol proposed is applied in the same manner as the MPLS label stack entry (LSE) shown in Fig. 2, but with specific label value calculations which involve both the indication of the LSP and the differential service (DiffServ). The calculation also includes a flow aggregate that requires network traffic to be marked and conditioned at the edges of the network, ensuring a different treatment for each of the tagged packets. The label value in TQS-LSE is:

$$L = \left[\alpha(N+1)\right] + P \,, \tag{1}$$

where α is the definite positive gain used for simple indication and for reducing conflicts in the tagging process. $N = 1, 2, \ldots, n$ is an LSP identification number, expressed as an integer, which represents the type of incoming data e.g. video or voice streaming. $P = 0, 1, 2, \ldots, n$ is denoted as a sub-LSP identification number for the bandwidth that is generated by the packet index at the edge of the node I when the number of channels for the D group of bandwidths satisfies the condition $I > D-1$. The P value can be obtained from:

$$P = \begin{cases} P = 0 & I > D-1 \\ P = I & D \leq D-1 \end{cases} . \tag{2}$$

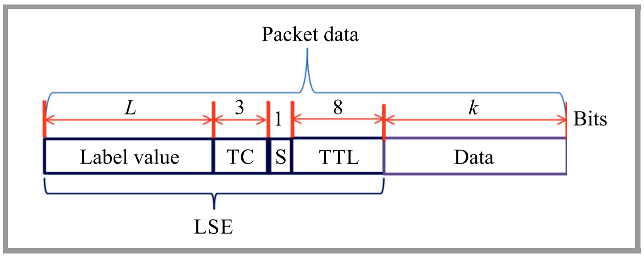


Fig. 2. LSE as a data tag in the Tag QoS switching protocol proposed in an RLN data distribution scheme.

As far as RFC 3270 is concerned, the experimental bits (EXP), renamed as traffic class (TC) bits [28], are used to encode and aggregate all per-hop behavior (PHB) bits from the data header to LSE [29]. In addition to that, three behaviors aggregating bits (BA) were used to encode with the DiffServ code point (DSCP) from the data header [30]. DSCP defines drop precedence in each type of class of services (CoS) for each data packet. MPLS-QoS encoding provides inferring CoS and drops precedence information from the data header to LSE. The bottom of stack (S) bit for the last entry of the label stack indicator and time-to-live (TTL) bits, as shown in Fig. 2, are the standard bit in the MPLS format, as defined in RFC3032 [11].

2.3. RLN Topology Model with TQS Protocol

For modeling and simulations relying on the proposed protocol, the nodes of switches/routers (robots) were categorized into two sections: tag edge mobile robot (TER) and tag switching robot (TSR). These switches/routers are the primary entities in RLN acting as transceivers for edges and switching, respectively, according to the TQS protocol data forwarding. As shown in Fig. 3, for simulation and analysis purposes, the RLN was modeled with several TERs and TSRs in OMNeT++. Here, the TER module was programmed to generate a raw/unlabeled data packet. The system consists of three sub-elements which are categorized as wire switch module (WSM), buffer switch module (BSM), and bandwidth switch module (BandSM), as pictured in Fig. 4. The usage of WSM in the TER module is the main feature that differentiates it from the TSR (Fig. 5). The switching process is applied when the label was swapped in the TSR instead of WSM, as shown in Fig. 6. The incoming tagged packets are buffered and then passed to the tag switching module (TSM) to swap the tag or label for the next hop bandwidth in BandSM. The swap process depends on the information label map (ILM) that has been programmed in the forwarding information table (FIT) of TSR (Fig. 6). Label value will always be swapped or replaced with a different number for the next hop bandwidth. The same goes for TTL as S bit values, they are also continuously updated. On the other hand, the information about bandwidth assignment is extracted in BandSM from the inverse calculation of Eq. (1) to get the information about the LSP and sub-LSP switched to the next hop of peer robot/node. BandSM, either in TER or TSR, is programmed to control the per-flow threshold according to the proposed bandwidth assignment scheme (BandAS), as presented in Table 1. The peak data rate (PDR) for premium/expedited forwarding CoS, or the committed data rate (CDR) for Olympic/assured forwarding CoS will discard the incoming data packets whenever the threshold is reached.

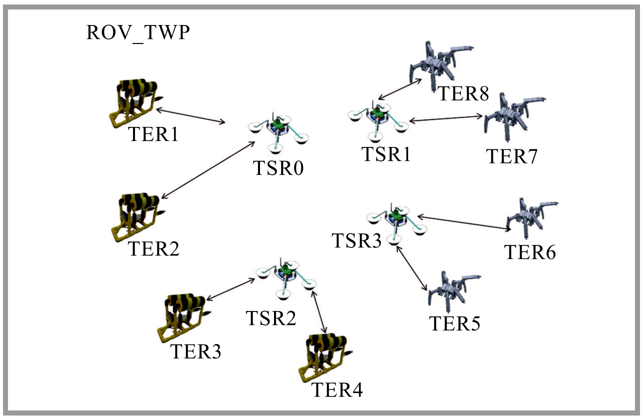


Fig. 3. RLN model topology for simulation and analysis in the OMNeT++ graphical runtime environment.

For the TER model, the process of forwarding the equivalent class (FEC) [11] to the next hop label forwarding entry (FTN) is applied in WSM as a labeling process, whereby the untagged data packet destination address is screened for the labeling process (Fig. 7). Initially, the data are generated with source address bits, destination/group address bits and hop limit values that are the same as those of packet data with the IP address. Moreover, the DSCP code in the generated packet data was aggregated to request the FEC code from the programmed FIT. Then, the label value with L bits (here 10 bits) was obtained (Fig. 7). The hop limit value and the DSCP code were then encoded to the TTL

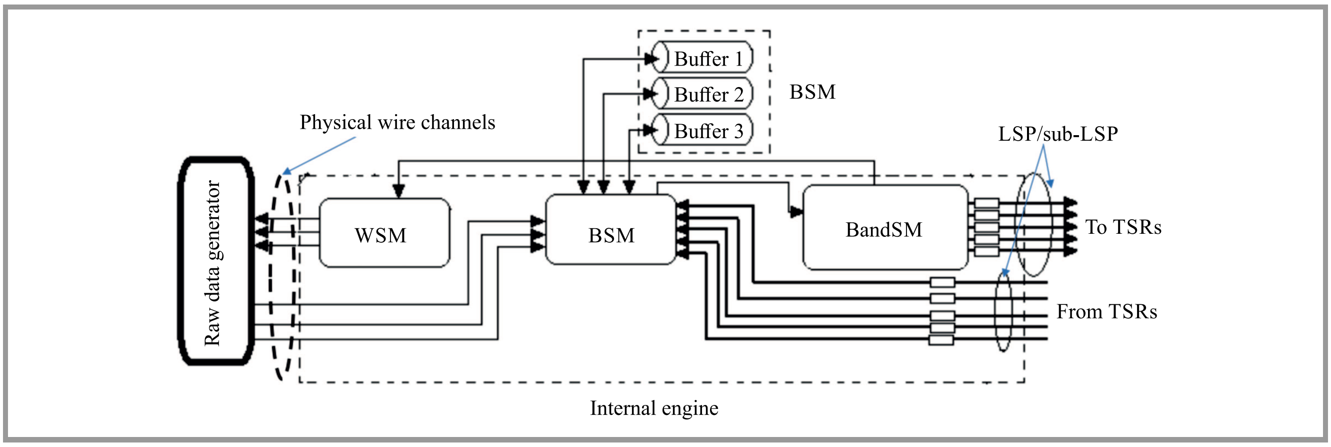


Fig. 4. TER model system engine with the TQS protocol proposed.

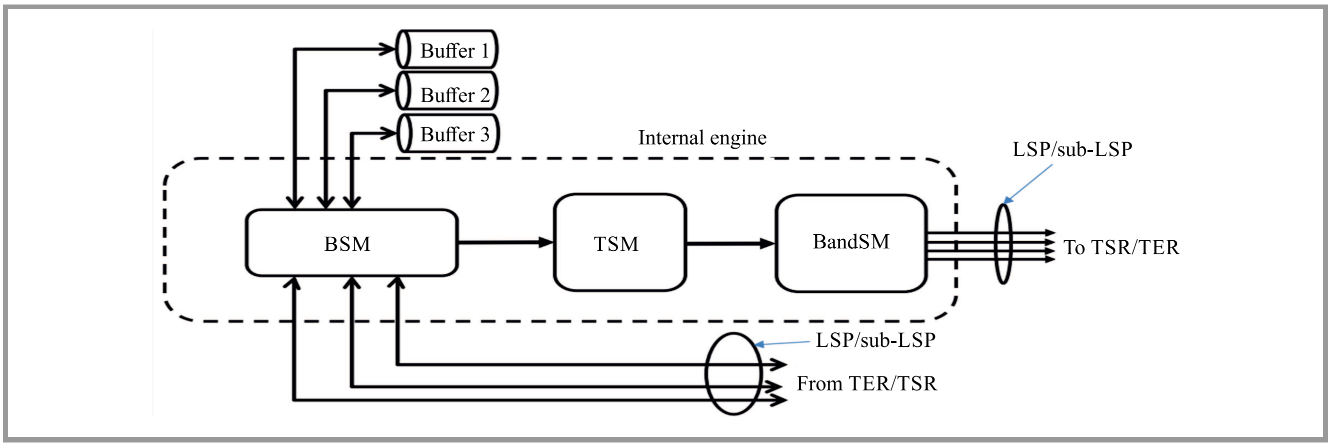


Fig. 5. TSR model system engine with the TQS protocol proposed.

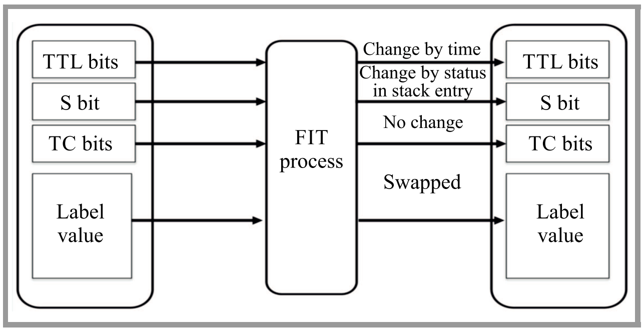


Fig. 6. Swapping process in BandSM.

Table 1
Bandwidth assignment scheme (BandAS)

Class of services		Channel bit rate
Premium		100% maximum bandwidth
Olympic	Gold	
	Silver	80% maximum bandwidth
	Bronze	60% maximum bandwidth
Best effort		40% maximum bandwidth

field and TC bits, respectively, using the MPLS-QoS [29] encoding method. The CoS for the applied QoS is determined according to the RFC 2597 draft, in which gold, silver, and bronze CoS of the Olympic CoS are applied [31].

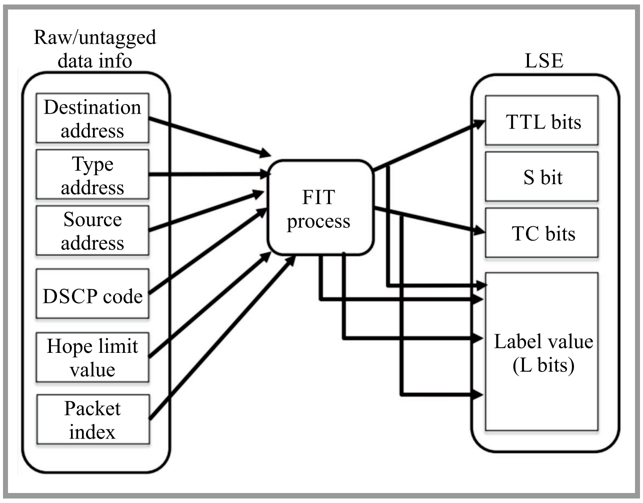


Fig. 7. FTN process for an unlabeled data packet in TM inside the TER model.

The drop precedence process is also conducted differently for all sub-CoS in the Olympic CoS. Hence, the channel bit rate was defined differently for the sub-CoS of the Olympic CoS in BandAS, as shown in Table 1. Gold CoS was the lowest drop precedence, and bronze CoS was the highest. With reference to the MPLS-QoS queuing process [29], as well as RFC2597, WSM in each TER and BandSM in TSR was programmed to perform marking through the labeling process on the untagged data packet. Data packet with low drop precedence was marked as a low priority packet to be discarded instead of the data packet with high drop precedence in BSM.

Table 2
Buffer assignment scheme

Buffer	Capacity (buffer length)
B0	20% × maximum buffer length
B1	40% × maximum buffer length
B2	Maximum buffer length

This first select-and-drop process (before BandAS) occurred when the number of packets in the allocated buffer reached its threshold limit, as allocated in Table 2, concerning the simple buffer assignment scheme [32]. The TC bit, as shown in Fig. 2, is an indicator for allocating the data to the particular buffer channel. The first-in-first-out (FIFO) principle was used in the BMS queuing process, where B0 and B2 for each TER/TSR were programmed to use the tail drop procedure, and B1 was programmed to use the random early detection (RED) procedure [33], mainly for Olympic CoS.

3. Simulation and Performance Analysis

Simulation and analysis of the RLN model were conducted and set up with the proposed TQS protocol, as shown in Fig. 3. The setup protocol was considered done in this simulation study, and all nodes were established with the least cost routing. The analysis relied on the same QoS class assessment program used in the model as applied in BandAS: premium, Olympic and best effort CoS. The analysis was performed by determining T1 as a focus line for performance evaluation, with other lines serving as disturbances for T1, as shown in Fig. 8. Moreover, TER1, TER2, TER3 and TER4 were modeled with proper underwater acoustic conditions, where the data rate for receiving/transmitting data packets equaled between 10 and 100 kbps. On the other hand, the TSR models were assumed to be positioned on the surface of the water as floating nodes that provided a link between TERs in the air/on land and under water. Communication between TSRs was modeled using ad-hoc communications, with arrows appearing when a link exists between individual peer nodes.

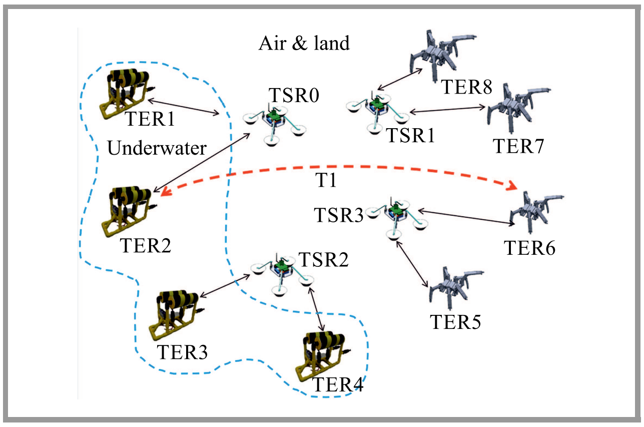


Fig. 8. Analysis notification for simulation purposes and analysis in the modeled RLN.

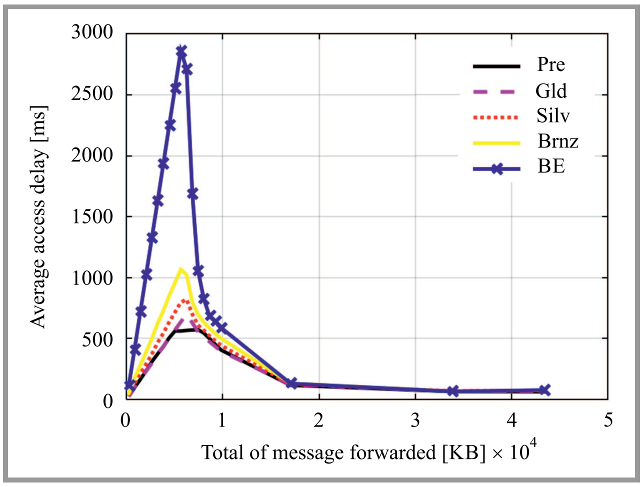


Fig. 9. Average access delay versus total of data packet streaming on T1.

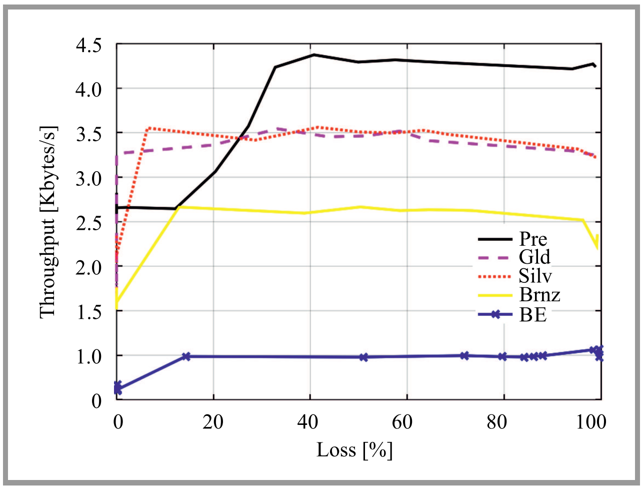


Fig. 10. Throughput versus percentage of data packet loss on T1.

The results for T1 data forwarding performance are presented in Figs. 9–12, with the data packet size randomly generated between 250 bytes and 1.5 Kbytes per packet. The results show a different performance of network parameters for different CoS: premium CoS (Pre), best effort (BE), and Olympic CoS members, namely gold CoS (Gld), silver (Silv) and bronze CoS (Brnz). Overall, Pre is the leader in terms of the performance of the majority of parameters.

Figure 9 shows that the average access delay for Pre was the lowest among all forwarding CoS, even though the BandAS channel bitrate was the same as that of the Gld channel, as shown in Table 1. Generally, the percentage of BE forwarding losses was the highest, whereas Pre forwarding was the lowest in both forwarding states (request and reply). This makes the throughputs for Pre CoS the highest for T1 communication, as illustrated in Fig. 10. In this simulation, other TERs were run to communicate with each other through the shortest path TSRs to provide traffic disturbance (to make the traffic randomly busy).

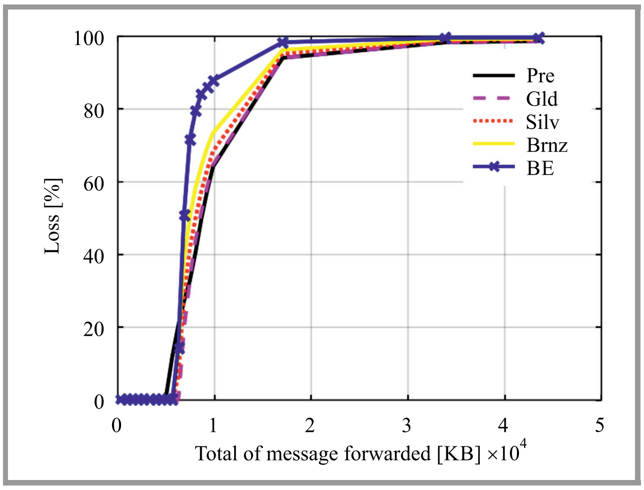


Fig. 11. Data packet loss percentage versus total data streamed on T1.

As far as the data packet loss percentage rate is concerned, Olympic CoS and Pre show similar growth trends and are shown in Fig. 11. BE shows the highest rate of data packet loss compared to other scenarios. The results are most evident when the overall data streaming rate on T1 was between 5000 and 20,000 KB. In this case, the loss percentage rate for BE is by about 61% higher than in the case of other CoS forwarding methods. Before a loss occurs (total of data streaming rate $<$ 35000 KB), access delay of Pre forwarding was by about 10% lower than in each of Olympic CoS forwarding members, and by almost 100% lower than in the default forwarding or BE. The differences in the performance of individual CoS, with increasing data forwarding rates in T1, are shown in Figs. 9 and 11.

As far as the comparison between access delay and data packet loss, as shown in Fig. 12 is concerned, BE forwarding shows a considerable decrease, as data packet loss rate approaches 100%. However, the average access delay of Pre forwarding is still the lowest and within the acceptable range as the value continues to drop with the increasing data packet loss. Similar results can be observed in the case of Olympic CoS members, where the performance of Gld data forwarding, in terms of access delay, was the best compared to Silv and Brnz, when the data packet loss rate increased. The results show that both Pre and Olympic CoS have minimal differences in data distribution and switching (Fig. 12). The throughputs to BE CoS forwarding started to decrease when the overall data streaming rate on T1 increased to more than 300,000 KB.

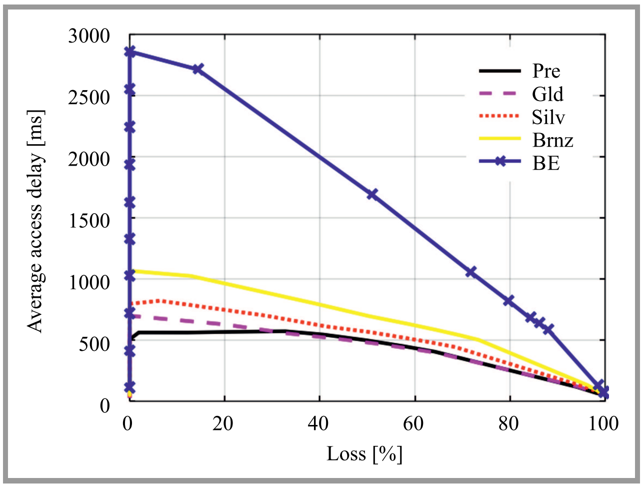


Fig. 12. Average access delay versus data packet loss percentage on T1.

4. Conclusion

The proposed TQS protocol in RLN was modeled and verified. The priority control via CoS was presented, with prioritization in data routing and switching deployed to achieve different performance outcomes. Pre shows the ability to control the average access delay, although the data packet loss decreased with the increase of the volume of data, since it is the highest priority CoS as far as minor access delay, high throughput and small data packet loss are concerned. Such a prioritization excludes the factor of different mediums used. Pre and Gld show good reliability for video and voice data forwarding, with the ability to control the number of data packets lost and to achieve a low average access delay, even though the network traffic is busy with handling the number of data packets that increases along with simulation time. The research will continue with the implementation of the proposed TQS protocol in swarm robot RLNs.

Acknowledgment

This study and research project is funded under the Universiti Malaysia Pahang (UMP) Research Grant for the Centre for Earth Resources Research & Management (CERRM) of UMP (RDU140351).

References

[1] G. Stamatescu, D. Popescu, and R. Dobrescu, “Cognitive radio as solution for ground-aerial surveillance through WSN and UAV infrastructure”, in *Proc. of the 6th Int. Conf. on Electron., Comp. and Artif. Intell. ECAI 2014*, Bucharest, Romania, 2014, pp. 51–56 (doi: 10.1109/ECAI.2014.7090210).

[2] B. P. Gerkey and M. J. Matarić, "A formal analysis and taxonomy of task allocation in multi-robot systems", *The Int. J. of Robotics Research*, vol. 23, no. 9, pp. 939–954, 2004 (doi: 10.1177/0278364904045564).

[3] G. A. Korsah, A. Stentz, and M. B. Dias, "A comprehensive taxonomy for multi-robot task allocation", *The Int. J. of Robotics Research*, vol. 32, no. 12, pp. 1495–1512, 2013 (doi: 10.1177/0278364913496484).

[4] L. Zhang, H. Zhong, and S. Y. Nof, "Adaptive fuzzy collaborative task assignment for heterogeneous multirobot systems", *Int. J. of Intelligent Systems*, vol. 30, no. 6, pp. 731–762, 2015 (doi: 10.1002/int.21725).

[5] C. Cheng, X. Y. Yu, L. L. Ou, and Y. K. Guo, "Research on multi-robot collaborative transportation control system", in *Proc. Chinese Control and Decision Conf. CCDC 2016*, Yinchuan, China, 2016, pp. 4886–4891 (doi: 10.1109/CCDC.2016.7531868).

[6] N. F. Bano, T. Roppel, and I. Gokhale, "Use of mobility models for communication in collaborative robotics", in *Proc. 42nd Southeastern Symp. on System Theory SSST 2010*, Tyler, TX, USA, 2010, pp. 143–146 (doi: 10.1109/SSST.2010.5442846).

[7] E. Kulla *et al.*, "Real World Emergency Scenario Using MANET in Indoor Environment: Experimental Data", in *Proc. 9th Int. Conf. on Complex, Intell., and Software Intensive Syst. CISIS 2015*, Blumenau, Brazil, pp. 336–341 (doi: 10.1109/CISIS.2015.49).

[8] P. E. Rybski, S. A. Stoeter, M. Gini, D. F. Hougen, and N. P. Papanikolopoulos, "Performance of a distributed robotic system using shared communications channels", *IEEE Transactions on Robotics and Automation*, vol. 18, no. 5, pp. 713–727, 2002 (doi: 10.1109/TRA.2002.803460).

[9] Q. Xiao, J. Baojun, D. Xingcheng, Z. Xiaotao, D. Yang, and L. Hui, "A robot remote control system based on VPN and TCP/IP protocol", in *Proc. IEEE Int. Conf. on Mechatron. and Autom. ICMA 2008*, Takamatsu, Japan, 2008, pp. 285–289 (doi: 10.1109/ICMA.2008.4798767).

[10] K. Kyung Jin, S. Il Hong, K. Sung Hoon, and O. Sang Rok, "A novel real-time control architecture for internet-based thin-client robot; simulacrum-based approach", in *Proc. IEEE Int. Conf. on Robot. and Autom.*, Pasadena, CA, USA, 2008, pp. 4080–4085 (doi: 10.1109/ROBOT.2008.4543838).

[11] E. Rosen, A. Viswanathan, and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, IETF, 2001 [Online]. Available: https://tools.ietf.org/html/rfc3031

[12] OMNeT++ Discrete Event Simulator [Online]. Available: https://omnetpp.org

[13] Z. Zeng, S. Fu, H. Zhang, Y. Dong, and J. Cheng, "A survey of underwater optical wireless communications", *IEEE Commun. Surveys & Tutorials*, vol. 19, no. 1, pp. 204–238, 2017 (doi: 10.1109/COMST.2016.2618841).

[14] J. Lloret, S. Sendra, M. Ardid, and J. J. P. C. Rodrigues, "Underwater wireless sensor communications in the 2.4 GHz ISM frequency band", *Sensors*, vol. 12, no. 4, p. 4237–4264, 2012 (doi: 10.3390/s120404423).

[15] D. Pompili and I. F. Akyildiz, "Overview of networking protocols for underwater wireless communications", *IEEE Communications Magazine*, vol. 47, no. 1, pp. 97–102, 2009 (doi: 10.1109/MCOM.2009.4752684).

[16] M. Dautta and M. I. Hasan, "Underwater vehicle communication using electromagnetic fields in shallow seas", in *Proc. Int. Conf. on Elec., Comp. and Commun. Engin. ECCE 2017*, Cox's Bazar, Bangladesh, 2017, pp. 38–43 (doi: 10.1109/ECACE.2017.7912875).

[17] H. Yoshida *et al.*, "Study on land-to-underwater communication", in *Proc. 14th Int. Symp. on Wirel. Personal Multim. Commun. WPMC 2011, 2011*, Brest, France, 2011, pp. 1–5.

[18] B. T. Doshi, R. Nagarajan, G. N. S. Prasanna, and M. A. Qureshi, "Future WAN architecture driven by services, traffic volume, and technology trends", *Bell Labs Tech. J.*, vol. 6, no.1, pp. 13–32, 2001 (doi: 10.1002/bltj.2261).

[19] M. C. Chuah, K. Medepalli, S. Y. Park, and J. Wang, "Quality of service in third-generation IP-based radio access networks", *Bell Labs Tech. J.*, vol. 7, no. 2, pp. 67–89, 2002 (doi: 10.1002/bltj.10006).

[20] O. Klopfenstein, "Rerouting tunnels for MPLS network resource optimization", *Eur. J. of Operat. Res.*, vol. 188, no. 1, pp. 293–312, 2008 (doi: 10.1016/j.ejor.2007.04.016).

[21] S. Ricciardi, F. Palmieri, A. Castiglione, and D. Careglio, "Energy efficiency of elastic frequency grids in multilayer IP/MPLS-over-flexgrid networks", *J. of Netw. and Comp. Appl.*, vol. 56, pp. 41–47, 2015 (doi: 10.1016/j.jnca.2015.06.014).

[22] M. Fathy, S. GholamalitabarFirouzjaee, and K. Raahemifar, "Improving QoS in VANET Using MPLS", *Procedia Comp. Science*, vol. 10, pp. 1018–1025, 2012 (doi: 10.1016/j.procs.2012.06.141).

[23] Z. Song, P. W. C. Prasad, A. Alsadoon, L. Pham, and A. Elchouemi, "Upgrading Internet service provider (ISP) network in multiprotocol label switching (MPLS) and border gateway protocol (BGP) environment", in *Proc. Int. Conf. on Adv. in Elec., Electron. and Syst. Engin. ICAEES 2016*, pp. 237–241, Putrajaya, Malaysia, 2016 (doi: 10.1109/ICAEES.2016.7888045).

[24] B. Genge and C. Siaterlis, "Analysis of the effects of distributed denial-of-service attacks on MPLS networks", *Int. J. of Crit. Infrastruc. Protect.*, vol. 6, no. 2, pp. 87–95, 2013 (doi: 10.1016/j.ijcip.2013.04.001).

[25] M. N. Soorki and H. Rostami, "Label switched protocol routing with guaranteed bandwidth and end to end path delay in MPLS networks", *J. of Netw. and Comp. Appl.*, vol. 42, pp. 21–38, 2014 (doi: 10.1016/j.jnca.2014.03.008).

[26] G. A. Mazhin, M. Bag-Mohammadi, M. Ghasemi, and S. Feizi, "Multi-layer architecture for realization of network virtualization using MPLS technology", *ICT Express*, vol. 3, no. 1, pp. 43–47, 2017 (doi: 10.1016/j.icte.2016.07.002).

[27] F. Francois, N. Wang, K. Moessner, S. Georgoulas, and R. de Oiveira Schmidt, "Leveraging MPLS backup paths for distributed energy-aware traffic engineering", *IEEE Trans. on Netw. and Serv. Manag.*, vol. 11, no. 2, pp. 235–249, 2014 (doi: 10.1109/TNSM.2014.2321839).

[28] L. Andersson, A. B. Acreo, and R. Asati, "Multiprotocol Label Switching (MPLS) Label Stack Entry: "EXP" Field Renamed to "Traffic Class" Field", RFC 5462, IETF, 2009 [Online]. Available: https://tools.ietf.org/html/rfc5462

[29] L. Wu *et al.*, "Multi-Protocol Label Switching (MPLS) Support of Differentiated Services", RFC 3270, IETF, 2002 [Online]. Available: https://tools.ietf.org/html/rfc3270

[30] N. Rouhana and E. Horlait, "Differentiated services and integrated services use of MPLS", in *Proc. 5th IEEE Symp. on Comp. and Commun. ISCC 2000.*, Antibes-Juan Les Pins, France, 2000, pp. 194–199 (doi: 10.1109/ISCC.2000.860638).

[31] J. Heinanen, F. Baker, W. Weiss, and J. Wroclawski, "Assured Forwarding PHB Group", RFC 2597, IETF, 1999 [Online]. Available: https://tools.ietf.org/html/rfc2597

[32] R. Pletka, P. Droz, and B. Stiller, "A Buffer-Management Scheme for Bandwidth and Delay Differentiation Using a Virtual Scheduler", in *Networking – ICN 2001: First International Conference on Networking Colmar, France, July 9–13, 2001 Proceedings, Part I*, P. Lorenz, Ed. *LNCS*, vol. 2093, pp. 218–234. Berlin, Heidelberg: Springer, 2001.

[33] R. Balakrishnan, *Advanced QoS for Multi-Service IP/MPLS Networks*. Wiley, 2012 (ISBN: 9781118621479).

Addie Irawan has been a Senior Lecturer at the Universiti Malaysia Pahang (UMP), Pahang, Malaysia since 2005, working at the Faculty of Electrical and Electronic Engineering (FKEE). His areas of interest include robotics, dynamic and motion control, as well as computers and networks. He specializes in network protocols and big data distributions. He is a Professional Engineer of the Board of Engineers Malaysia (BEM), a Charted Engineer and a Charted Marine Engineer under the British Engineering Council via the Institute of Marine Engineering Science and Technology (IMarEST), as well as a Senior Member of the Institute of Electrical and Electronics Engineers (IEEE).

E-mail: addieirawan@ump.edu.my
Robotics and Unmanned Systems (RUS) research group
Faculty of Electrical and Electronics Engineering
Universiti Malaysia Pahang
26600 Pekan, Pahang, Malaysia

Mohammad Fadhil Abas received his B.Sc. in Electric, Electronic and Computer Systems form Universiti Kebangsaan Malaysia, Malaysia in 2002, his M.Sc. in Power Electronics from University Putra Malaysia, Malaysia in 2005 and his Ph.D. in Artificial Science from Chiba University, Japan in 2013. Since 2004 he has been working at Universiti Malaysia Pahang (UMP), as a Senior Lecturer at the Faculty of Electrical and Electronic Engineering. In his work, he focuses mainly on the following areas: instrumentation, human-computer interface, human-robot interface, unmanned systems, automation, vision systems for robotics and control systems.

E-mail: mfadhil@ump.edu.my
Robotics and Unmanned Systems (RUS) research group
Faculty of Electrical and Electronics Engineering
Universiti Malaysia Pahang
26600 Pekan, Pahang, Malaysia

Nurulfadzilah Hasan received her B.Sc. degree in Computer Engineering from Universiti Teknologi Malaysia in 2003. She received her M.Sc. degree in Electrical Engineering from the same university in 2005. She has been working as a lecturer at Universiti Malaysia Pahang since 2006, where she has been involved in several research projects and has published numerous research papers. Her research areas include antenna design, wireless communications and computer engineering. Presently, she is pursuing her Ph.D. in antenna design at Universiti Malaysia Pahang.

E-mail: nurulfadzilah@ump.edu.my
Applied Electronics (AE) research group
Faculty of Electrical and Electronics Engineering
Universiti Malaysia Pahang
26600 Pekan, Pahang, Malaysia

Paper

Fuzzy Clustering with Multi-Constraint QoS Service Routing in Wireless Sensor Networks

Jayashree Agarkhed, Vijayalaxmi Kadrolli, and Siddarama R. Patil

P.D.A. College of Engineering, Kalaburagi, India

https://doi.org/10.26636/jtit.2019.127818

Abstract—This paper presents a fuzzy logic-based, service differentiated, QoS aware routing protocol (FMSR) offering multipath routing for WSNs, with the purpose of providing a service differentiated path meant for communication between nodes, based on actual requirements. The proposed protocol initially forms a cluster by fuzzy c-means. Next, the building of a routing follows, so as to establish multiple paths between nodes through the modified QoS k-nearest neighborhood, based on different QoS constraints and on optimum shortest paths. If one node in the path fails due to lack of residual energy, bandwidth, packet loss, delay, an alternate path leading through another neighborhood node is selected for communication. Simulation results show that the proposed protocol performs better in terms of packet delivery ratio, delay, packet drop ratio and throughput compared to other existing routing protocols.

Keywords—fuzzy logic, QoS, routing, WSN.

1. Introduction

Energy efficient routing is the main objective of wireless sensor networks (WSNs). In WSNs, sensor nodes collaborate with each other by communicating with neighboring nodes. They also perform basic computations based on the data collected and complete different tasks, such as neighborhood node discovery, smart sensing and optimal efficient routing – at every layer. The routing protocols are classified in terms of QoS aware protocols and performance [1]–[3]. In order to provide QoS in the applications, in most of the cases fuzzy logic-based selection of cluster head is used in the course of the routing process, which provides a non-probabilistic approach with two fuzzy variables: one is base station distance and the other is residual energy of the sensor nodes.

Multi-hop communication is used for the selection of cluster head (CH). This has the authority to communicate with other CHs and with BS. Various methods are used to identify the next forwarding node. The selection of nodes is based on different techniques, such as fuzzy logic, neuro fuzzy and the mobility of nodes [4].

Flooding is also used to set up possible routes to destinations which rely on bandwidth, node energy or link quality. As a consequence, these strategies may lead to unnecessary message transmissions, network jamming, longer delays and loss of packets. To avoid these problems, it is essential to come across the optimal path between nodes using the existing resources in the network. However, attempting to choose a route that satisfies many constraints may result in conflicts and the process may be complicated. Therefore, it is recommended to deploy multi-metrics in WSNs, with path and packet communication based on differentiated services [5].

An effective, optimal, multipath, service-differentiated routing protocol, known as the fuzzy-based service-differentiated QoS-aware routing protocol (FMSR), is proposed, which initially forms a cluster by fuzzy c-means and uses multiple metrics, such as link bandwidth, residual energy, packet loss and delay to choose the neighborhood nodes. Multiple paths are subsequently established between the source and the destination, leading through these neighborhood nodes, by means of the k-nearest neighborhood method, thus forming an optimal route for differentiated services. If a neighborhood node along the path fails due to lack of bandwidth or energy, an alternative path is established.

In Section 1 an introduction to the paper is presented. Section 2 contains a short survey of the existing routing protocols. The plan of the proposed work is described in Section 3. Finally, performance, simulation results and conclusions are summarized in Sections 4 and 5.

2. Related Work

Several studies have been conducted to attempt, with varying degrees of success, to address the problem of energy-efficient, delay-constrained routing in WSNs and multiple metrics are used for routing, considering link rate and packet loss, i.e. [6].

Soft computing methods have been truly helpful in a variety of areas and have shown capable outputs. Novel clustering algorithms are based on the fuzzy c-means concept, where the selection of the cluster head is based on its proximity to the middle of the cluster and to the node having the highest residual energy. Non-cluster head nodes broadcast sensed data to the cluster head perform data aggregation and transmit data straight to the base station [7], [8].

Fuzzy logic control, known as BOKHARI-SEPFL, based on distance of nodes form the base station, density of nodes and energy level, as well as traditional threshold values are used to enhance the process of cluster head election to improve the lifetime and throughput of the WSN [9].

An energy efficient adaptive routing is proposed in the form of a fuzzy-based clustering protocol that makes use of the direct transmission technique, depending upon the criticality and the location of sensor nodes [10].

Mamdani's fuzzy inference system is used to identify the ability of a sensor node to become a group cluster head, depending on the distance of the input parameter and the energy of the sensing nodes. The heuristic search algorithm is used to find the minimum path length from the source to the receiving node. The aggregated data packets are routed from the originating CH to the receiving node along the selected route [11].

Energy-aware routing protocols have been proposed for WSNs. Most of them are energy savers and there is not much focus on energy balancing. Though, the lifetime of WSNs severely depend energy use; so, energy management is a necessary job to be considered [11], [12]. The energy aware routing protocol – FEAR which balances energy and energy saving, is considered. It shows an appropriate trade-off between the saving of energy and energy balancing by a fuzzy set scheme. Based on examination of energy expenditure for the data transceiver, a single-hop forwarding system is proved to provide less energy than multi-hop forwarding [13], [14].

The main advantage of the fuzzy logic control-based QoS management (FLC-QM) method consists in changing traffic load. It utilizes a fuzzy logic controller, relying on the source sensor node to get the sampling period and the deadline miss ratio for the transmission of data from the sensor to the actuator [15], [16].

A protocol has been introduced that exchanges roles between regular nodes and cluster heads in a round robin manner, following the token ring methodology. The equidistribution of cluster head burden over all sensors in the same cluster reduces the need of expensive periodic re-clustering. The domain memberships of edge sensors are handled through fuzzy logic, based on the residual energy [17].

The rumor routing algorithm in WSNs allows the query source to distribute the query to identify a source which helps get a timely query message, but energy efficiency is improved by relying on the hierarchical clustering formation method, while the fuzzy logic method is a used to increase network efficiency [18].

Intelligent multipath routing has been used, which uses fuzzy stochastic multipath routing (FSMR) for providing hop count, battery power and signal strength. Nodes are stochastically forwarded with path selection, which results in automatic load balancing and fault tolerance [19], [20].

Geographic opportunistic routing (GOR) used to provide QoS with end-to-end reliability and delay restrictions in WSNs, for different opportunities, has been proposed to ensure multi control service quality in WSNs – a problem that may be formulated as one involving multi-objective optimization, i.e. selection and prioritization of a set of candidates for efficient forwarding. The solution is suitable for WSN in terms of efficiency energy, latency and temporal complexity [21].

An adaptive multi-constraint multipath routing protocol which minimizes loss rate, energy consumption and delay between clusters, based on a weighted cost function and on such parameters as loss rate, residual energy and delay, is presented in [18]. An approach to calculate approximately probabilistic timeliness guarantees end-to-end communication delivery delays in WSNs and is used at run-time to build a metric which estimates the probability density function of the end-to-end latency of a path [22]. In fuzzy stochastic multipath routing (FSMR), multiple metrics are used to determine hop count, battery power, signal strength and fuzzy logic is used to offer multiple optimal paths [23]. A novel relative mobility metric for mobile ad hoc networks (MANETs), which is based on power level ratios changing at every node due to consecutive receptions from its neighbors, is addressed in [24].

3. System Architecture and Methodology

The proposed work identifies multiple paths between nodes, leading through candidate nodes, based on different QoS constraints. The proposed architecture is shown in Fig. 1. It comprises 3 parts:

- cluster formation by using fuzzy c-mean,
- optimal shortest routing by using k-nearest neighborhood method,
- providing service differentiations based on service requirements.

QoS nodes based on different constraints are known as candidate nodes and are chosen from the CH set. These nodes are chosen based on residual energy, bandwidth, packet loss and delay. The distance of the CH from the sink is essential for energy efficiency and is crucial for balancing energy spending and network lifetime. Hence, the formation of clusters between sensor nodes is given priority. The fuzzy c-means clustering approach is used in cluster formation to determine the set of k clusters in d-dimensional space. In the network structure, each sensor node has maintained a routing table with neighbor node distance, residual energy, bandwidth and packet loss between nodes. Before starting any operation, each sensor node initializes each sensor by exchanging the routing table. After initiating, each node starts the transmission phase. Before the transmission, it checks the required bandwidth, delay, residual energy and packet loss by comparing it with the threshold value set. The routing path is a set of applicant nodes based on different QoS parameters. If any QoS applicant

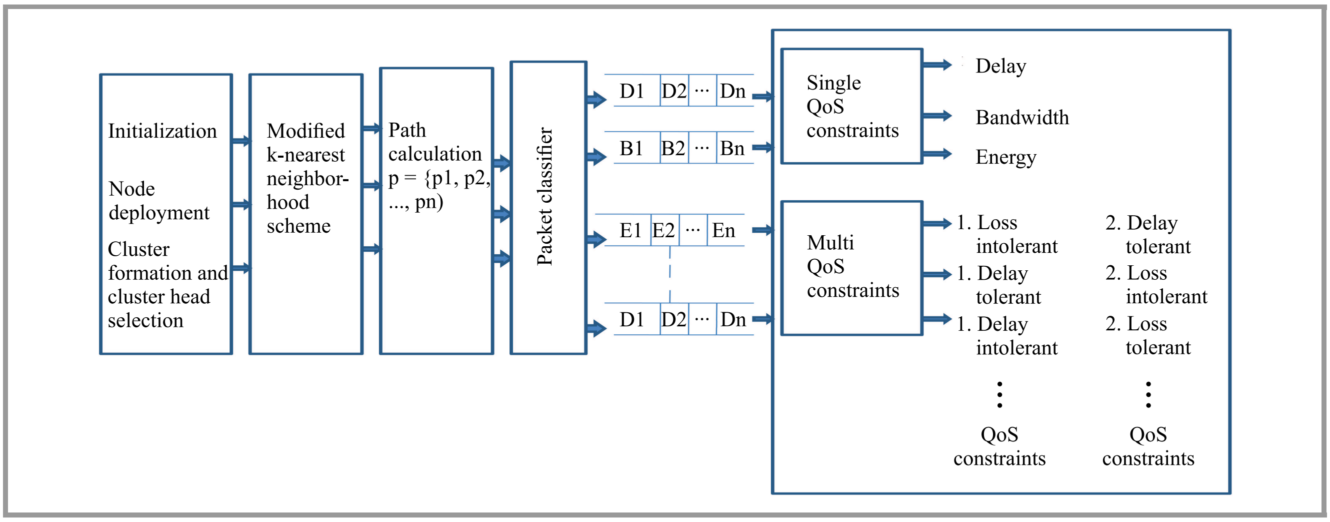


Fig. 1. System architecture.

node in the path fails because it does not meet the requirements related to bandwidth, residual energy, packet loss, an alternate path is established.

In the initialization phase, every sensor node telecasts a hello message to neighboring nodes, keeping in mind that the end goal is to have enough high quality information. Every sensor node maintains and updates a table with neighboring nodes during this stage. The table contains a list of the sensor node's neighboring nodes. While establishing each path, the sensor node sends a hello packet to another node. If the parameters are met, then the path is established (Fig. 2).

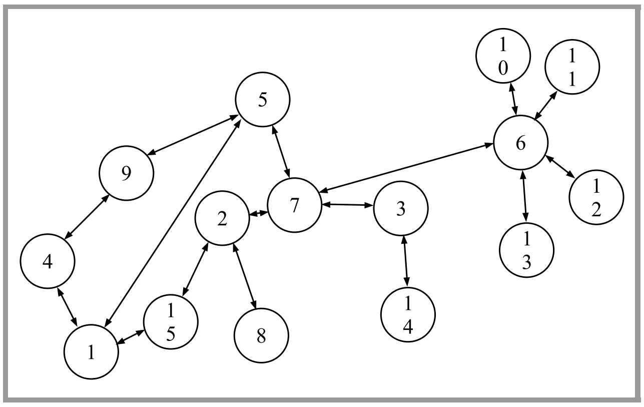


Fig. 2. Route initialization phase.

Next, cluster formation proceeds and the fuzzy c-means clustering algorithm is applied. It uses the information about its members to select the right option. Then, cluster head formation is calculated and the cluster head is selected for each cluster.

In the neighborhood scheme, a modified QoS k-neighborhood algorithm is used. The distance between CH and sink is calculated by the Euclidian distance (Algorithm 1). This demonstrates how to route data information based

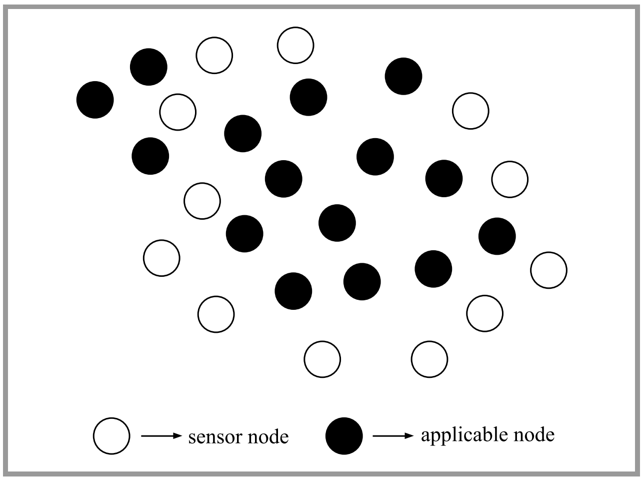


Fig. 3. Applicable cluster node selection.

on QoS parameters: residual energy, delay, bandwidth and packet loss. Within the network structure, each sensor node maintains a routing table that contains neighbor node distance, residual energy, bandwidth and packet loss experienced while communicating with other nodes, as shown in Fig. 3. Before commencing any operation, each censor node is initialized, i.e. it exchanges the routing table. After initiating the transmission phase, but before the transmission, it checks the minimum bandwidth, delay, residual energy and packet loss.

To validate the routing, four metrics are used:

- **Bandwidth**. When a node intends to transport data, it has to be conscious of the local bandwidth and the interference and transmission range of the neighboring nodes. Therefore, the node needs to monitor the channel and estimate local bandwidth (LBW) which depends on the idle-to-busy time ratio. Residual energy is computed by the node over a given period of time.

- **Packet loss rate (PLR)**. The packet loss rate is calculated as the ratio between the amount of data packets received and the total amount of data packets sent.

- **Delay** – is the time difference between the time when the packet was sent from source and he time when it arrives at the other node.

Algorithm 1 : Modified QoS k-nearest neighbor

```
Let $(X_i, C_i)$ where $i = 1, 2, \ldots, n$ be data points. Let $x$ be a point
related count class.
Calculate using k-nearest neighbor along with QoS parameters:
Step 1.  Compute $d(x, x_i)$, $i = 1, 2, \ldots, n$, where $d$ represents
         the Euclidean distance between the points along
         with the QoS parameter
Step 2.  Arrange the calculated $n$ Euclidean distance form
Step 3.  Let $k$ be a positive integer, take the first $k$ distances
         from this sorted list based on the QoS parameters
Step 4.  Find individuals $k$-points matching to these $k$-distances
         and QoS parameter considerations
Step 5.  Let $k_i$ denotes the number of points belonging to
         the $i$-th class among $k$ points, i.e. $k \geq 0$. Consider
         those nodes which satisfy the QoS based on a single
         constraint and multiple constraints
```

In the next step service differentiations are provided based on service requirements. Packet and path classifiers are determined as well (Fig. 4).

Packets are classified based on QoS constraints. If packets are delay sensitive, then they are sent to the delay sensitive path. If packets are bandwidth sensitive, then they are sent to the bandwidth sensitive path. If packets are energy sensitive, then they are sent to the energy sensitive path.

The best available path matching the service type is considered for routing the data. The path with the lowest energy consumption, delay, bandwidth and packet loss is considered to be the optimal path.

3.1. Path Discovery Phase

When a source node intends to broadcast a data packet to a target node, it initiates the multipath routing detection process between the source and the destination. This initiates the routing process. The source node must first check its routing table to determine whether the routing table contains information. If a route is established, the source will make use of the route to send the data packet instantly, or else, the source node will broadcast a route request (RREQ) packet.

The state of d node indicates whether the node is designated as a candidate node or a non-candidate node. When a node receives an RREQ packet, it will forward the packet to all its neighbors. When an intermediate node receives an RREQ packet, if it has previously received an RREQ packet with a similar series number and destination ID, it drops the unnecessary RREQ packet. When an intermediate node receives an RREQ packet for the first time, it updates its routing table with the source ID and destination ID and the previous hop node ID and its state and appends its state to the RREQ message in the node state field and analyzes the destination ID. When the destination node receives the RREQ message, it appends its state to the route reply (RREP) and unicasts the reply message on the reverse path toward the source. The destination performs this action for every RREQ it receives. At the same time, an intermediate node receives the RREP message, it appends its state to the message, updates its routing table and unicasts the RREP in the direction of the source using the formerly stored hop node information. The source chooses a path of applicable nodes to transmit the data packet.

3.2. Data Transmission Phase

Figure 5 shows the optimal path's source (thin line, path 3) and destination nodes, indicated by a square box. The ser-

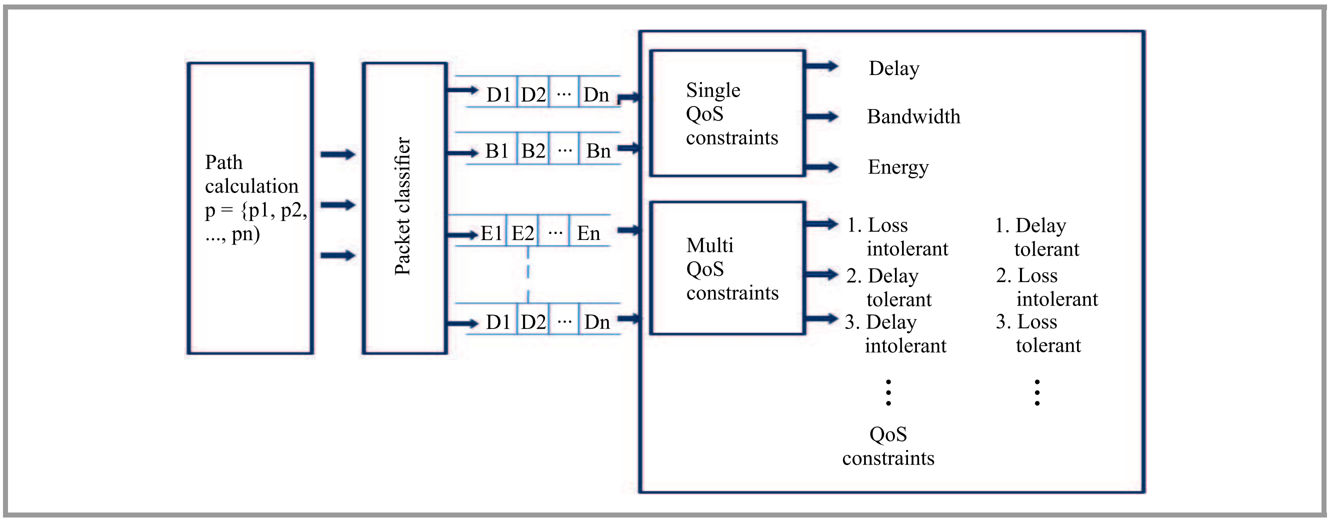


Fig. 4. Packet and path classifier.

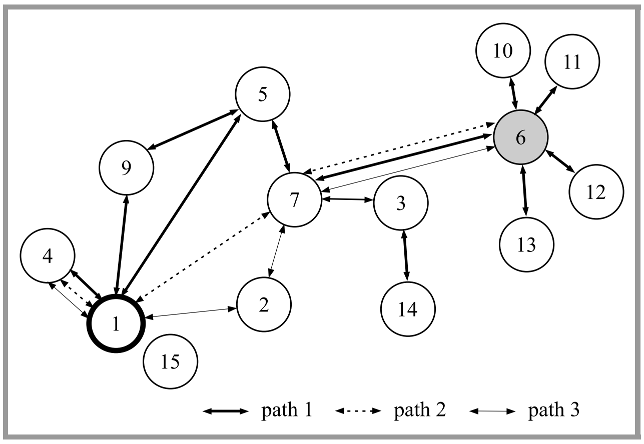


Fig. 5. Path discovery.

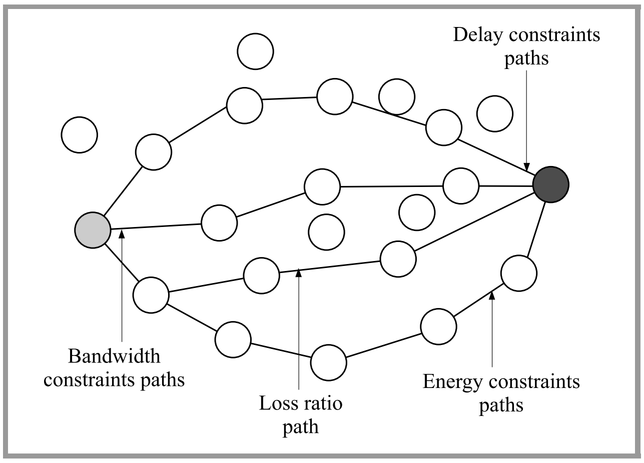


Fig. 6. Service differentiation paths.

vice differentiation paths are shown in Fig. 6. After receiving transferring of sense data it represented output results by the graph. For example, in the figure shown, node 6 wants to send data to node then it will find a destination node using multipath paths are For example in the figure, node 6 wants to send data to node then it will find a destination node using multipath paths are:

- 1.6-7-5-9-1 – in this path, the hop count is larger, it consumes more energy and node 9 is busier than in other paths,
- 2.6-7-1-4 – in this path the hop count is lower, less energy is consumed and BW is greater,
- 3.6-7-2-1-4 – in this path the hop count is higher, more energy is consumed and BW is lower compared to path 6.

Checking all parameters, select the optimized path used to transmit data from the source to the destination.

4. Simulation System and Parameters

A comparison of the proposed protocol with existing protocols is performed using the NS-2 network simulator [25]. The simulation models use a network of 100 nodes in a 500 × 500 m region, with the number of sensor nodes varying from 0 to 100. The average of 10 runs has been performed for simulation purposes. Each and every node randomly selects a position and moves in the direction of that position. Once the node arrives at the position, it stays there for a predefined period of time. After that time, it selects a new position and repeats the process. The simulations run lasts for 200 s.

The evaluation of the proposed protocol is based on relevant parameters - the number of packets dropped, delay, bandwidth and hop count.

4.1. Comparative Analysis

Figure 7 shows a comparison between the packet delivery ratio (PDR) in the proposed network and in the existing protocol [14], with varying times. As the time increases, the number of packets delivered increases in the existing protocol. The existing protocol's FSMR chooses the optimal route for transmission and finds an alternate path through the applicable nodes when the path breaks. When applicable nodes are selected, no other nodes are involved and they are sent to sleep mode. Consequently, it offers better PDR in terms of different QoS constraints.

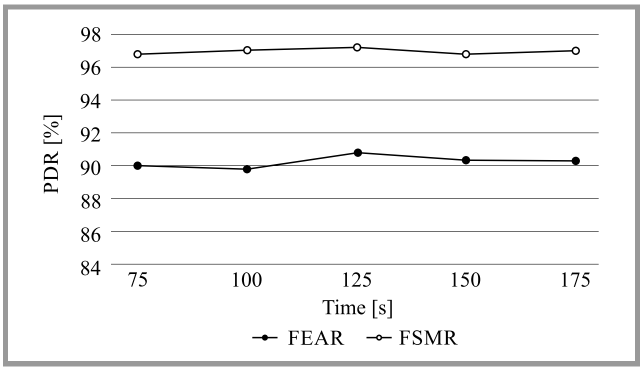


Fig. 7. PDR versus time.

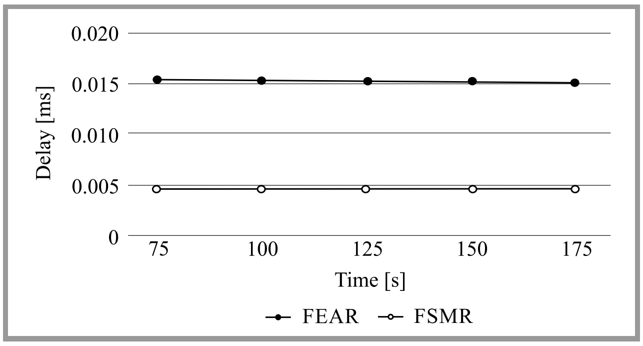


Fig. 8. Delay versus time.

Figure 8 shows that the delay in FSMR will be lower than in FEAR, because of the fact that applicable nodes have already considered the delay parameter, and if the link breaks, another applicable node will be chosen to establish the route. FSMR offers also a better service based on QoS constraints, and a separate path with a specific delay is maintained. It checks whether the QoS delay parameter is satisfied or not for each link. It separates all available paths only by considering the most prominent delay, and selects one optimal path based on the delay constraint and the shortest path to transmit the packet.

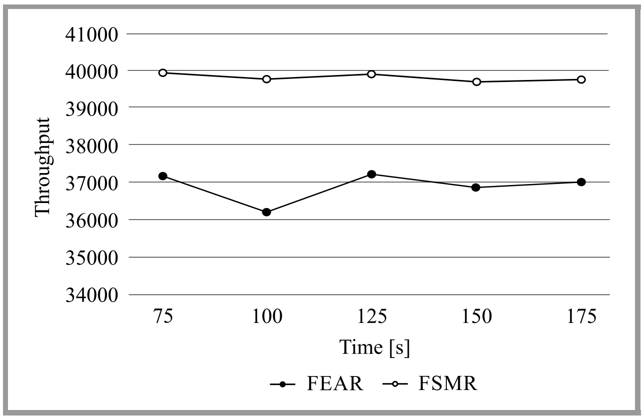


Fig. 9. Throughput versus time.

Figure 9 shows that the delay of FEAR increases as the time increases, because the longer time may deplete more energy. This may lead to packet loss, buffer overflow, degradation in throughput and frequent route breaks. Therefore, it results in a longer end-to-end delay. FSMR achieves the best path to transmit the data from the source to the designation before transmitting any packets, so the link selecting process has to be performed. The selected path should have a lower mean end-to-end delay.

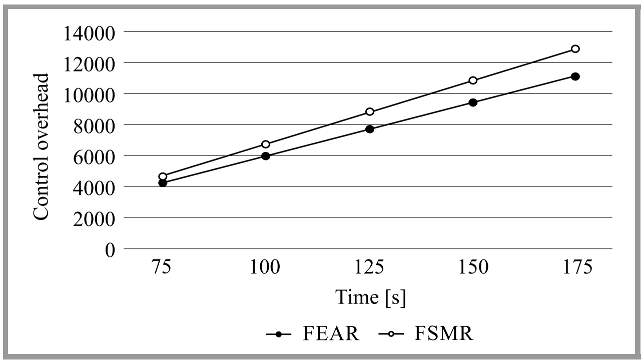


Fig. 10. Control overhead versus time.

Figure 10 shows the behavior of the proposed protocol according to the control overhead. It shows the packets that are sent and received with the use of a specific route. A different simulation time has been adopted here for each of the iterations, while the characteristics (initial power, node distribution and distance from sink) remain the same. The average of 10 simulation runs have been conducted to evaluate performance.

The overall control overhead is reduced for FSMR, because adopting fuzzy logic, and also by using the selected nodes only to establish the route for servicing the packets for particular services.

5. Conclusion

The purpose of the proposed protocol is to find the optimized path between nodes, using relevant nodes. The node selection process is performed using the modified QoS k-nearest neighborhood technique. Simulation performed in the NS2 simulator shows that the proposed technique reduces the delay, as well as increases the packet delivery ratio and throughput.

References

[1] S. Hasan, Z. Hussain, and R. K. Singh, "A survey of wireless sensor network", *Int. J. of Emerg. Technol. and Adv. Engin.*, vol. 3, no. 3, pp. 487–492, 2013 [Online]. Available: https://ijetae.com/files/Volume3Issue3/IJETAE_0313_83.pdf

[2] M. Asif, S. Khan, R. Ahmad, M. Sohail, and D. Singh, "Quality of service of routing protocols in wireless sensor networks: A survey", *IEEE Access*, vol. 5, pp. 1846–1871, 2017 (doi: 10.1109/ACCESS.2017.2654356).

[3] V. Kadrolli and J. Agarkhed, "Soft computing routing techniques in wireless sensor network", in *Proc. 2nd Int. Conf. on Adv. in Electric., Electron., Inform., Commun. and Bio-Inform. AEEICB 2016*, Chennai, India, 2016, pp. 748–751 (doi: 10.1109/AEEICB.2016.7538395).

[4] S. Boukerche *et al.*, "Routing protocols in ad hoc networks: A survey", *Comp. Networks*, vol. 55, no. 13, pp. 3032–3080, 2011 (doi: 10.1016/j.comnet.2011.05.010).

[5] F. Ahamad and R. Kumar, "Energy efficient region based clustering algorithm for WSN using fuzzy logic", in *Proc. IEEE Int. Conf. on Recent Trends in Electron., Inform. & Commun. Technol. RTEICT 2016*, Bangalore, India, 2016, pp. 1020–1024 (doi: 10.1109/RTEICT.2016.7807984)

[6] T. A. Muthupandian, J. G. Eanoch, and H. Robinson Yesudhas, "A survey on techniques for selection of forwarding node in wireless sensor networks", *Int. J. of Adv. in Comp. and Electron. Engin.*, vol. 2, no. 4, pp. 24–29, 2017.

[7] G. Santhi and A. Nachiappan, "Fuzzy-cost based multi constrained QoS routing with mobility MANETs", *Egyptian Informat. J.*, vol. 13, no. 1, pp. 19–25 2012 (doi: 10.1016/j.eij.2011.12.001).

[8] C. Intanagonwiwat, R. Govindan, and D. Estrin, "Directed diffusion: A scalable and robust communication paradigm for sensor networks", in Proc. of the 6th Ann. Int. Conf. on Mobile Comput. and Network. MobiCom'00, Boston, MA, USA, 2000, pp. 56–67 (doi: 10.1145/345910.345920).

[9] A. Nanda and A. K. Rath, "Mamdani fuzzy inference based hierarchical cost effective routing (MFIHR) in WSNs", in *Proc. IEEE 7th Int. Adv. Comput. Conf. IACC 2017*, Hyderabad, India, 2017, pp. 7–401 (doi: 10.1109/IACC.2017.0089).

[10] K. Singh and R. K. Singh, "An energy efficient fuzzy based adaptive routing protocol for wireless body area network", in *Proc. IEEE UP Section Conf. on Elec. Comp. and Electron. UPCON 2015*, Allahabad, India, 2015, pp. 1–6 (doi: 10.1109/UPCON.2015.7456680).

[11] S. V. Mallapur and S. R. Patil, "Fuzzy logic-based stable multipath routing protocol for mobile ad hoc networks", in *Proc. Ann. IEEE India Conf. INDICON 2014*, Pune, India, 2014, pp. 1–6 (doi: 10.1109/INDICON.2014.7030670)

[12] M. U. Bokhari, "Bokhari-SEPFL routing protocol based on fuzzy logic for WSNs", in *Proc. 5th Int. Conf. on Reliab., Infocom Technol. and Optimiz. ICRITO 2016 (Trends and Future Directions)*, Noida, India, 2016, pp. 38–43 (doi: 10.1109/ICRITO.2016.7784920).

[13] S. Souiki, M. Hadjila, and M. Feham, "Fuzzy based clustering and energy efficient routing for underwater wireless sensor networks", *Int. J. of Comp. Netw. & Commun. (IJCNC)*, vol. 7, no. 2, pp. 33–44, 2015 (doi: 10.5121/ijcnc.2015.7203).

[14] E. Ahvar, A. Pourmoslemi, and M. J. Piran, "FEAR: A fuzzy-based energy-aware routing protocol for wireless sensor networks", arXiv preprint arXiv:1108.2777 [Online]. Available: https://arxiv.org/ftp/arxiv/papers/1108/1108.2777.pdf

[15] H. Jiang, Y. Sun, R. Sun, and H. Xu, "Fuzzy-logic-based energy optimized routing for wireless sensor networks", *Int. J. of Distrib. Sensor Netw.*, vol. 9, no. 8, 2013 (doi: 10.1155/2013/216561).

[16] M. Omari, H. Abdelkarim, and B. Salem, "Optimization of energy consumption based on genetic algorithms optimization and fuzzy classification", in *Proc. 2nd World Symp. on Web Appl. and Network. WSWAN 2015*, Sousse, Tunisia, 2015
(doi: 10.1109/WSWAN.2015.7210317).

[17] F. Xia, W. Zhao, Y. Sun, and Y. C. Tian, "Fuzzy logic control based QoS management in wireless sensor/actuator networks", *Sensors*, vol. 7, no. 12, pp. 3179–3191, 2007 (doi: 10.3390/s7123179).

[18] R. V. Dharaskare and M. M. Goswami, "Intelligent multipath routing protocol for mobile ad hoc network", *Int. J. of Comp. Sci. and Appl.*, vol. 2, 2009, pp. 135–145.

[19] L. Cheng *et al.*, "QoS aware geographic opportunistic routing in wireless sensor networks", *IEEE Trans. on Parallel and Distrib. Syst.*, vol. 25, no. 7, pp. 1864–1875 (doi: 10.1109/TPDS.2013.240).

[20] J. Agrakhed, G. S. Biradar, and V. D. Mytri, "Adaptiv multi constraint multipath routing protocol in wireless multimedia sensor network", in *Proc. Int. Conf. on Comput. Sci.*, Phagwara, India, 2012, pp. 326–331 (doi: 10.1109/ICCS.2012.9).

[21] R. S. Oliver "Estimation of the probability density function of end-to-end delays in wireless sensor networks", Tech. Rep., Technische Universität Kaiserslautern, Kaiserslauter, Germany, Jan. 2009 [Online]. Available: https://rts.eit.uni-kl.de/fileadmin/publication_files/TR09_serna_oliver.pdf

[22] T.-S. Su, C.-H. Lin, and W.-S. Hsieh, "A novel QoS-aware routing for ad hoc networks", in *Proc. of the 9th Joint Int. Conf. on Inform. Sci. JCIS 2006*, Kaohsiung, Taiwan, China, 2006
(doi: 10.2991/jcis.2006.117).

[23] V. Rashiwal, S. Verma, and S. K. Bajpai, "QoS based power aware routing in MANETs", *Int. J. of Comp. Theory and Engin.*, vol. 1, no. 1, pp. 49–54 (doi: 10.7763/IJCTE.2009.V1.8).

[24] P. Basu, N. Khan, and T. D. C. Little, "A mobility based metric for clustering in mobile ad hoc networks", in *Proc. 21st Int. Conf. on Distrib. Comput. Syst. Worksh. ICDCS 2001*, Mesa, AZ, USA, 2001, pp. 413–418 (doi: 10.1109/CDCS.2001.918738).

[25] Network Simulator [Online]. Available:
https://www.isi.edu/nsnam/ns

Jayashree V. Agarkhed is currently working as professor at the Computer Science and Engineering Department of the Poojya Doddappa Appa College of Engineering, Gulbarga, Karnataka, India. She received her BE degree from Gulbarga University, Gulbarga, Karnataka, India, as well as M.Tech. and Ph.D. degrees from Visveshwaraya Technological University, Belagavi, Karnataka, India in 1999, 2003 and 2013, respectively. She supervised numerous BE and M.Tech. projects and guided more than 8 Ph.D. students. She is a lifetime member of the Indian Society for Technical Education, an India member of the Institute of Electrical and Electronics Engineers, a member of the Institute of Electronics and Telecommunication Engineers, India, and a member of the Institute of Engineers, India. Her research interests are in the area of wireless networking with QoS provisioning, as well as scheduling and routing algorithm design in sensor networks, Ad Hoc Networks and cloud computing. She has published more than 125 scientific articles in top-tier journals and conferences. She has also published 2 books. She has chaired various international conferences and is the reviewer for various National and International journals and Conferences. She is the Member of Board of Studies (BOS) and Board of Examiners (BOE) of computer science and engineering department and also the member of IEEE, ACM-W and IEI, life member of CSI and fellow member of ISTE and IETE.

https://orcid.org/0000-0003-3365-6498
E-mail: jayashreeptl@yahoo.com
Department of CSE
P.D.A College of Engineering
Kalaburagi, India

Vijayalaxmi Kadrolli is currently a Research Scholar at the Computer Science and Engineering Department at Poojya Doddappa Appa (PDA) College of Engineering, Kalaburagi, Karnataka, India, an autonomous institute affiliated to Visvesvaraya Technological University (VTU), Belagavi, Karnataka, India. She obtained her M.Tech. in Computer Science and Engineering from the P.D.A. College of Engineering, VTU, in 2004. Her main research areas are in wireless sensor networks, artificial intelligence and soft computing.

https://orcid.org/0000-0003-4349-519X
E-mail: udachanv@gmail.com
Research Scholar
Department of CSE
P.D.A. College of Engineering
Kalaburagi, India

Siddarama R. Patil received his B.E. degree in Electronics and Communication Engineering from Gulbarga University, Gulbarga, Karnataka, India, M.Tech. in Telecommunication Engineering and Ph.D. from the Indian Institute of Technology (IIT), Khargpur, India in 1990, 1999 and 2009, respectively. Currently, he is

a Professor and Dean Academics at Poojya Doddapa Appa College of Engineering, Kalaburgi, Karnataka, India. He has published more than 50 research papers in top-tier journals and conferences, including Springer journal, IEEE conference proceedings and Springer Book chapters. He has guided many BE and M.Tech. Projects and guiding more than 8 Ph.D. students. He is a life member of Indian Society for Technical Education (ISTE), India Member of Institute of Electrical and Electronics Engineers (IEEE), Member of Institute of Electronics and Telecommunication Engineers (IETE), India and Member of Institute of Engineers, India. His current research includes Information Theory and Coding, Turbo Codes, LDPC codes, Iterative decoding algorithms, wireless sensor network, Mobile Ad Hoc Network, Cognitive Radio.

https://orcid.org/0000-0002-7798-1359
E-mail: pdapatil@gmail.com
Department of E&CE
P.D.A. College of Engineering
Kalaburagi, India

Paper

Neuroplasticity and Microglia Functions Applied in Dense Wireless Networks

Łukasz Kułacz and Adrian Kliks

Faculty of Electronics and Telecommunications, Poznan University of Technology, Poznań, Poland

https://doi.org/10.26636/jtit.2019.130618

Abstract—This paper presents developments in the area of brain-inspired wireless communications relied upon in dense wireless networks. Classic approaches to network design are complemented, firstly, by the neuroplasticity feature enabling to add the learning ability to the network. Secondly, the microglia ability enabling to repair a network with damaged neurons is considered. When combined, these two functionalities guarantee a certain level of fault-tolerance and self-repair of the network. This work is inspired primarily by observations of extremely energy efficient functions of the brain, and of the role that microglia cells play in the active immune defense system. The concept is verified by computer simulations, where messages are transferred through a dense wireless network based on the assumption of minimized energy consumption. Simulation encompasses three different network topologies which show the impact that the location of microglia nodes and their quantity exerts on network performance. Based on the results achieved, some algorithm improvements and potential future work directions have been identified.

Keywords—ad-hoc network, brain inspired communication, glial cell, neurons.

1. Introduction

The human body has a great potential of adjusting itself to new, specific situations. There are various mechanisms which enable us to learn, become immune to disease and adapt to distinct settings. Such a straightforward observation could be, however, a source of great inspiration in realization of various network capabilities and features, such as learning capability, fault-tolerance, and self-organization. One may observe that almost all human functions are controlled or somehow affected by the central and peripheral nervous system. A closer look on these systems enables us to identify the neuroplasticity attribute which allows neural connections to adapt and reorganize. On the other hand, there are astrocytes which – among various functions they perform in the human body – join two separate, completely different systems enabling them to work together. Astrocytes are mainly involved in linking circulatory and nervous systems. Lastly, dedicated glial cells exist, with microglia being their peculiar type, having the ability to repair damaged neurons. Observations of all these features and capabilities of the human body (with the brain and the nervous system being their primary focus) lead to new proposals concerning their implementation in the context of dense wireless networks [1], [2].

In such a case, numerous transmission points (nodes), deployed randomly over the area concerned, transmit with relatively low power rating, thus communicating with their closets neighbors. The well-known examples include wireless sensor networks [3] and ad-hoc networks, widely explored over the past decades. In this work, however, we are targeting the problem of achieving high fault-tolerance (as may be observed in the human brain) in dense wireless networks, but based on two assumptions: that the overall amount of energy consumed is minimized to the extent possible, and that the complexity of communication between the nodes is reduced to the minimum required. Thus, one of the key assumptions is that the transmission power is minimized to a certain reasonable level (as discussed later), and that the number of nodes deployed within the network is large enough to model the link between the neighboring nodes as a line of sight with the dominance of additive white Gaussian noise (AWGN), and that the effect of multipath transmissions is neglected. Such an approach is necessary to relax the need for the application of advanced coding schemes and retransmission algorithms. We attempt to mimic the behavior of human brain whose energy efficiency in transmitting one bit of information is much lower than that of contemporary wireless systems, with a relatively high level of fault-tolerance being guaranteed.

In this paper we describe how a few inspirations based on the functionality of the human nervous system have been applied in the scenario considered, i.e. in a dense wireless network system, with the ultimate goal of achieving high reliability with ultra-low energy consumption. The paper is structured in the following way. First, in Section 2 we summarize the key capabilities and the selected features of the human brain and nervous system. In Section 3 we present our approach to potential implementation of these biological features in a dense wireless network. Simulation results are discussed and conclusions are drawn in Sections 4 and 6, respectively. Section 5 presents the authors' plan for the future in this topic.

2. Human Body Inspiration

In our investigations, we targeted highly energy-efficient and fault-tolerant dense wireless networks, where we attempted to follow our overall inspirations based on the human brain and nervous system. In this section, we recap the biological and medical information about the roles played by selected *components* of the human body. We indicate precisely, how these *components* inspire us in the context of the scenario considered.

2.1. Neuroplasticity

The observation that human brain and the entire nervous system optimize energy consumption through the course of the entire life is the leading idea behind the research conducted. The brain of an embryo that is a couple weeks old has a fully connected network of neurons. Later, as a result of synaptic pruning, a small child's brain uses 44–87% of the total energy consumed by the body, whereas the brain of an adult – 25% at the most [4]. The process of maintaining commonly used neural connections and removing the rarely used routes is called *neuroplasticity* and ensures better performance of the human brain and lower energy consumption. In addition, in the case of injuries caused by illness or accidents, the brain is capable of rewiring the connections (after long rehabilitation). This means that it has the ability to bypass the damaged parts of the neural network and create new connections to restore the functions affected, e.g. feeling in the limbs. Fault-tolerance in that case is not instant, but requires much time. Although in a real-life wireless network a repair lead time that is too long is typically not acceptable, neuroplasticity still constitutes an interesting mechanism that is worth considering. It may improve the fault- tolerance ability of the network while keeping the overall energy consumption at the desired level.

From the point of view of wireless communications, neuroplasticity may be treated as the ability to optimize the functioning of the network, and to guarantee fault-tolerant communication in the case of an emergency.

2.2. Neurons

Neurons play an important role in nervous system of mammals. A neuron is made up of the cell body, dendrites and an axon with synapses at its ends. Dendrites receive neurotransmitters by receptors. In consequence, the neuron may generate the so-called action potential (AP) in the axon's hillock. This AP moves from the cell body, through the axon, to the synapses that release other neurotransmitters and may activate another neuron. That is how information is transferred within the nervous system. It is worth noting that this type of communication is unidirectional, meaning that where some impulses are generated, no reception of direct response is possible. To understand how the brain knows that some actions have already been performed (like moving or shaking one's head), one needs to note that all information between the central and the peripheral nervous system pass through the spinal cord. In the spinal cord, there are 31 pairs of spinal nerves and each pair is made of afferent and efferent nerves. Afferent nerves transfer impulses from *sensory* neurons (e.g. receptors placed in the skin) to the brain. An analogy to uplink transmissions in wireless networks may be identified here. On the other hand, efferent nerves transfer signals from the brain to *motor* neurons, e.g. those placed in the muscles, which could correspond to the downlink transmission. Communication relying on afferent and efferent nerves is realized through different paths. This is why the same person may not be able to feel the touch with their hand, but may at the same be able to move their hand. Such symptoms may be the consequence of melanotic cancer, for instance [5].

The functioning of a neural network is continuously improved in the process known as neuroplasticity, which is based on a very simple rule: "neurons that fire together wire together" [6]. In this respect, a very important role is played by the myelin sheath which is formed on axons and ensures the acceleration of passing signals, as well as prevents unintentional leakage of impulses to other neurons. In simple words, myelin sheath protects information [7].

In the context of wireless networks, neurons may be treated as transmission points (nodes) responsible for the reception of and for relaying the message. Various neurons are responsible for different communication directions. Finally, the presence of myelin sheath may be understood as a way of boosting transmission in a specific direction and of protecting the information.

2.3. Microglia

Microglia are cells that play a key role in brain maintenance. They constantly monitor the neighboring (associated) neurons and eliminate the damaged or unnecessary neurons and synapses. Where a dangerous signal is detected, microglia switch into active mode. If the severity of the signal is moderate or low, they clean the debris, support regeneration and secrete substances needed in the process of remyelination [8]. But if the dangerous signal is intensive, microglia produce various types of substances to stop the cells that threaten neurons, and stimulate the production of new cells. Microglia are the primary form of the active immune defense system. It is also important to bear in mind that microglia have the ability to communicate with other microglia, nerves and astrocytes.

In the case of wireless networks, microglia may serve as a source of inspiration for creating specific devices which enabling the network to self-repair and activating in the state of emergency.

3. Bio-inspired Functions Applied in Wireless Networks

Once our inspirations originating from the particular elements of the nervous system have been presented and summarized, we intend to discuss, in the present section, details of the system model and the experimental scenario built based thereon. Selected algorithms have also been proposed and described.

3.1. System Model

It needs to be borne in mind that our ultimate goal is to investigate the solutions for wireless communications and data transfer, guaranteeing a high level of reliability with extremely low power consumption. The term *extreme* shall be considered in such a way that we intentionally want to eliminate all potentially unnecessary sources of power consumption. In particular, we intend to minimize processing in physical and medium access layers by relaxing the need for advanced message coding and decoding, sophisticated link adaptation, retransmissions, etc. Such an approach may be considered in a case where, for example, distances between the neighboring nodes are small enough to guarantee a line-of-sight transmission, and where the link may be effectively modeled as flat with AWGN dominance. In consequence, in this analysis we assume the presence of a dense network of simple (i.e. not complicated) wireless nodes deployed randomly over a certain area, and a set of users, also randomly placed on edges of this area, as presented in Fig. 1. Neurons are represented by distributed antenna systems consisting of four antennas (denoted in the figure by black dots), centrally connected by grey lines, and marked by ID. The neuron itself is described in detail in the following section and shown in Fig. 2. Microglia nodes are shown likewise as neurons, but are marked with grey dots. As black dots represent active antennas, grey color should indicate that microglia node antennas are not currently used for data transmission. Additionally, microglia nodes are marked by M prefix added to their ID. In our experiments, we are testing the behavior of the network in the case of various errors. Therefore, the neurons which will be considered damaged in later deliberations are highlighted by the use of red text. On the other hand, microglia nodes marked with blue text are the ones which will be used to repair the network. Users transmit messages between themselves, and the main role of the dense wireless network within the area considered is to forward data from the source user to the destination user. We assume that a unique ID is assigned to each user.

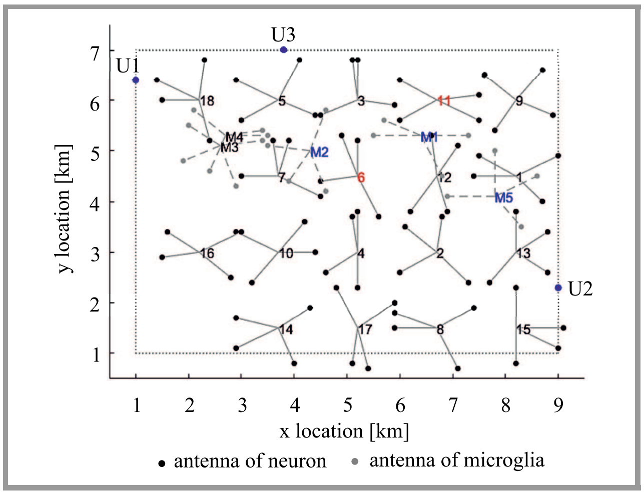


Fig. 1. Topology of the network considered: *Mi* denotes microglia nodes, blue color is used to represent those microglia nodes which will be used to repair the network, and red color is used to represent those neurons which will be considered as damaged. (For color pictures visit https://doi.org/10.26636/jtit.2019.130618)

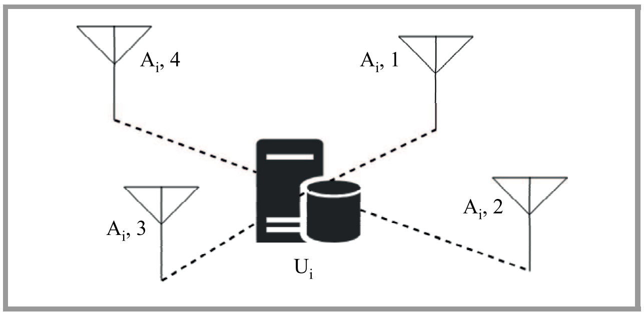


Fig. 2. A wireless neuron.

Following the analogy to the brain and nervous system, the considered network is composed of devices acting as wireless neurons and wireless microglia nodes (marked with the letter M letter before the index), as defined in the following subsections. By assumption, wireless neurons are very simple transceivers with learning (storing) ability - they can remember the approximate location of the user by associating their IDs with the nearest antenna. This information may be used to transmit data in the right direction, directly towards the specific user. The main goal of microglia nodes is to monitor the performance of the network. In the case of any network failure, these nodes may enable classic neuronal functionality (i.e. they can relay messages).

3.2. Wireless Neuron

Inspired by the functioning of the nervous system, we consider a transceiving device (we also refer to it as a wireless neuron) which mimics the behavior of a natural neuron. In particular, let us assume that i-th device n_i is equipped with a low-power distributed antenna system containing N_A antennas denoted as $A_{i,\#}$ and connected to the central processing unit U_i. An exemplary device with $N_A = 4$ antennas is shown in Fig. 2. The transmission power on each antenna is set to -1 dBm, and omnidirectional antennas are considered only. The wireless neuron is fired only when the strength (observed aggregated power) is above a certain threshold. Assuming constant noise power, this constraint may also be reflected by means of the minimum signal-to-noise ratio required, SNR_{min}. Fulfillment of this requirement guarantees also that message dropping functionality, existing in the human brain as well, is applied too. Finally, such a wireless neuron is able to learn and adjust itself (following the neuroplasticity functionality) in order to reduce total energy consumption in the network and to send messages directly towards the destination node, as shown in Algorithm 1. Neurons use the myelin functionality to reduce unnecessary interference they induce within the network by selecting the antenna which is nearest to the destination user of the message. They also reduce energy consumption of the single neuron (the message is transmitted only by a subset of all antennas). In consequence, neurons which are too far away from the best message route will not receive the message, thus the interference level will be reduced.

Algorithm 1: Neuron learning algorithm

```
Data: neuron with $N_A$ antennas
if SNR on any antenna is above the limit $\mathrm{SNR_{min}}$ then
    if neuron did not have message from this source yet then
        select antenna with highest SNR
        save pair of source and antenna index
    end
    if neuron had received messages previously from destination of current message then
        transmit message on saved antenna only
    else
        transmit message on all antennas
    end
end
```

The neuron that receives, for the first time, the message from a specific source relays this message using all antennas. This means that, at the initial phase, the network nodes broadcast all messages and, by doing that, they train themselves. Once trained, the neuron can utilize the distributed antenna system for a more precise message delivery directly towards the destination. Please note that the ultimate goal of the network is ensure that its functionality is realized with minimized energy consumption.

3.3. Wireless Microglia

Wireless microglia nodes, in this case, are the devices very similar to the wireless neurons, for example, they are also equipped with distributed antenna system, but they deliver other functions to the network. In particular, microglia nodes observe surroundings, and if some changes in message flow are detected, like neuron failure (neuron not responding), microglia nodes can enable inbuilt neuron functionality, which was inactive so far in order to reduce energy consumption, and transmit data.

Algorithm 2: Algorithm of enabling neuron functionality in microglia nodes

```
Data: microglia node with $N_A$ antennas
if received a message then
    save source and destination of the message
    observe SNR on any antenna
    if current SNR is drastically different from previous saved SNR values then
        enable neuronal functions
        neurons in range of this microglia node have to start learning from beginning
    end
end
```

In our case, each microglia node calculates the power (and in consequence estimated SNR) of each received message, and if the SNR value observed differs dramatically from previous values (or is even at the noise level), the microglia node switches into active mode. Microglia nodes represent some emergency devices, so if they are enabled, they do not learn the routes of messages and simply transmit the messages using all antennas. The detailed procedure based on which microglia nodes operate is presented in Algorithm 2. In what follows, we denote each of X microglia nodes as $Mi,\ i = 1, 2, \dots, X$.

4. Simulation Results

4.1. Simulation Setup

In order to evaluate the performance of the algorithm proposed in the considered scenario, we considered a dense network with 18 wireless neurons at random positions, with their main role being to transfer messages between users. In this case, three network users have been randomly deployed on the borders of the analyzed area, and they exchange messages between themselves. The transmission power of the users was set to 1 dBm and they have only one antenna. The AWGN channel and free space path loss have been considered only. All simulations were performed using the Matlab environment. The distances between the neurons' processing units and the antennas are approximately 1 km, and the distances between neurons equal at least 1.5 km.

Algorithm 3: Main simulation loop

```
Deploy all neurons, microglia nodes and users
for $Y$ times do
    generate message from random user $tx$ to random user $rx$
    while user rx does not receive message do
        foreach node, which has message and did not send it yet do
            select transmission antenna, based on own saved history
            foreach node $i$ do
                calculate SNR
                if $\mathrm{SNR} > \mathrm{SNR_{min}}$ then
                    if node i is neuron then
                        run Algorithm 1
                    end
                    if node i is microglia node then
                        run Algorithm 2
                    end
                end
            end
        end
    end
end
```

We assume the frequency of 3.5 GHz and the system bandwidth of 5 MHz. SNR_{min} was set at 5 dB.
The main simulation procedure is presented as Algorithm 3. The node represents any type of device: a user, a neuron or a microglia node. Note that the results of depend highly on the topology of the network.

4.2. Routing

Let us now observe the routing mechanism implemented in the network due to the application of two algorithms: for neuron learning (i.e. Algorithm 1) and for activation of the neural functions in the microglia node (Algorithm 2). At the beginning, when neurons do not have any knowledge about users' locations, messages simply flood the network. Later, once the learning phase is finished, we can observe the paths created to deliver messages between users.
Let us now analyze the following example of message routing from user 2 to user 1, as shown in Fig. 3. Green points mark the nodes that have already received the message, and red lines mark antennas of the specific nodes that are used to relay data to subsequent neurons. It needs to be noticed that the simulation shows that identification of a specific path between two users results in a significant reduction in power consumption, equaling approximately 85% (compared to the unlearned network).

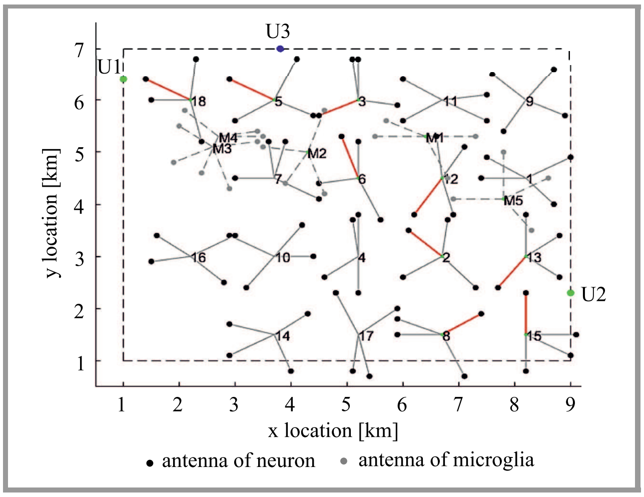


Fig. 3. Route of message from user 2 to user 1 after neuron learning.

After observing the failure of neuron n_6, the microglia node M5 activates its neuronal functionality. It is worth noting that in this example the microglia node M5 is activated and microglia nodes M1 and M2 are not, even though the latter are closer to the faulty neuron n_6. In the algorithm we did not consider direct communication between microglia nodes, so the first microglia node which identifies the problem is turned on. In the considered example, when message from user 2 is sent by neurons n_2 and n_{12}, it is also received by the microglia node M5. Then microglia node M5 waits for the confirmation message (as in saved history) sent by neuron n_6, but does not received anything (due to neuron failure), so it turns on its own neural functionality. Another reason for not activating the nearest microglia nodes stems directly from our algorithm. A microglia node must first receive a message to have the ability of resending it in the case of an emergency. In this specific coincidence, microglia nodes M1 and M2 do not receive any messages after the failure of neuron n_6. The new path for transferring the message transfer from user 2 to user 1 is shown in Fig. 4. It is important to point out that the path between users 1 and 3 did not change. Let us note that neurons n_1, n_9 and n_{11} could potentially transmit using one antenna, but after erasing their memory (due to the fact that the neuronal functionality of the microglia node is enabled) they will not receive any new messages from user 1. This means that this neuron does not know where the destination user of the message is. In order to solve this, periodic update messages from all users may be broadcast. More precisely, at some particular time stamps, each neuron will send a message using all antennas. This will result in an update of the network topology in every neuron.

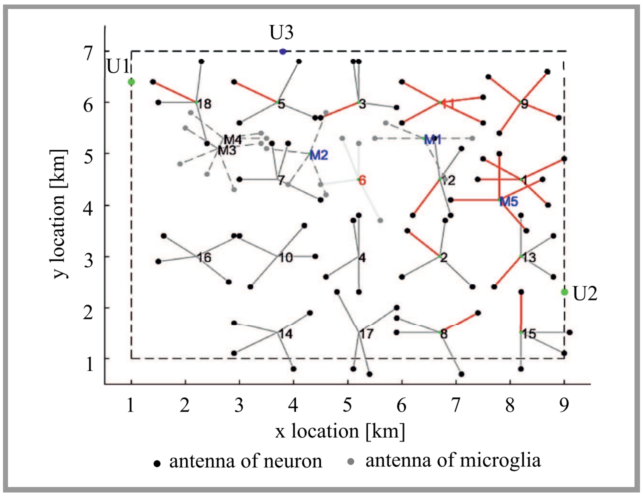


Fig. 4. Route of message from user 2 to user 1 after neuron learning and after neuron n_6 being damaged.

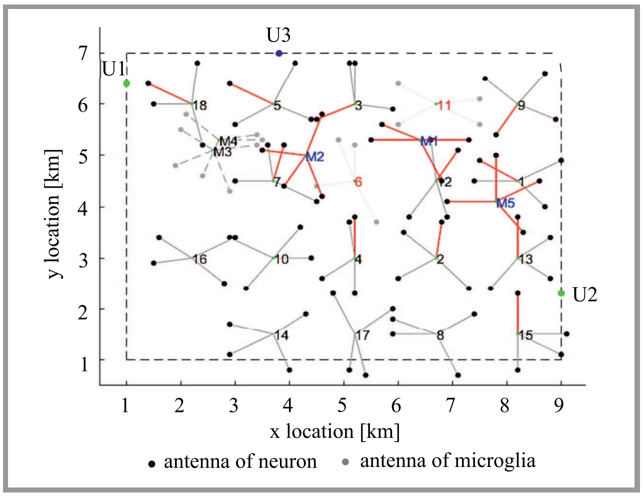


Fig. 5. Route of message from user 2 to user 1 after neuron learning and after neurons n_6 and n_{11} being damaged.

Let us now consider another failure that has happened in this network. After the failure of neuron n_{11} is detected, microglia nodes M1 and M2 activate their neuronal functionality. The new resultant path is shown in Fig. 5.

4.3. Network Fault Tolerance

To evaluate performance of the considered solution, we have analyzed the tolerance of the network to neuron faults. As in the previous subsection, the network topology (as shown in Fig. 1) comprises neurons n_6 and n_{11} (marked red) which stop working properly at one-third and two-thirds of the simulation period, respectively. These time stamps correspond to approx. 12 and 24 messages sent. The blue color of microglia nodes M1, M2 and M5 indicates those microglia nodes which enabled the transmission due to failure detection. In Fig. 6 energy consumption (by the radio portion) and the number of message hops in the network along the path between user 2 and 1 are presented. It may be noticed that energy consumption in the network is very high at the beginning, when neurons do not know where the users are located. The lowest energy consumption value is observed when network learning is completed and every device works properly. When neuron n6 stops working, the microglia node M5 turns on its neuronal functionality and the message still reaches its destination, but with higher energy consumption and with more hops. Without microglia nodes and in the presence of the same failure, messages from user 2 cannot reach their destination. On the other hand, it may be noticed that there is no difference in message flow between users 3 and 1, even when neurons n_6 and n_{11} stop working.

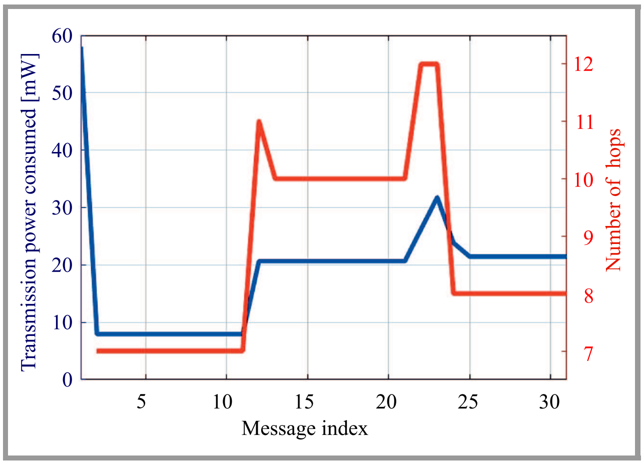


Fig. 6. Power consumption and number of hops on route of message from user 2 to user 1.

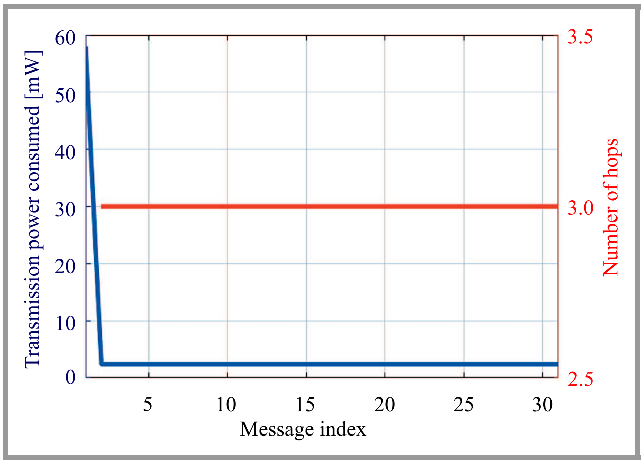


Fig. 7. Power consumption and number of hops on route of message from user 3 to user 1

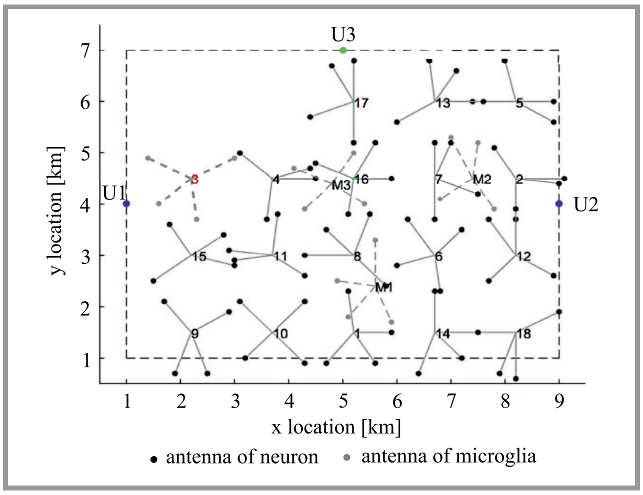


Fig. 8. Network topology (second scenario).

It is important to notice that the proposed algorithms substantially depends on network topology. To evaluate the potential problems and challenges, two other examples are analyzed. In Fig. 8 we can see that once the failure of neuron n_3 has occurred, the distance to the closest microglia node is too high. That is why no microglia node will activate its neuron functionality after the failure of this particular neuron. In consequence, user 1 is unable to communicate with other users. This situation shows that fault tolerance of the network depends highly on the location of microglia nodes. One possible solution to this problem assumes the deployment of microglia nodes closely to each neuron. This will offer a significant increase in fault tolerance, but a high additional hardware cost is required and a resultant increase in energy consumption is observed.

In the third scenario, illustrated in Fig. 9, 18 neurons and 10 microglia nodes have been deployed. Thus, we should achieve a better protection of neurons than in the previous scenario. In that case, once the failure of neuron n_{16} has occurred, we can observe activation of subsequent microglia nodes, resembling the flooding effect. The failure of this neuron has caused 8 microglia nodes activations, because each of the microglia nodes along the route of the message has observed some changes in the transmission, and has turned on their neural functionality. The distant microglia nodes (M6 and M10) are simply too far away from the message route, so they fail to observe the changes concerned. This behavior results in the network connections being repaired, enabling the messages to be once again exchanged

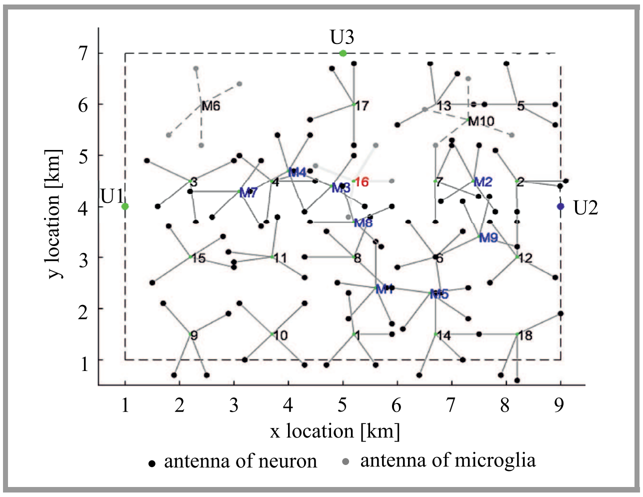


Fig. 9. Network topology (third scenario).

between users. However, in terms of energy efficiency, this is not the best solution. In order to cope with this problem, one option is to suspend the activation of other microglia nodes when the first (closest) one has already been activated (no flooding effect will be observed). This may be realized by the introduction of a dedicated *pause* message which is sent by the *just-activated* microglia nodes, through a dedicated channel, to all nearby microglia nodes, or as a control message. With that change, the failure of neuron n_{16} activates the transmission ability only in microglia node M3, and prevents other microglia nodes from activation. Let us now compare the transmit power consumption (i.e. with and without the *pause* message), and the number of hops in both scenarios, as shown in Fig. 10.

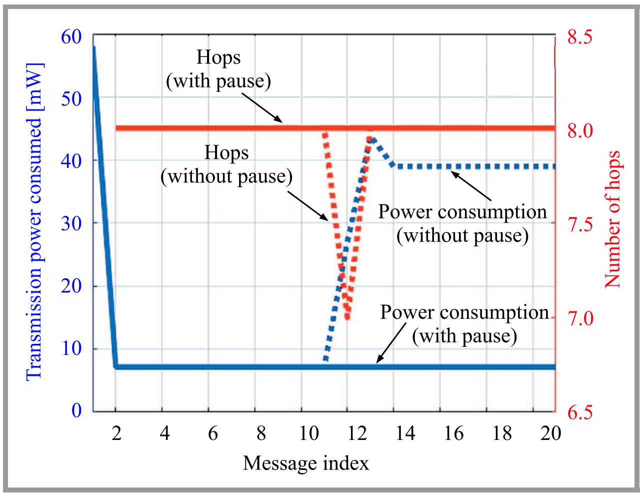


Fig. 10. Power consumption and number of hops on route of message form user 2 to user 1.

It can be seen that the number of hops changes only for a moment (shortly after failure), but in fails to change in the long-term, in both cases. However, energy consumption is much better in the scenario with the *pause* message. This shows the impact of microglia node redundancy in the network.

5. Future Work

5.1. Control Channel

Analysis of the simulation results leads to the conclusion that, conceptually, activation of other microglia nodes should be suspended once the right microglia node had turned on its neural functionality. Each activation of microglia nodes in the aftermath of a neuron failure changes the route along which the message passes between the users. From the point of view of other microglia nodes, such a change may be observed as a neuron failure, which is incorrect. There are two methods that seem to be worth considering in future research. First, a dedicated control channel between the microglia nodes may created, where various control messages (such "suspend activation for a specific period of time") could be transmitted. In such a case, control and data channels occupy various frequency bands. In the second approach, an in-band transmission of control type messages is envisaged, where control messages are mixed with user data on the same physical channel.

5.2. Multiple Activation

Another issue is related to the fact that in our experiment each microglia node activation is associated with dropping the routing memory of nearby neurons, but it can be currently activated only once, and another activation of an already activated microglia node is impossible. By assumption, the microglia node was set as a single use repair device, exactly like a microglia in the human body, where cells of this type are used as the first line defense deployed by the immune system. However, in practice, a dedicated mechanism reverting the microglia node to an idle state or enabling its multiple activations is necessary, and should be the object of future work.

5.3. Switch between Simplified and Advanced Data Transmission

In the considered scenario, very short distances between nodes in the network are considered, and, in consequence, the wireless link may be analyzed as one that is dominated by the additive white Gaussian noise with a dominant direct line of sight. In such a case, one may consider the relaxing of any advanced signal processing technologies (including coding). In specific cases, even a distinct type of an analog transmission could be considered to minimize, to the extent possible, the energy consumed by the node for signal processing and for removing the quantization noise. In this case, we can benefit from lower energy consumption due to a simpler transmitter and receiver structure. Therefore, in our opinion, it would be interesting to evaluate a mechanism for selection of when and where in the network advanced signal processing schemes could be switched off, leaving space for a fully simplified, analog-like transmission.

6. Conclusion

The simulation shows that the human brain and nervous system is a big source of inspiration for current and upcoming communication systems. The functionality observed seems to be useful and may be applied, in certain cases, in wireless networks as well. Wireless neurons in our system are stand-alone devices which do not require a central management unit, which provides scalability and easy reconfiguration for a dense wireless network. Moreover, the functionality of microglia nodes may be applied in order to increase the level of fault-tolerance of the system. The simulations conducted proved the correctness of our approach, showing that the application of additional, human brain-inspired solutions may lead to an increase in network performance. For example, we foresee that addition of the myelin sheath functionality may be a topic of future research. In order to reduce the transmission delay, neurons with myelin sheath could transmit with a higher power to reduce the number of hops within the network.

Acknowledgments

The work has been funded by the National Science Center in Poland, within the framework of the BioNets project, pursuant to decision DEC-2016/23/B/ST7/03937.

References

[1] M. Kamel, W. Hamounda, and A. Youssef, "Ultra-Dense Networks: A survey", *IEEE Commun. Surveys & Tutor.*, vol. 18, no. 4, pp. 2522–2545, 2016 (doi: 10.1109/COMST.2016.2571730).

[2] M. M. Mowla, I. Ahmad, D. Habibi, and V. Phung, "Energy efficient backhauling for 5G small cell networks", *IEEE Trans. on Sustain. Comput.*, 2018 (doi: 10.1109/TSUSC.2018.2838116).

[3] D. Goyal and M. R. Tripathy, "Routing protocols in wireless sensor networks: A survey", in *Proc. 2nd Int. Conf. on Adv. Comput. & Commun. Technol.*, Rohtak, Haryana, India, 2012 (doi: 10.1109/ACCT.2012.98).

[4] G. Gilli, L. Benso, and L. M. Schell (Eds.), *Human Growth from Conception to Maturity*. London: Smith-Gordon, 2002, pp. 36–49 (ISBN: 9781854632166).

[5] D. Purves, G. Augustine, and D. Fitzpatrick, *Neuroscience*, 2nd ed. Sunderland (MA): Sinauer Associates, Neural Circuits, 2001 (ISBN: 978-0-87893-742-0).

[6] S. Löwel and W. Singer, "Selection of intrinsic horizontal connections in the visual cortex by correlated neuronal activity", *Science*, vol. 255, no. 5041, pp. 209–212, 1992 (doi: 10.1126/science.1372754).

[7] A. J. Barkovich, "Concepts of myelin and myelination in neuroradiology", *Amer. J. of Neurorad. AJNR*, vol. 21, no. 6, pp. 1099–1109, 2000.

[8] A. London, M. Cohen, and M. Schwartz, "Microglia and monocyte-derived macrophages: functionally distinct populations that act in concert in CNS plasticity and repair", *Front. in Cell. Neurosci.*, 2013, vol. 7, article 34 (doi: 10.3389/fncel.2013.00034).

Łukasz Kułacz received his M.Sc. degree in Telecommunications from the Poznan University of Technology, Poland, in 2018, where he is currently pursuing the Ph.D. degree with the Chair of Wireless Communications, PUT. His main fields of interest include programming, wireless communications and algorithm design.

https://orcid.org/0000-0002-3434-1917
E-mail: lukasz.kulacz@put.poznan.pl
Faculty of Electronics and Telecommunications
Poznan University of Technology
5, M. Skłodowska-Curie Sq.
60-965 Poznań

Adrian Kliks received his M.Sc. and Ph.D. degrees in Telecommunications from the Poznan University of Technology, in 2005 and 2011, respectively. Since 2011, he has been an Assistant Professor at the Chair of Wireless Communications. His research interests cover a wide spectrum of wireless communications, in particular new waveforms for future wireless systems, including orthogonal, non-orthogonal and non-contiguous multicarrier schemes. He is also interested in the application of cognitive radio technology, advanced spectrum management, deployment and resource management in small-cells, as well as network virtualization.

https://orcid.org/0000-0001-6766-7836
E-mail: adrian.kliks@put.poznan.pl
Faculty of Electronics and Telecommunications
Poznan University of Technology
5, M. Skłodowska-Curie Sq.
60-965 Poznań

Paper

Empirical Approach in Topology Control of Sensor Networks for Urban Environment

Bartosz Musznicki

Chair of Communication and Computer Networks, Poznań University of Technology, Poznan, Poland

https://doi.org/10.26636/jtit.2019.129918

Abstract—Research into the topology control of Wireless Sensor Networks (WSNs) is geared towards modeling and analysis of methods that may be potentially harnessed to optimize the structure of connections. However, in practice, the ideas and concepts provided by researchers have actually been rarely used by network designers, while sensor systems that have already been deployed and are under continued development in urban environments frequently differ from the patterns and research models available. Moreover, easy access to diversified wireless technologies enabling new solutions to be empirically developed and popularized has also been conducive to strengthening this particular trend.

Keywords*—*empirical approach, node deployment, sensor network, topology control, urban environment, WSN.

1. Introduction

The need to use devices that provide measurement data and those that transport thus obtained information to a destination point usually located deep inside the network is an inextricable element defining the operation of any Wireless Sensor Network (WSN) [1]. Hence, with reference to WSN, a sensor is typically understood not only as a component that performs measurements, but also as an entire small-size network node. This small and highly specialized microcomputer should be equipped with measurement sensors, but also with its own power supply, a wireless communication module, a microcontroller or microprocessor and memory [2]. In addition, other components defining the applications area, e.g. a GPS signal receiver or relays that make the control of external actuators possible [3], may be required as well.

Initially, work on sensor networks involved the individual authors' own hardware designs, mainly due to the lack of commercial availability of dedicated products. As recently as 5–10 years ago, sensor platforms belonging to the MICA2, MICAz and TelosB families were considered to be the most advanced and were most widely accepted by researchers. Over the past few years, general-purpose embedded platforms, such as the Arduino UNO, equipped with the ZigBee module, or the Raspberry Pi 3 Model B which provides IEEE 802.11b/g/n and Bluetooth 4.1 wireless connectivity, seem to be used increasingly frequently. Because of their easy availability, affordable pricing and numerous configuration options, they enable researchers, enthusiasts and innovators alike to design and deploy sensor networks or sensor-like networks [4]–[8].

Another factor that supports the development and implementation of WSNs is the increasing coverage ensured by different wireless access networks [9]–[12] that can be used to transfer data. Moreover, Low-Power Wide-Area Network (LPWAN) technologies that support long-range low bit rate and energy efficient communication in sub-1 GHz frequencies are being developed (e.g. by LoRa Alliance, SigFox and Weightless SIG). In urban environments, such network infrastructure may be found, most frequently, on top of high-rise structures (see Fig. 1), i.e. mounted on masts or placed on rooftops of buildings, as well as at more unusual locations such as, for instance, on the branches of an artificial tree on the slope of a hill or in palm trees. Furthermore, components providing access to different local (short-range) wireless networks are common inside office and apartment buildings as well.

What becomes more and more apparent is the fact that objects which at first glance are a far cry from simple sensor nodes, here understood as devices with limited computational capabilities and battery-based power supply, are now being equipped with integrated sensor capabilities. In real-life applications, sensor functions are performed, ever more frequently, by vehicles and consumer devices, such as residential water meters, mobile phones and sports watches. They are often called smart objects and are considered to be capable of providing additional functionalities. Objects such as these may be combined to form an integrated system and are capable of cooperating to complete more complex and context-related tasks [13], [14]. They are often combined with additional analytical tools and distributed resources provided by cloud computing [15]. The ever-increasing potential in terms of the range of applications for sensor-based or sensor-like devices, which are already perceived as one of the components of the Internet of Things (IoT), is followed, within the domains of research and product marketing alike, by the need for their further differentiation. Subsequent subcategories with their particular functionalities and purposes clearly identified emerge [16], [17], including Vehicular Sensor Net-

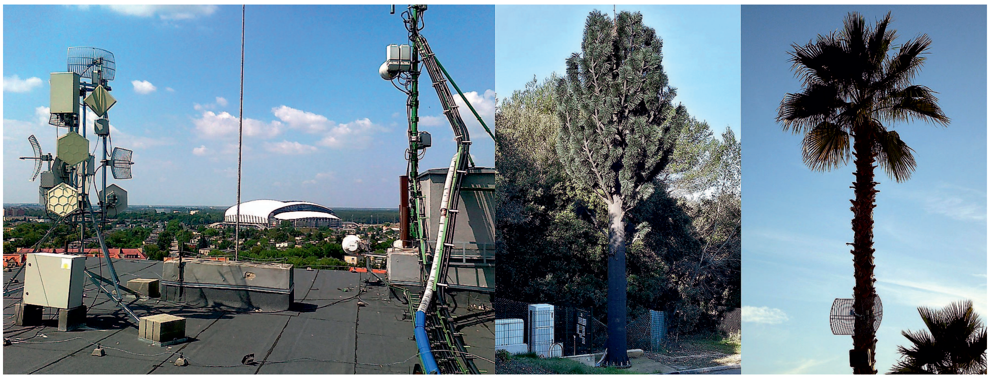


Fig. 1. A rooftop cellular and IEEE 802.11 networks base station (left) (Poznań, Poland, July 2011), an artificial tree used as a base station on a hillside (middle) (Sophia Antipolis, France, March 2016), a wireless station on a palm tree (right) (Athens, Greece, October 2013).

works (VSNs), Body Area Networks (BANs), home automation, smart factory or smart city, just to name a few. All of them, however, have one thing in common – they perform measurements and rely on wireless communication solutions [18].

In recent developments, attention is attached not only to functionality-related issues, but also to considerations related to the nature of the processes involved and to system of connections. The structure of connections and the way they are used depend on the environment and the conditions in which the network is operating, as well as on the assumptions made and the tasks that are to be performed. Such circumstances are often investigated in simulations of the required type [19]. It may be stated, based on the author's research, that the implementation phase is of a dual nature. It is sometimes preceded by a long-term research study (lasting for years), but more often simply follows the product development stage. This leads to a certain dissonance between research and actual implementation practices. On the one hand, the results of novel research related to optimization methods can be relied upon [20]. On the other, however, in response to changeable business needs, well-established technologies and solutions are continuously used, while the very vision of the sensor network is either being simplified or modified so that it meets the requirement of quick execution of ideas and commercial adaptation of the product to the needs of the market.

The above observation partly coincides with the opinion, as presented in [21], that "*although new topology control algorithms are presented on a regular basis, topology control has never made the breakthrough in real-world deployments*", and may be accompanied by a statement that the obstacles faced include the following: unrealistic assumptions, unsuitable graph structures, application agnosticism, unclear role in the stack and insufficient framework support. The arguments presented included insufficient utilization of graph-based methods for optimization of the structures of typical WSNs. In the author's opinion, with real-life implementations of sensor-type networks, it is rather the market situation and the way in which innovative products and services are created that is largely decisive for the approach adopted with regard to topology issues. As a consequence of this attitude, this paper provides a juxtaposition between theoretical and research-based views on the manner in which the structures of WSNs are managed on the one hand, and the empirical approach to deployment and implementation of sensor-based systems, as seen by a network architect with practical experience, on the other. The presentation is based on the author's experience with various networks and his involvement in tests pertaining to the networks under scrutiny.

Section 2 provides a definition of the notion of topology control, whereas Section 3 presents two types of physical arrangements of nodes that may be found in practice. Both categories are illustrated with examples of real systems. Conclusions are presented in Section 4.

2. General Objectives of Network Topology Control

Network topology is typically understood to be a model that describes the structure of connections between the elements within a given network [22] and is frequently presented in the form of a graph $\mathscr{G} = (\mathscr{V}, \mathscr{E})$, where $\mathscr{V}$ denotes the set of vertices (in other words nodes, such as sensors), whereas $\mathscr{E}$ is the set of edges (i.e. connections) between vertices [23]. This notation may refer to both physical relations (relative arrangement of nodes and connections between them, directly stemming from the properties of the transmission medium applied) and logical relations (an operational configuration based method for transmitting data via the network from the starting point to the end point, between the elements of the network's infrastructure [24]. As a result, physical and logical topologies may be distinguished [25].

Topology control is a related notion and in general encompasses different aspects related to planning, maintenance and adaptation of the system of connections within a given network [26]. Topology management is an alternative term used on some occasions [27], [28].

In its broadest and the most diverse scope, topology control is closely connected with wireless networks, not only due to the variable character of radio communications, but also due to the particular features of devices that rely on wireless transmission. Topology control is distinctively illustrated with regard to the ad-hoc networks [29]–[31], and becomes of special significance within the WSN context. As the development of sensor networks progresses, this issue is gaining in importance and reflects the increasingly more detail-oriented and extensive scrutiny of each of the aspects influencing the network structure.

Santi states that "*topology control is the art of coordinating nodes' decisions regarding their transmitting ranges, in order to generate a network with the desired properties, e.g. connectivity, while reducing node energy consumption and/or increasing network capacity*" [26]. Labrador and Wightman point out, more broadly, that "*topology control is the reorganization and management of node parameters and modes of operation from time to time to modify the topology of the network with the goal of extending its lifetime while preserving important characteristics, such as network and sensing connectivity and coverage*" [32]. At the same time, they emphasize that the above definition refers not only to the control of the transmitting power of sensors, but also to turning on and shutting off nodes depending on current needs. Aziz *et al.* provide the following definition "*topology control is a technique that uses any controlled network parameter to generate and maintain a topology for the benefits of reducing energy consumption and achieving a desired property for an entire network*" [33]. Li *et al.* present, in turn, a view that topology control is a fundamental benchmark "*which characterizes how well a sensing field is monitored and how well each pair of sensors is mutually connected in WSNs*" [34]. Consequently, studies related to topology control may include investigations into operational management and transmitting power control of radio modules [35], [36], energy-harvesting [37], interference prediction [38], as well as sensor placement [39], network coverage [40], logical network structure and message routing [41], [42], node functional diversification and hierarchy (e.g. chaining [43], and clustering [44]).

Having closely examined the development of this particular domain and taking advantage of the definitions presented above, it can be generalized that the term *topology control* covers all activities intended to influence the physical or logical structure of a network in order to optimize the way the network executes its tasks while retaining the expected properties.

3. Sensor Nodes Deployment

Node deployment is the basic element that influences the way the topology of a sensor network is controlled [1], [45]. Moreover, the empirical practices related to topology control in commercially-oriented ventures, as discussed in the following sections, seem to focus, first and foremost, on the distribution of nodes.

In real applications, depending on particular requirements or environmental conditions in which a given sensor network operates, random [40] or deterministic [46] distribution of network nodes may be distinguished. This distribution may be predicted at the designing stage, or can be partly or totally random in a dynamically changing working environment. In many real applications it is difficult or even impossible to assign a given type of node distribution within a network (or a part thereof) to just one of two categories. One should not forget that each sensor structure is characterized by a certain degree of determinism (therefore also a degree of randomness) that will vary along with changes in the external environment, though frequently to an extent that can be neglected in a given application.

The following subsections discuss both types of node deployment schemes that may be encountered in urban environments, i.e. random deployment and deterministic deployment, and provide examples of their implementations.

3.1. Random Deployment

Since the very beginning of work on WSNs, a general view has prevailed in the literature of the subject that random deployment of nodes [40] is the fundamental approach, serves as the point of departure while constructing WSNs and is typical of this group. One of the basic areas of application for such a network is monitoring the parameters of the natural environment (e.g. temperature or pressure) [47]. Much attention has been then given to methods for random deployment of sensors, while one of the most frequently mentioned examples illustrating the above would be a situation in which sensor devices are dropped from an aircraft over the area to be monitored [48].

The sheer multitude of potential applications of sensor networks that has been identified over nearly two past decades has led to numerous complex concepts related to the construction of the networks' physical topologies. Many of them depart from randomness of node deployment, understood in the direct and unconditional manner, by introducing some kind of order. In the author's opinion, one of the most interesting scenarios is a network in which the sensors are deployed in an unknown working environment, with the process carried out according to a predefined algorithm and based on information obtained during actual deployment [49].

Other networks that should also be noted within this context include VSNs, in which a sensor network typically covers the intended area – a road and its closest surroundings – and may be spread over tens or even hundreds of kilometers. In the case of such networks, one may speak of a combination of randomness and determinism, i.e. a certain portion of the sensors are deployed permanently alongside the road, while the sensors that communicate with them are those that may be deployed in vehicles. Their distribution is random in such a case and they frequently remain beyond the control of sensor networks' operators [50]. Vehicles may be then viewed as mobile agents that perform not only tasks assigned to them, but also additional functions that relate to stationary sensors [51]. Currently, it is rare that nodes in such systems communicate between one another, because typically the exchange of information is performed with a central point only (e.g. a control center or a data collection software). Furthermore, in the case of VSN, it is common to omit issues related to limitations of energy sources or computational power – so important in traditional WSNs [52]. This is why the largest number of implemented examples of sensor-type and sensor-like networks may be found in the group of systems related to transport and presented in the subsequent subsections. It is worthwhile noticing, at this point, that the largest systems of this type, based on smartphones that serve as mobile agents, form today's most widely used sensor networks and that their number and scope of functionalities are depicted by a continuous upward trend.

3.1.1. Mobile Measurement Agent within INEA Network

Following the research studies initiated by the author and carried out together with the associates from INEA, a regional Polish telecommunications operator, it was possible to perform measurements related to the operation of radio networks that are based on the IEEE 802.11 family of communication standards [53]. A detailed description of the tests and an analysis of the results obtained are presented in [11], while the conclusions from this work are presented below.

The first group of experiments was performed using access points providing wireless Internet access, with the use of the 2.4 GHz band, to the passengers of 330 public transport vehicles in Poznań and Konin, two cities in the Greater Poland region. The movement of those agents was beyond any control of the telecommunications network operator because it was the transport operator that decided about the movement of the vehicles involved, at the same time impacting the topology of the network. By relying on devices known as RouterBoard RB751U, deployed in trams and buses and equipped with a 2.5 dBi antenna and a 4G cellular network modem-based uplink (providing Internet connection), a collection of samples was performed in each vehicle once every 15 minute. The important parameters included noise floor level (background noise), expressed in dBm, and the values of the Received Signal Strength Indicator (RSSI) related to each of the connected users were recorded. In this way, reliable around-the-clock distributions for the urban environment were obtained, thus allowing a technical and a business analysis useful for INEA to be performed.

The around-the-clock distribution of the average noise floor that occurred over the period of one month was particularly interesting. In order to verify the distributions observed, they were compared with the results obtained with the help of 10 stationary INEA access points operating in the 5 GHz band, located on masts and on rooftops. It turned out that both trends were nearly identical, which is clearly visible in Fig. 2. Each point in the graph corresponds to the average value from a given month, whereas confidence intervals correspond to standard deviation. Because the comparison involved a mobile network and a stationary network, both operating in different environmental conditions and with the application of antennas with a much higher gain and on other frequencies, the conclusion drawn is the changes in noise floor were mainly caused by external factors and were not typical of human activity. Further studies were conducted in 2017 with the use of a stationary 3.5 GHz IEEE 802.16e WiMAX network, yielding similar characteristics and suggesting that they may result, in addition to other factors that are yet to be identified, from ambient temperature [54] changes.

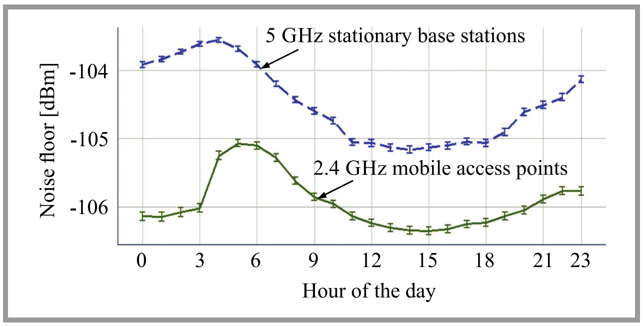


Fig. 2. Average noise floor observed with the use of INEA's IEEE 802.11 access points.

The other group of experiments involved issues related to the operation of more than 20 thousand fixed residential Wi-Fi hotspots, i.e. IEEE 802.11 access points that were located at INEA customers' homes. Each INEA subscriber takes advantage of a community Wi-Fi service based on home routers connected to the INEA_HotSpot_WiFi network.

The experience of mobile users was verified using a smartphone equipped with an IEEE 802.11a/b/g/n/ac (2x2 MIMO) radio module, GPS and GLONASS navigation receiver, as well as measurement and communication freeware. The test to be carried out involved measurements of signal parameters and extraction of technical information for each Basic Service Set Identifier (BSSID) observed. During the test, the user was moving, with the smartphone, on the sidewalk, between multi-story apartment buildings in council housing estates built in the 1970s. It was observed that 313 out of 1874 BSSIDs used the Service Set Identifier (SSID) with the name INEA HotSpot WiFi. The tests performed indicate that 59% of INEA residential hotspots could have been used for conversations with the use of Voice over IP (VoIP), provided that the user standing on a sidewalk was connected to the access point and that the strength of the signal received was not lower than −75 dBm.

3.1.2. Yanosik Driving Assistant and notiOne Location Beacon

The development of mobile radio-location systems, as well as increasingly common packet data transmission has led

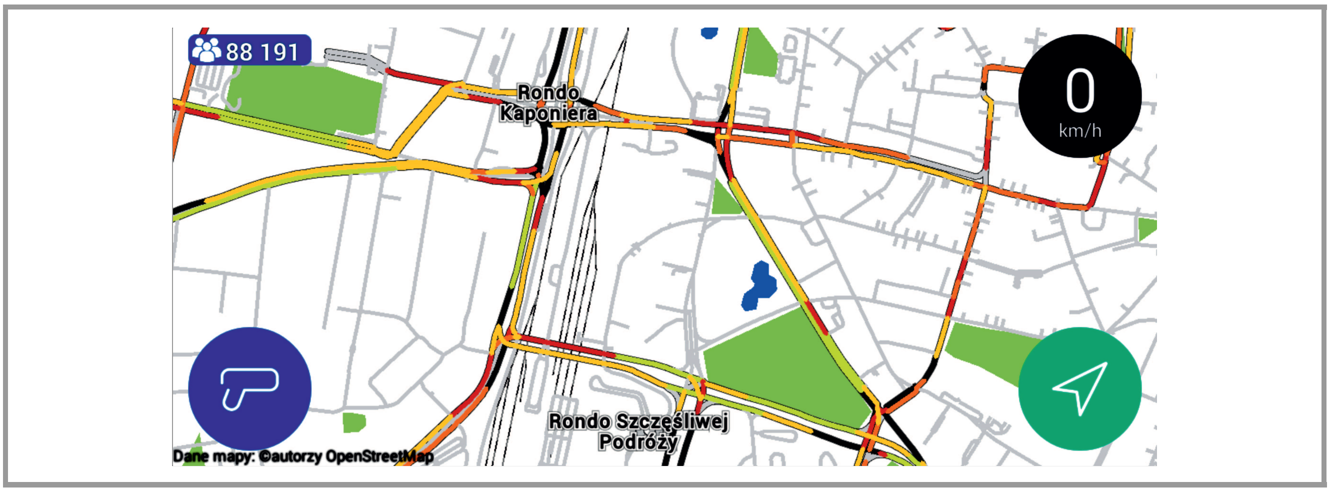


Fig. 3. Poznań city center road traffic map presented by Yanosik (ver. 3.1.1.1), 9 May 2017.

to widespread use, among drivers, of sensor systems based on mobile agents. No statistical comparisons are available that would illustrate the popularity of particular solutions, but, based on market observations, one may come to a conclusion that the Yanosik driving assistant, operated by a Poznań-based company named Neptis, is the leading platform among Polish drivers. On the global scale, such applications as Google Maps, HERE WeGo and Waze are based on similar concepts.

The Yanosik driving assistant is primarily a system used for exchanging information and warnings between drivers, relying on dedicated devices and smartphones (a special free app has to be installed). Both devices need to be wirelessly connected to the Internet (usually via a cellular network) to act as mobile agents. Vehicles (devices) do not communicate directly with other road users because the entirety of the exchange of information is performed on the central operator's host platform. This network has a dynamically variable random topology of the logical star type, in which the location of nodes depends on the situation on the road.

Drivers who report events, such as road accidents, road works or a police patrol, are the source of information. Along with a report, additional information is forwarded on the user's location. A smartphone with the Yanosik app may be also used as a navigation system with real-time traffic service. As indicated in a release issued to the author by Neptis in April 2017, at its peak times, the system is used by over 150,000 concurrent users, which means that it comprises the same number of sensor nodes. The further processing of data makes it possible to develop real time traffic intensity maps (Fig. 3). In addition, the author was shown the results of an investigation that demonstrated the use of information gathered from sensors embedded in smartphones (accelerometers and gyroscopes) in order to evaluate the quality of roads and driving comfort. This enabled data on vibrations experienced by drivers and vehicles, while in motion, to be analyzed. The smoothness of traffic flow and the average speed were studied as well.

In more complex network topologies, smartphones offer wireless interfaces of different types and make it possible to connect with devices of other types to execute additional agent functions. This provides a basis for the operation of devices known as notiOne, which are the so-called beacons (see Fig. 4). These simple and small-scale mobile transmitters broadcast signal that includes a device identifier and may be received by smartphones that happen to be in the vicinity. In this way, the location of the beacon is approximated based on the accurate location of the smartphone that received the beacon's signal. Effectiveness and accuracy of such geolocation depends on the number of nearby smartphones on which software co-operating with the system has been installed.

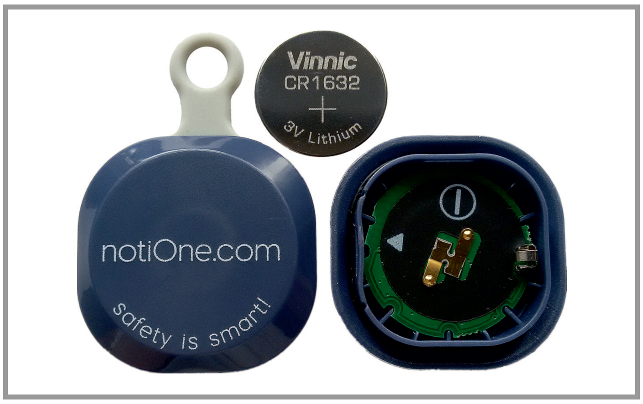


Fig. 4. Opened notiOne beacon with a CR1632 button cell battery.

The beacon may be attached to a key pendant or a dog collar, so that if lost, it makes it possible to easily obtain information on its location by means of a dedicated app. The location highlighted on the map will indicate the spot where the beacon's signal was received for the last time. The notiOne device is powered by a button cell battery that allows the device to work for nearly one year, with the range up to 90 m using Bluetooth 4.0 Low Energy connectivity.

3.1.3. The Veniam System

The activity of Veniam serves as a good example of a successful implementation of research work on mobile sensor networks in a commercial product. The company's founders, Barros, Cardote, and Sargento, were previously involved in research on mobile and sensor networks, among others [55]–[58]. In 2003, as a result of their work, commissioned by city authorities, Veniam launched in Porto, Portugal, a mobile wireless network targeting to collect the results of measurements performed in the urban environment and improving the operational efficacy of the city's transport and service utility vehicles. Recently, the network comprised nearly 350 vehicles (buses, police cars, garbage trucks, taxi cabs) [59]. These vehicles are equipped with a device that serves as an access point (with a network interface, etc.) known as NetRider. To provide connectivity, radio base stations are used, operating in the bandwidth of 5.9 GHz according to the IEEE 802.11p standard for the mobile environment, in particular for the so-called Intelligent Transport Systems (ITS) [60]. Each access point serves as an IEEE 802.11g Wi-Fi hotspot and makes the Internet available to passengers. Sensors are placed inside vehicles to make environmental measurements and to monitor the fleet. The vehicles may communicate with one another and serve as mobile transceivers (called mobile relays) [28], making indirect, real time communication between the control system and the out of range vehicles possible. In turn, when communication in real time is not possible or suffers from delays, thus forming, in fact, a delay tolerant network (DTN), a function of data extraction and temporary buffering, most frequently called data MULE (Mobile Ubiquitous LAN Extension) [61], is performed. If a vehicle that passed near a stationary sensor located at the edge of the road is just outside the base station coverage, the reading is taken (receiving a portion of data using IEEE 802.11 or Bluetooth) by using the local memory of the mobile agent, with data forwarded to the central repository once the connection with the base station has been reestablished [62]. The system employs a complex and variable topology of connections that is successfully created by means of different models and methods for wireless communication.

The author has had a chance to examine the Veniam network's control and management panel. The virtual environment software makes it possible to monitor the network in real time and to store and analyze the data extracted. The location of any vehicle is presented, just as are the estimated range of the hotspot and the live traffic intensity map for the area covered.

3.1.4. Automated Road Passenger Transport

Transport automation is another area of sensor applications that is currently under development. One of the leading European projects in this area was CityMobil2, launched in 2012 and concluded in August 2016, co-financed under the European Union's Seventh Framework Programme [63]. The goal was to create a pilot automated passenger road transport platform with automated and autonomous vehicles, and to carry out tests in a number of European urban environments [64]. The platform was made up of electric vehicles, i.e. mini buses equipped with wireless interfaces for communication with the control center, as well as with sensors necessary for the unmanned vehicles to operate. The entire system used a central controller that gathered data transmitted by the vehicles and controlled the vehicle movement, hence the physical network topology. In addition, the personnel was capable of override the control system.

In 2016, the author had a chance to visit a demonstration route used by autonomous vehicles in the French Sophia-Antipolis technology park, and took a ride in the EasyMile EZ10 autonomous vehicle shown in Fig. 5. The one-kilometer test lane with five stops was used by three autonomous shuttle service vehicles, each with the capacity of 9 passengers. By using a GPS receiver, proximity detectors and accelerometers, the vehicles were capable of adjusting the driving speed to other road users and avoided obstacles or, alternatively, stopped before them if avoiding the obstacle was impossible. When this was the case,

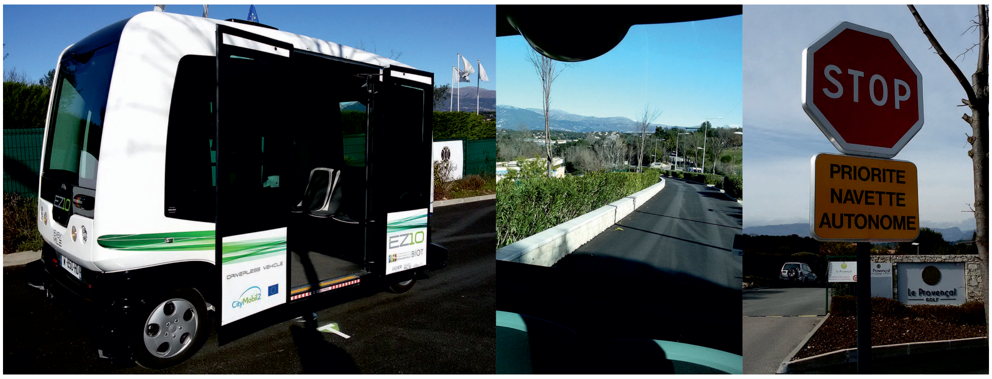


Fig. 5. EasyMile EZ10 (left), dedicated bus lane (middle), and autonomous vehicle precedence sign (right) (Sophia Antipolis, France, 2016).

the vehicles were sending information to the control center, requesting operator's intervention.

The primary reason for the operator's presence in the autonomous vehicle was, according to the operator, the restriction imposed by applicable French legal regulations that do not allow vehicles to be admitted to streets and roads without a person authorized to drive them on board. The other reason was the occurrence of a potentially dangerous situation, due to the pilot stage of the project. In fact, this turned out to be necessary when, for example, the board computer crashed, when a situation occurred on the road that had not been foreseen in the control algorithms applied, or when an uncontrollable panic attack took place among the passengers.

During the tests, a collision of a car driven by a human and one of the unmanned vehicles occurred at the only crossing of the dedicated bus lane with a general traffic road, and as a result the presence of the operator also turned out to be necessary. Following this accident, to make the test and demonstration route more conspicuous to other road users, it was additionally marked with noticeable posters, and a STOP traffic sign with a visible note "*priorité navette autonome*" (French for "*priority for the autonomous shuttle bus*"), as shown in Fig. 5, was installed.

The tests have shown how diversified problems need to be foreseen and predicted while designing autonomous systems operating in a dynamically changing environment. Moreover, more integrated and complex information is to be extracted by means of different sensors, enabling a fast and reliable interpretation of the road traffic situation.

3.2. Deterministic Deployment

Deterministic deployment of nodes in WSNs is the second category identified in the process of creating physical topologies. For example, some of industrial WSNs are capable of using fixed or controllable node deployment (distribution) schemes. This type of a network would be tasked, for example, with monitoring the vibration signatures to predict maintenance needs [65]. Manually deployed sensor networks with cameras and microphones [66] are also being considered for implementation. In addition, in various research projects related to natural environment monitoring, sensors are deployed manually. This makes it possible to adjust network topology to the nature of phenomena observed and to the assumptions based on which the experiments are to be carried out. As often as not the location of nodes remains deterministic (in most cases it changes slowly or remains fixed) during their service life (e.g. this is the case with investigations concerning volcanic phenomena [45]). The deployment of nodes or the range of their relocation makes it possible to prolong the operating time of a network. This can be also achieved, for example, by securing such distances between sensors that would enable the routing mechanisms applied to remain operable in the most effective way, without the need to manipulate the power of transceivers [1], [39].

In consumer applications, home automation systems (otherwise known as smart home systems) become increasingly popular. Usually, they have the features of a small-scale sensor network and take advantage of wireless communication protocols, such as Bluetooth Low Energy, ZigBee, Z-Wave and 6LoWPAN [67]. Their physical topology is usually determined at the installation stage due to the operational range being limited to just one property, whereas network communication takes place predominantly directly between the node and the base station (control center), and only occasionally with the use of intermediary network nodes [68]. In multi-family houses, radio-enabled electricity meters [69] or water meters [70] may also be found, often placed inconspicuously but effective in performing their measuring functions.

3.2.1. Sensor Network in Cisco openBerlin Innovation Center

The sensor network launched in Berlin, Germany, at the Cisco openBerlin Innovation Center, is an example of the deterministic deployment scheme. The network is used both as a backbone of a smart home system and as a testbed on which research projects of companies affiliated with openBerlin are evaluated [71] (Fig. 6).

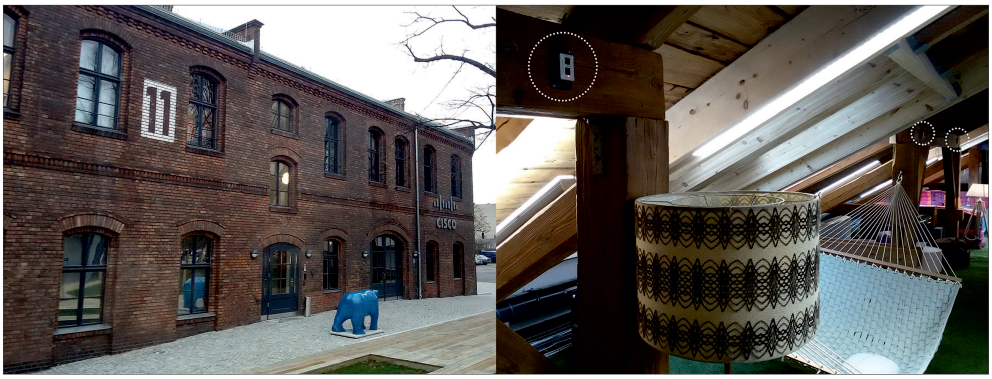


Fig. 6. Cisco openBerlin Innovation Center (left) and Bosch XDK110 based sensor network (right) (Berlin, Germany, 2016).

The author had a chance to examine the components of the test and demonstration set, as well as to verify the topology of the network deployed within the building that consisted of several hundred nodes. The node hardware is based on the Bosch XDK platform. The devices are located at different places, including cable support ceiling systems in the office section or near ceiling joists in the recreational section, as shown in Fig. 6, and are marked with dotted circles. The Bosch XDK110 node is a hardware component equipped with a 32-bit ARM Cortex-M3 microcontroller, 1 MB Flash memory, 128 KB RAM, a Micro SD card reader, Bluetooth 4.0 Low Energy and IEEE 802.11b/g/n modules, 560 mAh rechargeable battery and contains 8 sensors: an accelerometer, a gyroscope, a magnetometer, as well as humidity, pressure, temperature, acoustic and light sensors. The network is used as a source of data for the system developed by an IoT company known as Relayr, and allows temperature and lighting inside the building to be controlled.

3.2.2. Fibaro Home Automation System

The Fibaro home automation system comprises a host controller that wirelessly manages the attached sensors and actuators (Fig. 7). The sensors include smoke, flood and motion detectors, door or window opening sensors, as well as a universal device that allows any sensor with a binary output to be added to the system. In addition, such components as a switch-key in the electric wall socket enclosure with an energy consumption measurement functionality, roller and gate shutters, lighting controller or relay switches may also be used. The system is capable of co-operating with a home weather station, wireless speakers or cameras.

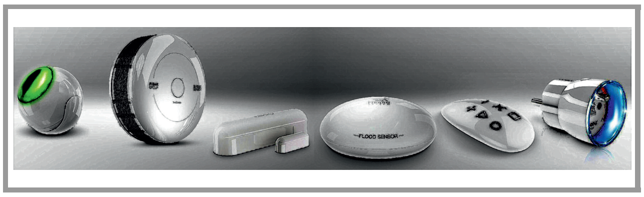


Fig. 7. Components of FIBARO home automation system.

It should be stressed that components of the Fibaro systems communicate via the Z-Wave protocol, and not via IEEE 802.15.4 ZigBee [72] that is in common use in research environments. The Z-Wave protocol was initially developed by the ZenSys, and was then largely used by the standardization organization ITU-T for the development of the G.9959 recommendation [73]. Radio communication relying on the Z-Wave protocol uses, in Europe, primarily the 868.42 MHz frequency band, but systems for the 2.4 GHz band are available as well. Data is routed between network nodes by assigning identifiers, whereas the throughput is not more than 40 kbps [74]. For a commercially available device to be capable of making use of the protocol, it has to be equipped by the manufacturer with a Z-wave communication module, sold separately, and then certified for interoperability to comply with the license agreement.

4. Conclusions

Sensor networks continue to raise the interest of researchers who pursue further improvements to complex problems and propose potential innovative applications. Network technologies and hardware platforms are being developed in parallel, while more and more types of devices are equipped with sensor components. By striving to optimize the operation of WSNs, it is possible to make use of or draw inspiration from new ingenious concepts and products. However, despite the availability of a number of new solutions, in today's real-life sensor network deployments that enjoy an established reputation among their users and are commercially successful in urban environments, one may primarily list only those that treat the idea of WSNs in a rather simplified manner, or those that have been even developed in isolation from elaborate and complex scientific research. Hence, sensor-based systems and products are often more loosely termed as IoT, smart home, etc. This is particularly visible in the area of topology control that can be analyzed and executed as a complex and multifaceted research problem, while it is still hard to find good examples of its implementations in which it would constitute one of the main issues. It seems that, in practice, the prevailing model of implementation is empirical, i.e. draws from experiments, past experience, best practices and intuition. In some implementations, network topology is even unknowingly or deliberately pushed aside from the areas of interest or remains beyond any significant interest of network designers involved in a project – as a component of industry standards or licensed protocols implemented in building blocks relied upon.

The primary or exclusive aspect of topology control is then reduced in its essence merely to the deployment of nodes. In applications related to transport, randomness and variability of node deployment, and hence the physical topology of the network, still remain the dominant element. In applications encompassing building automation systems, in turn, it is the deterministic and static distribution of nodes established during the installation of each of the components that remains dominant.

Although sensor products and services available today still remain at different, frequently early stages of development, they allow given assumptions to be verified and lines of action to be corrected, so that they would meet the expected needs in the best possible way. This might be one of the factors facilitating the introduction of sensor or sensor-like networks into common use. This situation brilliantly illustrates the often overlooked significance of an appropriate and fruitful combination of research activities and market operability. In the area of sensor networks, a number of research investigations outpace, by decades, the current market needs or the implementation capabilities available. As a result, despite their research excellence, they might never be adopted in practice in their full and extensive forms. Innovative business enterprises may not be able to undertake a risk or accept costs of an implementation of too complex and seemingly expendable ideas. No wonder then that they tend to choose ready-made and easily available components and proven solutions that make it possible for them to focus on functionality issues rather than on details of all-technical aspects. This, however, is often done at the cost of getting attached to license-fee paying technologies that are not fully open and, more importantly, are developed by some other companies. As a result, business enterprises chose to provide Minimum Viable Product (MVP) as quickly as possible, i.e. such a product that would in the most favorable way satisfy the expectations of the first group of users and would allow the product to be further developed [75]. It is important then that scientists in their research efforts are able to follow the market developments and try to understand current and future needs of prospective users and, wherever possible, check and streamline their new ideas in close cooperation with operators of already existing networks and systems [76]. Moreover, socially-oriented and valuable results may be achieved when research investigations provide opportunities to transform them into real systems. Then, by gaining practical relevance, they will have a chance to enter the mainstream and be appreciated by standardization organizations, thus, at least to a certain degree, be in a position to shape the way sensor networks are implemented in the future.

References

[1] B. Musznicki and P. Zwierzykowski, "Performance evaluation of flooding algorithms for wireless sensor networks based on EffiSen: The custom-made simulator", in *Simulation Technologies in Networking and Communications: Selecting the Best Tool for the Test*, A.-S. K. Pathan, M. M. Monowar, and S. Khan, Eds. Boca Raton, FL, USA: CRC Press, 2015, pp. 433–458 (doi: 10.1201/b17650-21).

[2] H. Karl and A. Willig, *Protocols and Architectures for Wireless Sensor Networks*. Chichester: Wiley, 2005 (ISBN: 978-0-470-09510-2).

[3] I. Akyildiz and M. C. Vuran, *Wireless Sensor Networks*. Chichester: Wiley, 2010 (ISBN: 9780470036013).

[4] R. Faludi, *Building Wireless Sensor Networks: with ZigBee, XBee, Arduino, and Processing*. Sebastopol, CA, USA: O'Reilly Media, 2010 (ISBN: 978-0596807733).

[5] S. Ferdoush and X. Li, "Wireless sensor network system design using Raspberry Pi and Arduino for environmental monitoring applications", *Procedia Comp. Sci.*, vol. 34, pp. 103–110, 2014 (doi: 10.1016/j.procs.2014.07.059).

[6] F. Leccese, M. Cagnetti, and D. Trinca, "A smart city application: a fully controlled street lighting isle based on Raspberry Pi card, a ZigBee sensor network and WiMAX", *Sensors*, vol. 14, no. 12, pp. 24408–24424, 2014 (doi: 10.3390/s141224408).

[7] C. P. Kruger, A. M. Abu-Mahfouz, and G. P. Hancke, "Rapid prototyping of a wireless sensor network gateway for the internet of things using off-the-shelf components", in *Proc. IEEE Int. Conf. on Industr. Technol. ICIT 2015*, Seville, Spain, 2015, pp. 1926–1931 (doi: 10.1109/ICIT.2015.7215378).

[8] A. D. Deshmukh and U. B. Shinde, "A low cost environment monitoring system using Raspberry Pi and Arduino with Zigbee", in *Proc. Int. Conf. on Invent. Comput. Technol. ICICT 2016*, Coimbatore, India, 2016, vol. 3, pp. 1–6 (doi: 10.1109/INVENTIVE.2016.7830096).

[9] A. Gupta and R. K. Jha, "A survey of 5G network: architecture and emerging technologies", *IEEE Access*, vol. 3, pp. 1206–1232, 2015 (doi: 10.1109/ACCESS.2015.2461602).

[10] K. Kowalik *et al.*, "Lessons learned from WiMAX deployment at INEA", *J. of Telecommun. and Inform. Technol.*, no. 3, pp. 34–41, 2014.

[11] B. Musznicki, K. Kowalik, P. Kołodziejski, and E. Grzybek, "Mobile and residential INEA Wi-Fi hotspot network", in *13th Int. Symp. on Wirel. Commun. Syst. ISWCS 2016*, Poznań, Poland, 2016 (arXiv:1608.06606).

[12] A. Kliks, B. Musznicki, K. Kowalik, and P. Kryszkiewicz, "Perspectives for Resource Sharing in 5G Networks", *Telecommun. Syst.*, vol. 68, no. 4, pp. 05–619, 2018 (doi: 10.1007/s11235-017-0411-3).

[13] L. Atzori, A. Iera, and G. Morabito, "From "smart objects" to "social objects": the next evolutionary step of the Internet of Things", *IEEE Commun. Mag.*, vol. 52, no. 1, pp. 97–105, 2014 (doi: 10.1109/MCOM.2014.6710070).

[14] C. Perera, A. Zaslavsky, P. Christen, and D. Georgakopoulos, "Context aware computing for The Internet of Things: A Survey", *IEEE Commun. Surv. & Tutor.*, vol. 16, no. 1, pp. 414–454, 2014 (doi: 10.1109/SURV.2013.042313.00197).

[15] M. Díaz, C. Martín, and B. Rubio, "State-of-the-art, challenges, and open issues in the integration of Internet of Things and cloud computing", *J. of Netw. and Comp. Appl.*, vol. 67, pp. 99–117, 2016 (doi: 10.1016/j.jnca.2016.01.010).

[16] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions", *Future Gener. Comp. Syst.*, vol. 29, no. 7, pp. 1645–1660, 2013 (doi: 10.1016/j.future.2013.01.010).

[17] C. P. Kruger and G. P. Hancke, "Implementing the Internet of Things Vision in Industrial Wireless Sensor Networks", in *Proc. 12th IEEE Int. Conf. on Industr. Informat. INDIN 2014*, Porto Alegre, Brazil, 2014, pp. 627–632 (doi: 10.1109/INDIN.2014.6945586).

[18] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A survey on enabling technologies, protocols, and applications", *IEEE Commun. Surv. & Tutor.*, vol. 17, no. 4, pp. 2347–2376, 2015 (doi: 10.1109/COMST.2015.2444095).

[19] B. Musznicki and P. Zwierzykowski, "Survey of simulators for wireless sensor networks", *Int. J. of Grid and Distrib.Comput.*, vol. 5, no. 3, pp. 23–50, 2012.

[20] M. Głąbowski, B. Musznicki, P. Nowak, and P. Zwierzykowski, "An in-depth discussion of challenges related to solving shortest path problems using ShortestPathACO based algorithms", in *Information Systems Architecture and Technology; Knowledge Based Approach to the Design, Control and Decision Support*, J. Świątek, L. Borzemski, A. Grzech, and Z. Wilimowska, Eds. Wrocław, Poland: Oficyna Wydawnicza Politechniki Wrocławskiej, 2013, pp. 77–88 (ISBN: 978-83-7493-802-0).

[21] M. Stein, T. Petry, I. Schweizer, M. Brachmann, and M. Mühlhäuser, "Topology control in wireless sensor networks: What blocks the breakthrough?", in *Proc. 41st Conf. on Local Comp. Netw. LCN 2016*, Dubai, United Arab Emirates, 2016, pp. 389–397 (doi: 10.1109/LCN.2016.67).

[22] D. Chaładyniak and J. Grzybowski, "Wybrane metody diagnozowania nieprawidłowości działania sieci teleinformatycznych", *Zeszyty Naukowe Warszawskiej Wyższej Szkoły Informatyki*, vol. 6, no. 8, pp. 61–76, 2012 [in Polish].

[23] M. Głąbowski, B. Musznicki, P. Nowak, and P. Zwierzykowski, "Shortest path problem solving based on ant colony optimization metaheuristic", *Int. J. of Image Process. & Commun., Special Issue: Algorithms and Protocols in Packet Networks*, vol. 17, no. 1–2, pp. 7–17, 2012 (doi: 10.2478/v10248-012-0011-5).

[24] Ł. Skibniewski and J. Furtak, "Zdalne laboratorium sieciowe", *Biuletyn Instytutu Automatyki i Robotyki*, vol. 32, 18, pp. 3–22, 2012 [in Polish].

[25] M. Stasiak and M. Michalski, "Algorytmy wspomagające projektowanie pierścieniowych sieci optycznych", in *Poznańskie Warsztaty Telekomunikacyjne PWT 2003*, Poznań, Poland, 2003, pp. 121–126 [in Polish].

[26] P. Santi, *Topology Control in Wireless Ad Hoc and Sensor Networks*. Chichester: Wiley, 2005.

[27] C. Schurgers, V. Tsiatsis, S. Ganeriwal, and M. Srivastava, "Topology management for sensor networks: exploiting latency and density", in *Proc. 3rd ACM Int. Symp. on Mob. Ad Hoc Network. & Comput. MobiHoc 2002*, Lausanne, Switzerland, 2002, pp. 135–145 (doi: 10.1145/513800.513817)

[28] M. Younis, I. F. Senturk, K. Akkaya, S. Lee, and F. Senel, "Topology management techniques for tolerating node failures in wireless sensor networks: A survey", *Comp. Netw.*, vol. 58, no. 1, pp. 254–283, 2014 (doi: 10.1016/j.comnet.2013.08.021).

[29] G. Sosnowski, "Przegląd algorytmów dynamicznego zarządzania topologią w bezprzewodowych, ruchomych sieciach ad hoc", in *Poznańskie Warsztaty Telekomunikacyjne PWT 2005*, Poznań, Poland, 2005 [in Polish].

[30] J. Zhao and G. Cao, "Robust topology control in multi-hop cognitive radio networks", *IEEE Trans. on Mob. Comput.*, vol. 13, no. 11, pp. 2634–2647, 2014 (doi: 10.1109/TMC.2014.2312715).

[31] K. Moon, D.-S. Yoo, W. Lee, and S.-J. Oh, "Receiver cooperation in topology control for wireless ad-hoc networks", *IEEE Trans. on Wirel. Commun.*, vol. 14, no. 4, pp. 1858–1870, 2015 (doi: 10.1109/TWC.2014.2374617).

[32] M. A. Labrador and P. M. Wightman, *Topology Control in Wireless Sensor Networks: With a Companion Simulation Tool for Teaching and Research*. Springer, 2009 (ISBN: 978-1-4020-9584-9).

[33] A. A. Aziz, Y. A. Sekercioglu, P. Fitzpatrick, and M. Ivanovich, "A survey on distributed topology control techniques for extending the lifetime of battery powered wireless sensor networks", *IEEE Commun. Surv. & Tutor.*, vol. 15, no. 1, pp. 121–144, 2013 (doi: 10.1109/SURV.2012.031612.00124).

[34] M. Li, Z. Li, and A. Vasilakos, "A survey on topology control in wireless sensor networks: taxonomy, comparative study, and open issues", *Proc. of the IEEE*, vol. 101, no. 12, pp. 2538–2557, 2013 (doi: 10.1109/JPROC.2013.2257631).

[35] E. Niewiadomska-Szynkiewicz, P. Kwaśniewski, and I. Windyga, "Comparative study of wireless sensor networks energy-efficient topologies and power save protocols", *J. of Telecommun. and Inform. Technol.*, no. 3, pp. 68–76, 2009.

[36] Y. Huang, J.-F. Martínez, V. H. Díaz, and J. Sendra, "A novel topology control approach to maintain the node degree in dynamic wireless sensor networks", *Sensors*, vol. 14, no. 3, pp. 4672–4688, 2014 (doi: 10.3390/s140304672).

[37] I. Yoon, D. K. Noh, and H. Shin, "Energy-aware hierarchical topology control for wireless sensor networks with energy-harvesting nodes", *Int. J. of Distrib. Sensor Netw.*, vol. 11, no. 6, 2015 (doi: 10.1155/2015/617383).

[38] B. Chen and L. Wang, "An interference prediction-based topology control algorithm for 3-D wireless sensor networks", *J. of Computat. Inform. Syst.*, vol. 7, no. 4, pp. 1198–1205, 2011 (doi: 10.1109/ICISE.2010.5689350).

[39] S. S. Dhillon and K. Chakrabarty, "Sensor placement for effective coverage and surveillance in distributed sensor networks, in *Proc. IEEE Wirel. Commun. and Network. WCNC 2003*, New Orleans, LA, USA, 2003, vol. 3 (doi: 10.1109/WCNC.2003.1200627).

[40] J. Ai and A. A. Abouzeid, "Coverage by directional sensors in randomly deployed wireless sensor networks", *J. of Combinator. Optimiz.*, vol. 11, no. 1, pp. 21–41, 2006 (doi: 10.1007/s10878-006-5975-x).

[41] B. Musznicki, M. Tomczak, and P. Zwierzykowski, "Dijkstra-based localized multicast routing in wireless sensor networks", in *Proc. 8th Int. Symp. on Commun. Syst., Netw. and Digit. Sig. Process. CSNDSP 2012*, Poznań, Poland, 2012 (doi: 10.1109/CSNDSP.2012.6292692).

[42] M. E. M. Campista and M. G. Rubinstein, *Advanced Routing Protocols for Wireless Networks*. Chichester: Wiley, 2014 (ISBN: 978-1-848-21627-3).

[43] Q. Mamun, "A qualitative comparison of different logical topologies for wireless sensor networks", *Sensors*, vol. 12, no. 11, pp. 14887–14913, 2012 (doi: 10.3390/s121114887).

[44] X. Liu, "A survey on clustering routing protocols in wireless sensor networks", *Sensors*, vol. 12, no. 8, pp. 11113–11153, 2012 (doi: 10.3390/s120811113).

[45] G. Werner-Allen *et al.*, "Deploying a wireless sensor network on an active volcano", *IEEE Internet Comput.*, vol. 10, no. 2, pp. 18–25, 2006 (doi: 10.1109/MIC.2006.26).

[46] H. Zhang and J. C. Hou, "Is deterministic deployment worse than random deployment for wireless sensor networks?", in *Proc. 25th IEEE Int. Conf. on Comp. Commun. INFOCOM 2006*, Barcelona, Spain, 2006, pp. 1–13 (doi: 10.1109/INFOCOM.2006.290).

[47] S. N. Simić and S. Sastry, "Distributed environmental monitoring using random sensor networks" in *Information Processing in Sensor Networks: Second International Workshop, IPSN 2003, Palo Alto, CA, USA, April 2003, Proceedings*, F. Zhao and L. Guibas, Eds. *LNCS*, vol. 2634, pp. 582–592. Berlin Heidelberg: Springer, 2003.

[48] K. Römer and F. Mattern, "The design space of wireless sensor networks", *IEEE Wireless Commun.*, vol. 11, no. 6, pp. 54–61, 2004 (doi: 10.1109/MWC.2004.1368897).

[49] A. Howard, M. J. Matarić, and G. S. Sukhatme, "An incremental self-deployment algorithm for mobile sensor networks", *Autonomous Robots*, vol. 13, no. 2, pp. 113–126, 2002 (doi: 10.1023/A:1019625207705).

[50] H. Gao *et al.*, "High speed data routing in vehicular sensor networks", *J. of Commun.*, vol. 5, no. 3, pp. 181–188, 2010 (doi: 10.4304/jcm.5.3.181-188).

[51] M. Wooldridge and N. R. Jennings, "Intelligent agents: theory and practice", *The Knowl. Engin. Rev.*, vol. 10, no. 2, pp. 115–152, 1995 (doi: 10.1017/S0269888900008122).

[52] U. Lee and M. Gerla, "A survey of urban vehicular sensing platforms", *Computer Netw.*, vol. 54, no. 4, pp. 527–544, 2010 (doi: 10.1016/j.comnet.2009.07.011).

[53] "IEEE Standard for Information technology – Telecommunications and information exchange between systems – Local and metropolitan area networks – Specific requirements, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications", IEEE Std 802.11™-2012, 29 March 2012 [Online]. Available: https://www.iith.ac.in/ tbr/teaching/docs/802.11-2007.pdf

[54] K. Kowalik, A. Kliks, B. Musznicki, M. Kołodziejski, and P. Kryszkiewicz, "Observation of WiMAX Radio Parameters to Enhance Spectrum Utilisation in Mixed Environment", *J. of Telecommun. and Inform. Technol.*, no. 1, pp. 42–50, 2018 (doi: 10.26636/jtit.2018.123917).

[55] H. Conceição, M. Ferreira, and J. Barros, "On the urban connectivity of vehicular sensor networks", in *Distributed Computing in Sensor Systems, 4th IEEE International Conference, DCOSS 2008, Santorini Island, Greece, June 11-14, 2008, Proceedings*, S. E. Nikoletseas, B. S. Chlebus, D. B. Johnson, and B. Krishnamachari, Eds. *LNCS*, vol. 5067, pp. 112–125. Berlin Heidelberg: Springer, 2008 (doi: 10.1007/978-3-540-69170-9_8).

[56] M. Boban, T. T. Vinhoza, M. Ferreira, J. Barros, and O. K. Tonguz, "Impact of vehicles as obstacles in vehicular ad hoc networks", *IEEE J. on Selec. Areas in Commun.*, vol. 29, no. 1, pp. 15–28, 2011 (doi: 10.1109/JSAC.2011.110103).

[57] A. Cardote, S. Sargento, and P. Steenkiste, "On the connection availability between relay nodes in a VANET", in *Proc. IEEE Globecom Worksh. GC Wkshps 2010*, Miami, FL, USA, 2010, pp. 181–185 (doi: 10.1109/GLOCOMW.2010.5700255).

[58] C. Ameixieira *et al.*, "HarborNet: A real-world testbed for vehicular networks", *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 108–114, 2014 (doi: 10.1109/MCOM.2014.6894460).

[59] "Creating The World's Largest Network of Connected Vehicles for Smart Cities" [Online]. Available: https://www.worldwifiday.com/wp-content/uploads/2016/05/3.-PortoCaseStudy_Letter_2016-04-15.pdf

[60] "IEEE Standard for Information technology – Telecommunications and information exchange between systems – Local and metropolitan area networks – Specific requirements", IEEE Std 802.11p™-2010, 15 July 2010.

[61] R. C. Shah, S. Roy, S. Jain, and W. Brunette, "Data MULEs: Modeling and Analysis of a Three-tier Architecture for Sparse Sensor Networks", *Ad Hoc Netw.*, vol. 1, no. 2, pp. 215–233, 2003 (doi: 10.1016/S1570-8705(03)00003-9).

[62] P. Santos *et al.*, "Demo abstract: Experiments on using vehicles as data mules for data collection from urban sensors", in *Proc. 12th Eur. Conf. on Wirel. Sensor Netw. EWSN 2015*, Porto, Portugal, 2015, pp. 17–18.

[63] "CityMobil2 Experience and Recommendations" [Online]. Available: https://www.polisnetwork.eu/CityMobil2%20booklet%20web%20final_17%2011%202016.pdf

[64] A. Alessandrini, A. Cattivera, C. Holguin, and D. Stam, "CityMobil2: Challenges and opportunities of fully automated mobility" in *Road Vehicle Automation*, G. Meyer and S. Beiker, Eds. Springer, 2014, pp. 169–184 (doi: 10.1007/978-3-319-05990-7).

[65] L. Krishnamurthy *et al.*, "Design and deployment of industrial sensor networks: experiences from a semiconductor plant and the North Sea", in *Proc. 3rd Int. Conf. on Embed. Network. Sensor Syst. SenSys 2005*, San Diego, CA, USA, 2005, pp. 64–75 (doi: 10.1145/1098918.1098926).

[66] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey", *Computer Netw.*, vol. 52, no. 12, pp. 2292–2330, 2008 (doi: 10.1016/j.comnet.2008.04.002).

[67] C. Gomez and J. Paradells, "Wireless home automation networks: A survey of architectures and technologies", *IEEE Commun. Mag.*, vol. 48, no. 6, pp. 92–101, 2010 (doi: 10.1109/MCOM.2010.5473869).

[68] T. Zachariah, N. Klugman, B. Campbell, J. Adkins, N. Jackson, and P. Dutta, "The Internet of Things has a gateway problem", in *Proc. 16th Int. Worksh. on Mobile Comput. Syst. and Appl. HotMobile 2015*, Santa Fe, New Mexico, USA, 2015, pp. 27–32 (doi: 10.1145/2699343.2699344).

[69] V. C. Gungor, D. Sahin, T. Kocak, S. Ergut, C. Buccella, C. Cecati, and G. P. Hancke, "Smart grid technologies: Communication technologies and standards", *IEEE Trans. on Industr. Informat.*, vol. 7, no 4, pp. 529–539, 2011 (doi: 10.1109/TII.2011.2166794).

[70] L. Quan-Xi and L. Gang, "Design of remote automatic meter reading system based on ZigBee and GPRS", in *Proc. 3rd Int. Symp. on Comp. Sci. and Computat. Technol. ISCSCT 2010*, Jiaozuo, China, 2010, vol. 2, pp. 186–189.

[71] Cisco openBerlin Innovation Center [Online]. Available: https://www.cisco.com/c/m/de_de/innovationcenter/berlin.html

[72] "IEEE Standard for Local and metropolitan area networks – Part 15.4: Low-Rate Wireless Personal Area Networks (LR-WPANs)", IEEE Std 802.15.4-2011, 5 September 2011 [Online]. Available: https://standards.ieee.org/standard/802_15_4-2011.html

[73] "Series G: Transmission Systems and Media, Digital Systems and Networks – Access networks – In premises networks – Short range narrow-band digital radiocommunication transceivers – PHY, MAC, SAR and LLC layer specifications", Recommendation ITU-T G.9959, 2015 [Online]. Available: https://www.itu.int/rec/T-REC-G.9959/en

[74] C. Gomez and J. Paradells, "Wireless home automation networks: A survey of architectures and technologies", *IEEE Commun. Mag.*, vol. 48, no. 6, pp. 92–101, 2010 (doi: 10.1109/COM.2010.5473869).

[75] D. R. Moogk, "Minimum viable product and the importance of experimentation in technology startups", *Technol. Innov. Manag. Rev.*, vol. 2, no. 3, pp. 23–26, 2012 (doi: 10.22215/timreview/535).

[76] P. Walkowiak, R. Szalski, B. Musznicki, D. Dudek, K. Kowalik, and P. Zwierzykowski, "Evaluation of CARMNET System in INEA HOTSPOT Network", in *Proc. IEICE Inform. and Commun. Technol. Forum ICTF 2014*, Poznań, Poland, 2014 (doi: 10.13140/2.1.2751.4567).

Bartosz Musznicki has been expanding his experience in the area of management and network architecture at INEA – a telecommunications operator active in the Greater Poland region. He is currently working for Capgemini in one of Europe's major commercial network projects. He is pursuing Ph.D. at the Poznan University of Technology, Poland. His main research interests include topology control and routing in wireless sensor networks. He is an author of four book chapters, eight journal articles and six conference papers.

https://orcid.org/0000-0002-7529-8898
E-mail: rsrch@musznicki.com
Chair of Communication and Computer Networks
Faculty of Electronics and Telecommunications
Poznan University of Technology
Polanka 3
61-131 Poznań, Poland

Paper

A P2P-based Communication Framework for Geo-Location Oriented Networks

Takumi Miyoshi[1], Yusuke Shimomura[2], and Olivier Fourmaux[3]

[1] *College of Systems Engineering and Science, Shibaura Institute of Technology, Saitama, Japan*
[2] *Graduate School of Engineering and Science, Shibaura Institute of Technology, Saitama, Japan*
[3] *Laboratoire d'Informatique de Paris 6, Sorbonne Université, Paris, France*

https://doi.org/10.26636/jtit.2019.131119

Abstract—This paper proposes a novel peer-to-peer communication framework to implement geographical location oriented networks, called G-LocON. Location-based services have been gaining in popularity, as proven by ridesharing and mobile games. Although these services have to construct geolocation oriented networks based on their users' geographical locations, they completely rely on client/server models to communicate with neighboring terminals. G-LocON provides geolocation oriented device-to-device communication only with the current wireless technologies, such as LTE and Wi-Fi, cooperating with the global positioning system and peer-to-peer overlay networking. G-LocON will serve as a type of a mobile ad-hoc network in which devices located within the focusing area are capable of communicating directly. We developed a primitive Android application to implement the G-LocON framework. Evaluation of the solution's performance has verified the usefulness of the proposed system that offers an admissive transmission delay. Moreover, to confirm the application-related potential of the G-LocON framework, we also show a practical map software in which all neighboring mobile devices present in the focusing area are displayed.

Keywords*—*geolocation oriented network, location-based service, overlay, peer-to-peer.

1. Introduction

In the past decade, smartphones have been expanding the range of their functionalities and have been gaining in popularity - not only mobile phones, but also as portable computing devices. Some reports forecast that the number of smartphone users in the world will reach 2.7 billion and will exceed 50% of all mobile users in 2019 [1], [2]. One of the remarkable smartphone functions is the global positioning system (GPS). GPS is capable of pinpointing the geographical location of users based on signals received from four or more GPS satellites, without any data transmission from GPS receivers. Ohmae defines a generation-based evolution model for the location information (LI) business [3]. The history of LI started in the 1990s with digital map services, including car navigation and digital map services on personal computers. In this generation, LI was used personally and privately to display the user's location on the map. In the 2000s, the advent of social networking services (SNS) and smartphones has brought about the next generation solution – LI 2.0. Users commonly provide their own LI to smartphone applications, some of which send LI to servers, where it is registered, and may therefore enjoy geolocation-based services (LBS). Ohmae estimates that we are now opening the door to the LI 3.0 era, in which LBS will be much improved and enhanced by Internet of Things (IoT) devices and sensors around us.

As far as geolocation-based applications are concerned, digital maps came first, route search and navigation still remain the areas in which the service is much more popular and useful. SNS applications, such as Facebook, Twitter, and Instagram, are also installed on most user smartphones. By tagging and sharing the users' locations, trajectories and data on where they are/were or what they intend to visit, the applications may easily find useful information, such as reputations and recommendations of shops and popular places. Ridesharing services, such as Uber and Grab, integrate LI and SNS. Digital map applications show the user's position, as well as the vehicles available around the user, and social networks establish trust and accountability between passengers and drivers [4], [5]. Furthermore, augmented reality (AR) has become, in recent years, a popular application that relies on the user's LI. Sekai Camera [6] has brought about a new world where, through the smartphone camera, we can see the messages and photos that are saved and linked to the real world by means of LI tags assigned by other users. Niantic [7] launched megahit games, Ingress and Pokemon Go, and is the world's leader in AR gaming.

Following in the footsteps of the LI services referred to above, inter-vehicular communication is one of the promising core technologies in intelligent transportation systems (ITS) [8], [9]. In general, vehicular-related communication forms a vehicular ad hoc network (VANET) among neighboring devices by utilizing radio propagation-based direct communication [10] and infrastructure-based communication [11]. VANET thus realizes a close-range direct or multi-hop communication among vehicles and other peripherals [12].

As described above, geolocation-oriented communications have good prospects for the future. However, there are some problems and challenges in LBS. Firstly, the current location-based applications are mainly realized as server-based systems. User devices and their LI are registered on and maintained at the location servers, and each device has to connect to the servers to obtain information about objects around the user. The computation- and storage-related load would be therefore concentrated on the servers. Secondly, although multi-hop communication has certainly been a promising technology for more than 20 years now, no one knows when it is due to become widespread and installed on consumer devices. Several research groups have attempted, in recent years, to evaluate VANET communications using IEEE 802.11p [13]–[15]. However, most conventional studies focus solely on fundamental characteristics of packet transmissions between vehicles, such as packet delivery ratio, transmission delay and jitter. Moreover, IEEE 802.11p devices have not yet become common in consumer products. The VANET system will work well after all vehicles have been equipped with wireless communications solutions, and it has to be borne in mind that people do not tend to replace their cars with new ones very frequently.

In this paper, we propose G-LocON – a novel communication framework that may be relied upon to establish geographical location oriented networks. G-LocON constructs a peer-to-peer (P2P) logical overlay network around each user. We implement a location tracker, which maintains LI of the user's devices and helps discover neighboring devices. Therefore, the proposed framework provides geolocation oriented device-to-device communication. Since the system relies solely only on such current technologies as LTE, Wi-Fi and GPS, the public may enjoy this service of the future on their smartphones without any delay.

The remainder of the paper is organized as follows. Section 2 briefly discusses previous works related to LBS and location-based P2P networks. In Section 3, we explain the G-LocON framework and describe the proposed protocols, as well as an implementation method. Section 4 shows a primitive application to realize the G-LocON framework and reports the results of its experimental evaluations in the actual environment, from the viewpoint of transmission delay. In Section 5, in order to confirm the potential of the G-LocON framework, we will demonstrate a practical Android application in which all neighboring mobile devices positioned within the focusing area are displayed on a peripheral street map. Conclusions and suggestions for future work are provided in Section 6.

2. Related Work

LBS is one of the recent and hot topics and it has succeeded in attracting much attention from numerous researchers [16]–[24]. Popular research objectives are mainly related to LBS frameworks [16]–[18] and LBS applications [19], [20], [23], [24]. In the former category, authors propose new network architectures or frameworks to provide location-based services. The latter category, on the other hand, develops new applications, such as mobile positioning or navigation systems.

Some articles utilize the P2P approach, which we also focus on in this paper, to construct logical overlay networks to discover neighbor peers [21]–[24]. Kaneko *et al.* propose a tree-based P2P overlay structure called LL-Net [21]. They assume a mixed environment where both fixed and mobile devices coexist, and try to form a three-layer tree structure on a traditional grid-cutting region model. LL-Net was evaluated for content retrieval by means of a computer simulation. Kovačević *et al.* also propose a hierarchical tree-based P2P overlay called Globase.KOM [22]. They focus on a world-wide spanning network, which is divided into non-overlapping rectangular zones. Simulation results show that Globase.KOM achieves full retrievability of area searches and a short response time. Wang *et al.* develop a P2P mobile navigation system to guide visitors in a flora exposition with 3D renders [23], [24]. P2P technology is used for clustering neighboring users who have similar interests and for downloading data from them. They also validate faster transmission rates of the proposed system by computer simulation, assuming all devices are connected to Wi-Fi.

The conventional location-based P2P systems mentioned above still suffer from certain problems. Both LL-Net and Globase.KOM target wide-area content retrievals and then form hierarchical tree-based overlay networks. This approach is costly in terms of finding super peers and constructing hierarchical networks when the focus is placed on a small area around each user. Moreover, all systems mentioned above seem not to consider real-time communications between users. In addition, they can hardly be implemented as-is in the current mobile network environment due to the private addressing system they rely on. We believe that a hybrid P2P system in which some servers exist for maintaining peers helps identify neighbor peers in a quicker manner. For reference, a hybrid P2P system is a certain kind of an implementation form and is very commonly used in popular P2P file sharing and video streaming applications, such as BitTorrent, PPTV and PPStream.

3. G-LocON Communication Framework

3.1. Overview of G-LocON

This paper proposes a P2P-based communication framework to realize a geographical location oriented network, and is called *G-LocON*. An outline image of the proposed system is shown in Fig. 1. Here, mobile devices, i.e. peers, are assumed to connect to various networks: they connect to cellular networks, Wi-Fi and so on. G-LocON provides a hybrid P2P network by implementing a location tracking server called *L-tracker*. L-tracker maintains the peers' information that consists of peer identifiers and locations, and helps each peer discover the neighbor peers around it.

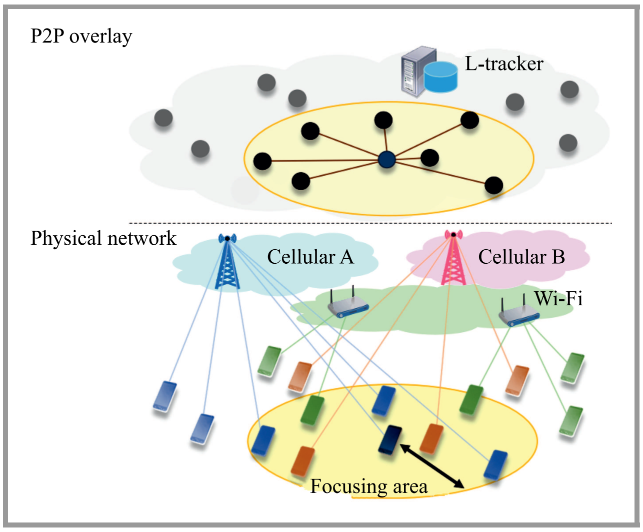


Fig. 1. G-LocON framework: forming a P2P overlay managed by L-tracker on interconnected physical networks.

G-LocON, by nature of a hybrid P2P system, may flexibly manage such unstable logical overlay networks formed by the dynamically moving focusing areas. Furthermore, each user is free to decide the size of its focusing area, since the range within which communication with neighbor peers may be established is unlimited. All peers may connect to each other through the Internet, independently of user density. This is a major advantage over ad hoc networks in which mobile devices hardly form a network in an area with sparsely located users, unless they are located mutually within the range of their wireless communication systems.

When a new peer joins the proposed system, it sends, firstly, a join message with peer information and its LI to the L-tracker. The peer then requests neighbor peers information from the L-tracker, and finally establishes P2P connections to its neighbors. In the proposed system, every peer periodically registers its LI with the L-tracker by reusing the join message, obtains renewed information about neighbor peers and updates the connections to its neighbors. After establishment of P2P connections, the peer directly communicates with its neighbors without relaying data through any servers. This is a typical behavior of hybrid P2P systems.

3.2. How to Form the G-LocON Overlay

To establish a direct connection between two mobile devices, a peer needs to know the translated IP address and port number of the destination peer. Such information is called *address bindings* and is automatically allocated when a packet crosses a NAT gateway. In this paper, a STUN server [25] is deployed to resolve the bindings information. Moreover, peers also have to exchange metadata to coordinate communication. Therefore, we introduce a signaling server to share the information about peers and to coordinate connections between them.

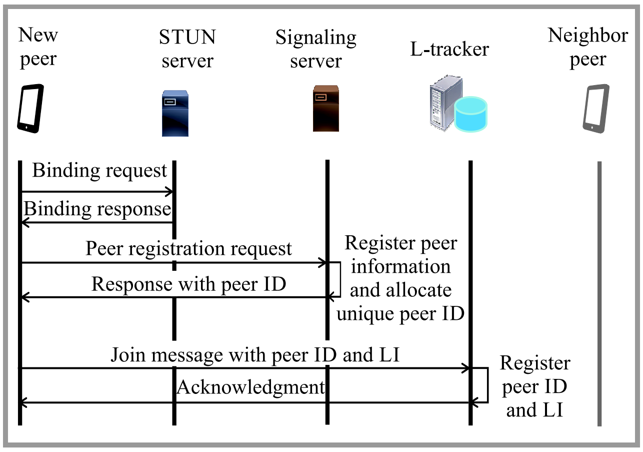


Fig. 2. Process of joining G-LocON.

The process of joining G-LocON. Figure 2 shows the communication process that takes place when a new peer joins the G-LocON overlay network. The sequence of process is described as follows:

- Firstly, the new peer sends a binding request to the STUN server. The server provides a binding response that consists of the IP address and port number, as observed from the server's perspective;
- The new peer then sends a registration request to the signaling server. The request message contains the peer's IP address and port number resolved by the STUN server. The server allocates a unique identifier to the peer (peer ID) and registers these three types data in the database. The server also sends back the peer ID to the peer as a response;
- The peer sends a join message to the L-tracker. The message includes the peer ID and LI that is obtained by the GPS module. The L-tracker then registers them in the database and replies with an acknowledgment. This step will be periodically repeated for updating LI at intervals of T_{int}.

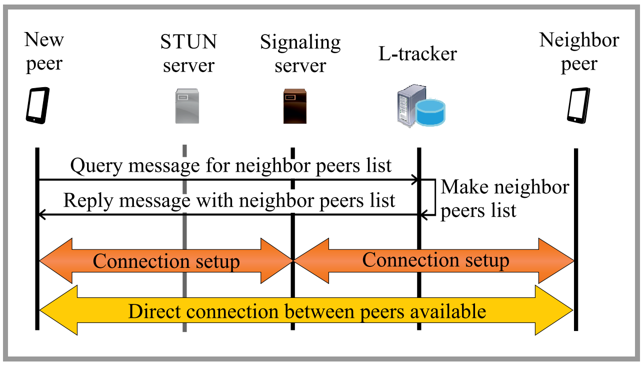

New peer
STUN server
Signaling server
L-tracker
Neighbor peer
Query message for neighbor peers list
Reply message with neighbor peers list
Make neighbor peers list
Connection setup
Connection setup
Direct connection between peers available

Fig. 3. P2P connection process.

Establishment of P2P connection between peers. After the joining process has been completed, the peer becomes ready to establish P2P connections. Figure 3 shows the communication process. When a peer sends a neighbor search query to the L-tracker, the focusing area to be searched has to be clearly indicated. In this paper, we determine that each focusing area is a circular form that is defined by a center and a radius. The center point may be the peer's location, and the radius represents the size of the focusing area. The sequence of the P2P connection process is described as follows:

- The peer sends a query message to the L-tracker to discover neighbor peers. The message includes the focusing area information, such as the peer's LI and a radius. The L-tracker makes a list of neighbor peers found in the focusing area and then sends it back to the peer, as a reply message. The peer maintains a list which consists of peer IDs only;

- Based on the list of neighbor peers, the peer sets up P2P connections with its neighbor peers, relying on the assistance of the signaling server. This process enables the peer to obtain the IP addresses and port numbers of each neighbor peer. The server sends another signaling message to each neighbor peer to inform it of the connection request from the new peer, as well as to provide its IP address and port number. Herewith, the two peers may establish the P2P connection and may subsequently communicate with each other any time while the connection remains alive. Each peer maintains the list of P2P connections;

- The peer closes its P2P connections if the neighbor peers that have already been connected to disappear from the list of neighbor peers.

If there is a particular type of NAT between two peers, the connection cannot be established by the above sequence. In this case, the peers try to connect through a TURN server [26]. Moreover, the P2P connection process is executed after each periodic repetition join messages sent to L-tracker. The interval time is thus T_{int}.

The process of leaving G-LocON. Figure 4 shows the communication process when a peer leaves the G-LocON overlay network. The sequence of the leaving process is described as:

- The peer first closes all the P2P connections with its neighbor peers, by sending connection closing messages;

- The peer sends a leave message to the L-tracker. The L-tracker then deletes relevant information from the database;

- The peer sends a peer unregistration request to the signaling server. The server then deletes relevant information from the database.

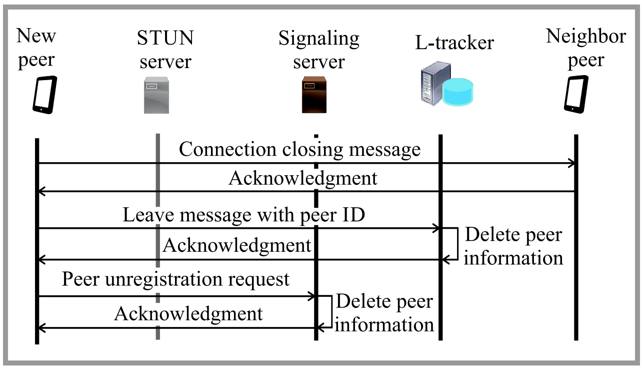


Fig. 4. The process of leaving G-LocON.

Avoidance of multiple P2P connections. If two peers attempt, virtually simultaneously, to initiate P2P connections with each other, multiple connections may be established occasionally. We therefore prepare a process to avoid such multiple connections. When a peer detects a new P2P connection with a neighbor peer, the peer checks if another connection with the same peer ID exists in the list of P2P connections. If it does, the new connection is immediately closed. Otherwise, the list is updated by adding information about the new connection.

3.3. Implementation

We implemented the G-LocON framework to establish geolocation-based P2P overlay networks. The L-tracker was developed on NIFCLOUD, a public cloud computing service provided by Fujitsu Cloud Technologies [27]. We also used SkyWay API, a WebRTC platform provided by NTT Communications [28]. SkyWay also offers the STUN, TURN, and signaling services. The mobile software was developed with the use of Android Studio with NIFCLOUD and SkyWay APIs. Asus ZenFone 3 / 3 Laser / 2 Laser, LG Nexus 5 / 5X, Samsung Galaxy S6, and Sony Xperia ZL2 were used as Android-based smartphones.

In the application developed, each peer obtains its LI from the GPS module every second, as well as registers and updates LI to the L-tracker every five seconds: $T_{\text{int}} = 5$ s. After establishing P2P connections with neighbor peers, each peer tries to share and send LI to its neighbors as soon as LI has been updated, i.e., every second. Each peer will maintain LI received from its neighbor peers.

4. Performance Evaluation

4.1. Experimental Settings

To evaluate the performance of G-LocON, we performed experiments to form a P2P overlay network comprising several Android smartphones operating in the real-world environment. In the experiments, we focused primarily on communication and processing delays. Here, t_l, t_s, and t_r denote the times when a peer obtains its LI from the GPS module, when the peer starts sending LI to each neighbor peer, and when the neighbor peer receives LI, respectively.

These times were recorded in each smartphone by relying on the GPS signal and its internal clock that had been adjusted beforehand by the network time protocol (NTP). $T_{\text{send}} = t_r - t_s$ represents the transmission delay required for a peer to send its LI to each neighbor peer. Meanwhile, $T_{\text{total}} = t_r - t_l$ represents the total delay time from when a peer obtains its LI to when LI arrives at each neighbor. In this paper, we evaluate T_{send} and T_{total}.

Two situations have been considered, namely a static and a dynamic scenario. In the static case, two to nine smartphones were placed in the laboratory, next to a window, to enable GPS signals to be received without any problems, and were connected to our university's Wi-Fi network. The G-LocON overlay established by those devices formed a full mesh topology since all peers were within the focusing areas of the other peers. Therefore, the smartphones mutually regarded each other as their neighbor peers. In one experiment set, each peer sends its LI to the remaining peers 300 times. Three sets have been performed and the mean values of T_{send} and T_{total} were calculated. In the dynamic case, two to five persons randomly moved with their smartphones, walking with a specified open-air area. The devices connected to LTE cellular networks provided by well-known Japanese mobile service providers, au (KDDI) and IIJ mobile (MVNO on NTT DOCOMO). The form of the G-LocON overlay was not always of the full-mesh variety, since some peers occasionally moved out of the focusing areas of other peers. We performed only one experimental set and calculated the mean values of T_{send} and T_{total}.

4.2. Results

Figures 5 and 6 present the average transmission delay T_{send} and the average total delay T_{total} in the static case, respectively. In a similar fashion, Figs. 7 and 8 show the results in the dynamic case. In each figure, we indicate three types of delay values: the overall average delay (blue line and circles), the average delay when peers obtain the updated list of neighbor peers from the L-tracker (red line and squares), and the average delay when the updated list is not obtained (green line and triangles).

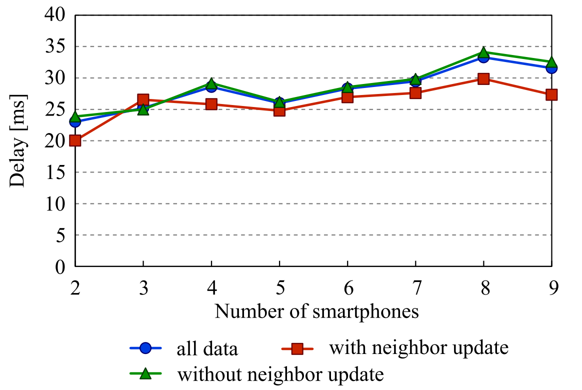


Fig. 5. Average transmission delay T_{send} (static case).

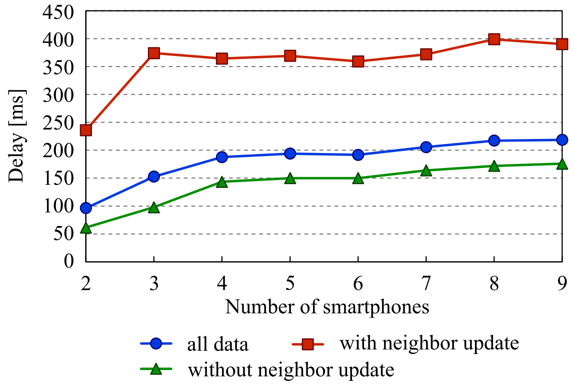


Fig. 6. Average total delay T_{total} (static case).

As shown in Fig. 5, the average transmission delay T_{send} becomes gradually longer with an increase in the number of devices. This is because both transmission and processing load of each peer become large as the number of neighbor peers increases. On the other hand, the delay when peers receive the updated list of neighbor peers (red line) is lower than the delay when the list is not obtained (green line). For updating the neighbor peers, each peer connects to the L-tracker at intervals of T_{int}. Since this process is executed sequentially in our current implementation, the peer does not communicate with its neighbor peers; and consequently the processing load will be reduced. From Fig. 6, it is observed that the average total delay T_{total} becomes much larger than the transmission delay T_{send}. In particular, the red line is roughly twice as high as the blue line. The total delay represented by the red line includes the processes required to obtain the list of neighbor peers from the L-tracker and to establish P2P connections with the neighbors. These processes take 170 to 270 ms, with the said values obtained as differences between red and green lines. Based on the results presented above, we are of the opinion that the performance of G-LocON depends strongly on communication with L-tracker and on the establishment of P2P connections.

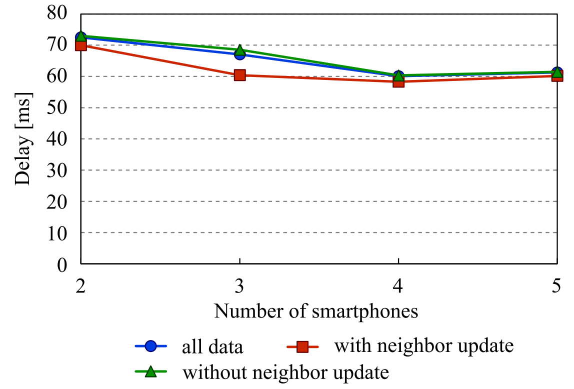


Fig. 7. Average transmission delay T_{send} (dynamic case).

Next, we turn to the dynamic case. It may be seen in Fig. 7 that the average transmission delay T_{send} reaches more than twice the value observed in the static case. This is because a longer network delay was added, since all smartphones were connected to LTE networks. However, the value does not deteriorate along with an increase in the number of devices. This phenomenon may be potentially explained by the fact that the cellular network connectivity varied by the hour, and the bandwidth available changed accordingly. To be honest, however, further experiments and thorough measurements are needed with more smartphone devices in the dynamic case.

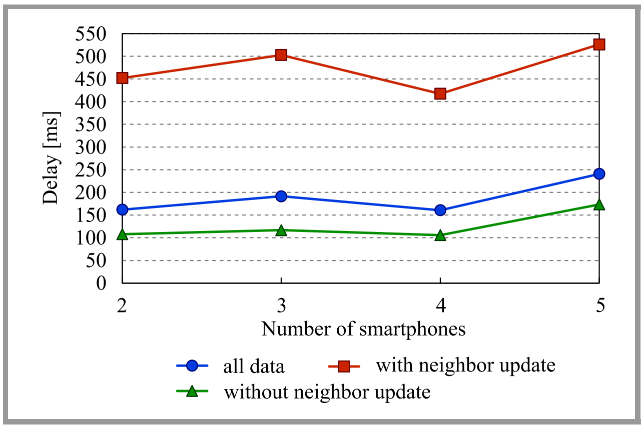


Fig. 8. Average total delay T_{total} (dynamic case).

As shown in Fig. 8, the average total delay T_{total} has similar characteristics to those observed in the static scenario: The total delay when the neighbor peers list was updated is much longer than in the case without the update. This also makes it clear that communication with the L-tracker and the establishment of P2P connections take a relatively long time compared with direct P2P communications between the peers.

As far as this aspect is concerned, we carefully examined the log data related to a situation in which three smartphone devices formed the G-LocON overlay in the dynamic case. We focused our attention on how long it takes for a peer to completely establish a P2P connection after it receives the list of its neighbor peers. From the observation, we found that cases existed in which the time required equaled several seconds. The time scale of this waiting time cannot be neglected when the G-LocON framework is used for real-world applications, such as inter-vehicular communications. Since the delay time will probably increase with the number of peers, implementation of the protocol to communicate with the L-tracker and the processing methods relied upon to establish P2P connections should be improved considerably.

5. Practical Application of G-LocON Framework

5.1. Application Design

To confirm the potential of the proposed G-LocON framework, we developed practical software for Android smartphones. We believe that G-LocON may be applied in inter-vehicular or in vehicle-to-pedestrian communications. In the developed application, assuming that both drivers and pedestrians use their own smartphones, the locations and moving speeds of neighbor peers present in the focusing area are shown on a peripheral street map.

The principal operation sequence of the developed software is based on the processes explained in Subsection 3.2, as described below:

1. A peer first joins the G-LocON overlay network. The joining process is repeated every five seconds: $T_{\text{int}} = 5$ s;
2. The peer sends a query message to the L-tracker to discover neighbor peers, and then obtains the list of neighbor peers in the circular focusing area whose radius equals 50 m;
3. The peer sets up P2P connections with its neighbors. They exchange their actual locations and moving velocities, relying on P2P communication, every second. The moving velocity of peer i, denoted by $\boldsymbol{v}_i(t)$, can be calculated as the difference between current and previous position vectors,

$$\begin{aligned}\boldsymbol{v}_i(t) &= \boldsymbol{P}_i(t) - \boldsymbol{P}_i(t-\Delta t)\\ &= \big[x_i(t) - x_i(t-\Delta t),\;\; y_i(t) - y_i(t-\Delta t)\big],\end{aligned}$$

   where $\boldsymbol{P}_i(t) = (x_i(t),\; y_i(t))$ represents position of peer i's at time t. Δt represents the interval in which LI is obtained the GPS module, and $\Delta t = 1$ s.
4. The peer's current location and circular focusing area are displayed on the peripheral map by relying on the Google map API. All neighbor peers are plotted concurrently as pins on the map. If the moving speed of a neighbor peer is v_{th} faster than own speed, the color of pin becomes red to alert the user about a moving vehicle approaching in the vicinity. The other neighbors are shown as green pins. In this implementation, v_{th} is set to 10 km/h.
5. When the user taps a pin that represents a neighbor peer, a popup balloon (known as *toast* in Android) appears to show the following information: peer's name, location, and moving speed.

5.2. Field Experiment

To confirm the behavior of the application, we performed a small field experiment. The software developed was installed on seven Android smartphones. All devices connected to the Internet using LTE cellular networks or Wi-Fi tethering via other mobile phones' hotspots.

Figure 9 shows the experimental environment in a residential area. In the experiment, two vehicles and five pedestrians used the application simultaneously. They can recognize one another as a neighbor peer if they enter the focusing areas of other peers. Two vehicles are approaching the intersection from outside of the pedestrians' focusing

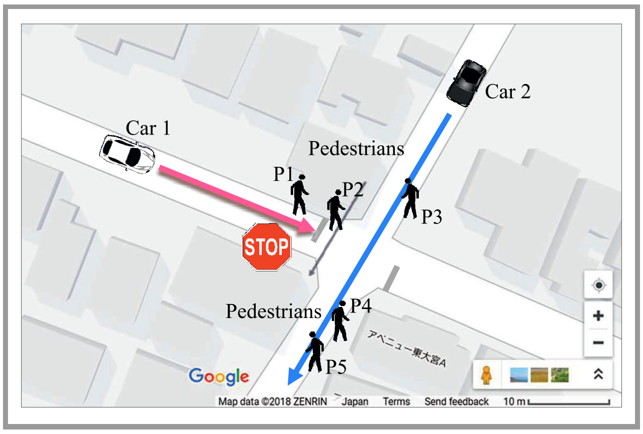


Fig. 9. Experimental environment for practical applications.

areas. Car 1 has to stop at a stop sign and wait for car 2 to by. Both vehicles obviously have to be careful and mind the pedestrians.

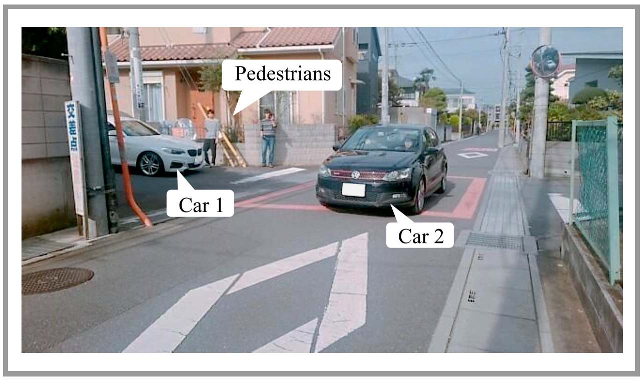
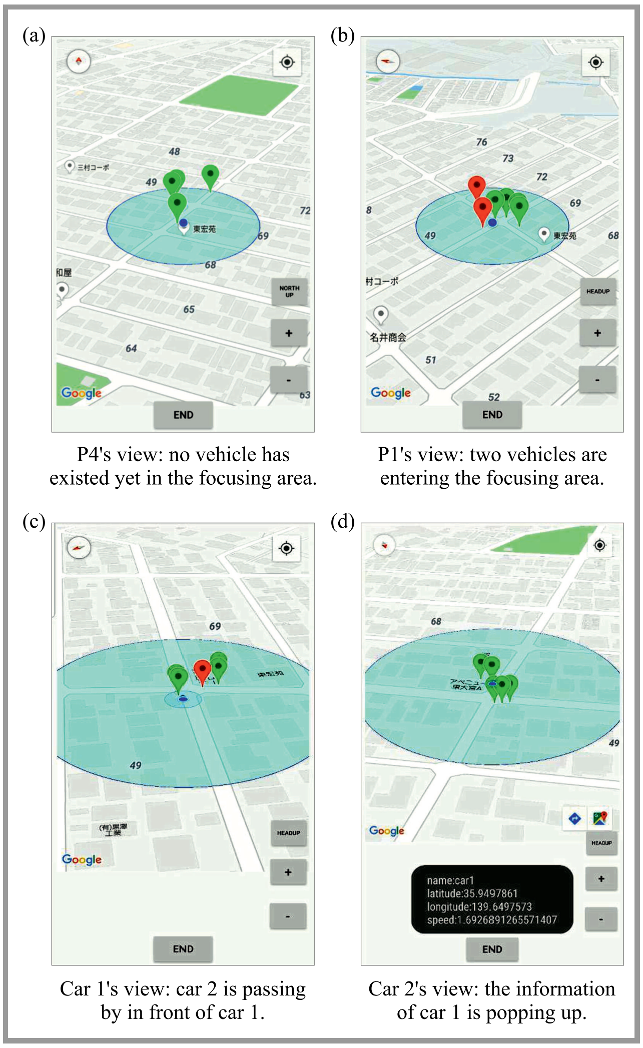


Fig. 10. A scene of the experiment when car 2 is passing an intersection: pedestrian P4's view.

The field experiment is shown in Fig. 10, with the image taken from the location of pedestrian P4. In this figure, car 2 is running through the intersection in front of car 1 which stops at the stop sign. The screenshots of Android smartphones are shown in Fig. 11. Figure 11a is P4's screenshot, and it only displays four other pedestrians because the vehicles have not entered the focusing area yet. On the contrary, Fig. 11b shows that two vehicles are entering the focusing area and are plotted by red pins. The users are therefore aware of the fact that moving vehicles are approaching. Figure 11c shows that the driver of car 1 is aware of car 2 running from left to right, as well as of some pedestrians. This application will certainly help drivers recognize other vehicles and pedestrians even at intersections with poor visibility. Finally, in Fig. 11d, a popup balloon appears to show information about car 1 when the user taps the pin of car 1. The peer's name, location and moving speed are obtained through P2P communication between car 1 and car 2.

Although the application developed was rudimentary in nature, the screenshots shown above clearly suggest the potential of the G-LocON framework. Namely, the proposed P2P mobile network overlay successfully conducts geolocation-based device-to-device communication relying only on current technologies. We are confident that the G-LocON framework will soon bring us to future LI services, such as inter-vehicular or vehicle-to-pedestrian communications.



P4's view: no vehicle has existed yet in the focusing area.

P1's view: two vehicles are entering the focusing area.

Car 1's view: car 2 is passing by in front of car 1.

name:car1
latitude:35.9497861
longitude:139.6497573
speed:1.6926891265571407

Car 2's view: the information of car 1 is popping up.

Fig. 11. Smartphone screenshots.

6. Conclusions

In this paper, we proposed a G-LocON framework to establish a P2P-based geolocation-oriented network. By using the proposed scheme, each mobile device can easily discover its neighbors around and may then form a G-LocON overlay network. We firstly developed a primitive Android application to implement the G-LocON framework by utilizing public cloud computing and WebRTC platform services. The results obtained from the evaluation experiments can be summarized as follows:

- The average transmission delay between peers, represented as T_{send}, is of the order of tens of milliseconds, and gradually increases with the number of neighbor peers;

- The average total delay from the moment a peer obtains its LI from the GPS module to the moment LI arrives at each neighbor, represented as T_{total}, is of the order of hundreds of milliseconds, and gradually increases with the number of neighbor peers;
- When peers obtain the updated list of neighbor peers from the L-tracker, the total delay becomes much longer than that without the update. Several seconds are occasionally needed to initiate a P2P connection after the neighbor peers list arrives.

We also developed another practical Android application based on the G-LocON framework, which realizes P2P-based direct neighbor communications to exchange information about peers between vehicles and pedestrians. Although the functionalities of the application are elementary, the small field experiment with vehicles and pedestrians has given an indication of the future potential of the G-LocON framework.

By analyzing the results of experiments described above, we have identified some problems concerning the G-LocON framework that require to be solved. Firstly, the protocol relied upon to communicate with the L-tracker and the processing methods establishing P2P connections should be much improved for larger-scale real-time applications. We will strive to introduce parallel processing in the communications between the L-tracker and neighbor peers. Secondly, public cloud services probably have their limits, such as the maximum number of connections, processing power and transmission speeds. We are now trying to implement our own L-tracker, which operates independently, without any cooperation with public cloud platforms. Finally, since the practical Android application developed in this paper was of the rudimentary nature, it was only capable of exchanging basic information between neighbors. We will strive to develop a more sophisticated application that may be used in a wide variety of environments.

In addition, we will evaluate the characteristics of the proposed system when user density changes.

Acknowledgment

This study was supported by JSPS Kakenhi Grant Number 17K06441.

References

[1] Statistica, "Number of smartphone users worldwide from 2014 to 2020", 2018 [Online]. Available: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide

[2] Statistica, "Number of mobile phone users worldwide from 2013 to 2019", 2017 [Online]. Available: https://www.statista.com/statistics/274774/forecast-of-mobile-phone-users-worldwide

[3] K. Ohmae, "Business model in location information 3.0 era", Kenichi Omae's special lecture, biblion [Online]. Available: https://biblion.jp/articles/RTuqy [in Japanese]

[4] A. Amey, J. Attanucci, and R. Mishalani, "Real-time ridesharing – the opportunities and challenges of utilizing mobile phone technology to improve rideshare services", *Transportation Res. Record: J. of the Transport. Res. Board*, vol. 217, no. 1, pp. 103–110, 2011 (doi: 10.3141/2217-13).

[5] S. Ma, Y. Zheng, and O. Wolfson, "Real-time city-scale taxi ridesharing", *IEEE Trans. Knowl. & Data Eng.*, vol. 27, no. 7, pp. 1782–1795, 2015 (doi: 10.1109/TKDE.2014.2334313).

[6] Tonchidot, "Introduction of sekai camera", Youtube video, Nov. 2010 [Online]. Available: https://www.youtube.com/watch?v=oxnKOQkWwF8

[7] Niantic, Inc., homepage [Online]. Available: https://www.nianticlabs.com

[8] H. Hartenstein and K. P. Laberteaux, "A tutorial survey on vehicular ad hoc networks", *IEEE Commun. Mag.*, vol. 46, no. 6, pp. 164–171, 2008 (doi: 10.1109/MCOM.2008.4539481).

[9] B. T. Sharef, R. A. Alsaquor, and M. Ismail, "Vehicular communication ad hoc routing protocols: a survey", *J. of Netw. & Comp. Appl.*, vol. 40, pp. 363–396, 2014 (doi: 10.1016/j.jnca.2013.09.008).

[10] D. Jiang and L. Delgrossi, "IEEE 802.11p: towards an international standard for wireless access in vehicular environments", in *Proc. IEEE Veh. Technol. Conf. VTC-Spring*, Singapore, 2008 (doi: 10.1109/VETECS.2008.458).

[11] Qualcomm, "Creating a digital 6th sense with LTE direct", 2015 [Online]. Available: https://www.qualcomm.com/media/documents/files/creating-a-digital-6th-sense-with-lte-direct.pdf

[12] L. Bariah, D. Shehada, E. Salahat, and C. Y. Yeun, "Recent advances in VANET security: a survey", in *Proc. 82nd IEEE Veh. Technol. Conf. VTC2015- Fall*, Boston, MA, USA, 2015 (doi: 10.1109/VTCFall.2015.7391111).

[13] F. A. Teixeira, V. F. e Silva, J. L. Leoni, D. F. Macedo, and J. M. S. Nogueira, "Vehicular networks using the IEEE 802.11p standard: an experimental analysis", *Veh. Commun.*, vol. 1, no. 2, pp. 91–96, 2014 (doi: 10.1016/j.vehcom.2014.04.001).

[14] M. E. Renda, G. Resta, P. Santi, F. Martelli, and A. Franchini, "IEEE 802.11p VANets: experimental evaluation of packet inter-reception time", *Comp. Commun.*, vol. 75, no. 1, pp. 26–38, 2016 (doi: 10.1016/j.comcom.2015.06.003).

[15] T. T. Almeida, L. C. Gomes, F. M. Ortiz, J. G. R. Júnior, and L. H. M. K. Costa, "IEEE 802.11p performance evaluation: simulations vs. real experiments", in *Proc. 21st IEEE Int. Conf. Intelli. Transport. Syst. ITSC 2018*, Maui, Hawaii, USA, 2018, pp. 3840–3845 (doi: 10.1109/ITSC.2018.8569676).

[16] D. Wang, Z. Li, and Y. Chen, "Design and implementation of a location based service business management platform", in *Proc. 4th Int. Conf. on Syst. & Informat. ICSAI 2017*, Hangzhou, China, 2017, pp. 1631–1635 (doi: 10.1109/ICSAI.2017.8248545).

[17] C.-M. Huang, D.-T. Dao, and C.-M. Mai, "Location-based service (LBS) data sharing using the k-member-limited clustering mechanism over the 4G and WiFi hybrid wireless mobile network", in *Proc. Int. Conf. Inform. Netw. ICOIN 2017*, Da Nang, Vietnam, 2017 (doi: 10.1109/ICOIN.2017.7899550).

[18] S. Wang *et al.*, "N-in-one: a novel location-based-service", *IEEE Trans. on Veh. Technol.*, vol. 67, no. 6, pp. 5274-5286, 2018 (doi: 10.1109/TVT.2017.2737017).

[19] S. Ishida *et al.*, "Implementation of on-demand indoor location-based service using ad-hoc wireless positioning network", in *Proc. 11th Int. Conf. Ubiquitous Intell. & Comput. (UIC 2014) and 11th Int. Conf. on Autonom. and Trust. Comput. and 14th Int. Conf. on Scal. Comput. and Commun. and its Assoc. Worksh.*, Bali, Indonesia, 2014, pp. 34–41 (doi: 10.1109/UIC-ATC-ScalCom.2014.96).

[20] A. Mena *et al.*, "Interactive geo-location based service application as pervasive computing through mobile devices", in *Proc. Chilean Conf. on Elec., Electron. Engin., Inform. & Commun. Technol. CHILECON 2017*, Pucon, Chile, 2017 (doi: 10.1109/CHILECON.2017.8229529).

[21] Y. Kaneko, K. Harumoto, S. Fukumura, S. Shimojo, and S. Nishio, "A location-based peer-to-peer network for context-aware services in a ubiquitous environment", in *Proc. Symp. Appl. & Internet Worksh. SAINT-W 2005*, Trento, Italy, 2005 (doi: 10.1109/SAINTW.2005.1620013).

[22] A. Kovačević, N. Liebau, and R. Steinmetz, "Globase.KOM – a P2P overlay for fully retrievable location-based search", in *Proc. 7th IEEE Int. Conf. on Peer-to-Peer Comput. P2P 2007*, Galway, Ireland, 2007, pp. 87–94 (doi: 10.1109/P2P.2007.18).

[23] C.-S. Wang, W.-D. Chen, and C.-L. Chen, "Location-based P2P mobile navigation system", in *Proc. Int. Symp. Compu. Sci. & Soc. ISCCS 2011*, Kota Kinabalu, Malaysia, 2011 (doi: 10.1109/ISCCS.2011.104).

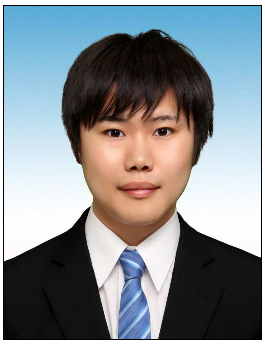
[24] C.-S. Wang, C.-L. Chen, and D.-J. Deng, "P2P-based mobile navigation system with location service", *Peer-to-Peer Network. and Appl.*, vol. 8, no. 1, pp. 22–31, 2015 (doi: 10.1007/s12083-013-0204-8).

[25] J. Rosenberg, R. Mahy, P. Matthews, and D. Wing, "Session traversal utilities for NAT (STUN)", RFC 5389, IETF, Oct. 2008.

[26] R. Mahy, P. Matthews, and J. Rosenberg, "Traversal using relays around NAT (TURN): relay extensions to session traversal utilities for NAT (STUN)", RFC 5766, IETF, Apr. 2010.

[27] Fujitsu Cloud Technologies, Ltd., NIFCLOUD, homepage [Online]. Available: https://cloud.nifty.com

[28] NTT Communications Corp., Enterprise Cloud SkyWay, homepage [Online]. Available: https://webrtc.ecl.ntt.com



Yusuke Shimomura received his B.E. degree in Electronic Information Systems from the Shibaura Institute of Technology, Tokyo, Japan, in 2017. He is presently a master's course student at the Graduate School of Engineering and Science, Shibaura Institute of Technology, Tokyo, Japan. His research interests are in peer-to-peer networks and their applications.

E-mail: mf17037@shibaura-it.ac.jp
Graduate School of Engineering and Science
Shibaura Institute of Technology
Saitama, Japan

Olivier Fourmaux has been an Associate Professor at Sorbonne Université, France, since 2003. Before, he was an Assistant Professor at Institut Galilee, Université Paris 13, France. He received his Ph.D. degree in Computer Networking in 1998 and his M.Sc. degree in computer systems in 1995, both from UPMC. His research interests cover content delivery networks, P2P networks, active networks and multimedia in high-speed networks. He is a member of the Network and Performance group of the LIP6 Laboratory (CNRS-Sorbonne Université)

E-mail: olivier.fourmaux@sorbonne-universite.fr
Laboratoire d'Informatique de Paris 6
Sorbonne Université
Paris, France

Takumi Miyoshi – for biography, see this issue, p. 22.

Paper

LoCO: Local Cooperative Data Offloading System Based on Location Information

Taku Yamazaki, Kazuma Asano, Satoshi Arai, Yusuke Shimomura, and Takumi Miyoshi

Shibaura Institute of Technology, Saitama, Japan

https://doi.org/10.26636/jtit.2019.130518

Abstract—The development of high speed mobile networks and the widespread use of smartphones have enabled users to easily obtain large data volumes via the Internet. This causes a heavy consumption of network resources, a burden on the available bandwidth. To solve such problems, a data offloading method with a wireless LAN access point has been used to distribute traffic from mobile to fixed networks. However, the method using wireless LAN access points can only change the communication paths but cannot reduce the overall traffic. This paper proposes a local cooperative data offloading system (LoCO) that reduces the overall traffic by sharing data, with direct communication between neighbors based on their location-related information. Moreover, the authors implemented the LoCO system on Android smartphones and clarified its performance in comparison with a traditional client/server system through experiments to download data in a real-world environment.

Keywords—cooperative offloading, distributed download, load balance, load reduction, location information, peer-to-peer.

1. Introduction

The development of high speed mobile networks and the widespread use of smartphones have enabled users to easily obtain a large volume of data via the Internet, regardless of the user's location. However, Cisco's forecast [1] predicts that the overall traffic from mobile terminals in 2021 will increase by approximately 7 times from that of 2016. In particular, it is also predicted that, in 2021, video content traffic will account for approximately 78% of all mobile data traffic. The Ericsson mobility report [2] also predicts that the overall mobile data traffic will increase by approximately 8 times from that of 2017. In 2023, video content traffic will account for approximately 73% of all mobile data traffic. Thus, on the basis of the reports, to alleviate the effect of large volume data, such as a video content, the heavy use of mobile networks' resources must be dealt with.

For solving this problem, data offloading methods to distribute and reduce mobile traffic have been discussed [3]. As one of the major data offloading methods, a method with a wireless LAN access point [4]–[10], known as Hotspot [11], which changes the communication path and distributes data traffic from mobile networks to fixed networks via the wireless LAN access point, referred to as vertical handover [12], has been widely used. As a different approach from the above, cooperative data offloading, downloading, and sharing methods in which terminals download data to cooperate with their neighbors [13]–[18] have been proposed.

Although the data offloading method with a wireless LAN access point can only distribute data traffic to move it from mobile networks to fixed networks, the overall data traffic is the same. Hence, the method is not able to reduce the data traffic generally. The cooperative offloading and downloading methods can reduce the data traffic. However, the methods require modifications to access point and terminal firmware. In addition, the discovery process of cooperation users has not been researched and discussed in many cases. The cooperative download method with terminals passed by opportunistically determines the range of data for a partial download based on mobility prediction. However, its sharing efficiency may decrease under actual environments, since the method strongly depends on the accuracy of mobility prediction.

This paper proposes a local cooperative data offloading system (LoCO) which is able to download cooperatively and shares data among neighbors via direct communications based on location information, relying on peer-to-peer communications. In addition, LoCO is likely to be able to configure and select a cooperative download method. This paper also proposes cooperative download methods referred to as leader election-based cooperative download (LCD), distributed cooperative download (DCD) and enhanced DCD (eDCD). Moreover, this paper implements the LoCO system on smartphones that run it at the application-level, which means that it does not require any firmware-level modifications.

2. Related Work

Currently, a data offloading method with a wireless LAN access point is widely used to distribute data traffic from mobile networks to fixed networks [4], [8]–[10]. The deployment methods of wireless LAN access points have also been discussed in [5]–[7] for improving the data offloading effect. In the access point-based method, users who connect to the Internet via mobile networks change their own connection to wireless LAN, such as public wireless LAN services. Hence, the method may change the communication path from mobile to other networks. However, since this method only focuses on the data traffic distribution between mobile networks and fixed networks, it is not able to reduce data traffic. It is only applied to specified areas, because it only enables an area where there is a wireless LAN access point. Hence, the distribution effect of traffic strongly depends on the location.

The cooperative download method, where users download partial data and disseminate it to each other by passing it opportunistically based on a mobility prediction, has been proposed in [17], [18]. Here, each terminal sends a control message to a server. The message includes its own current position, destination, departure time and information about own partial data. Upon receiving the control message, the server generates a reply message which includes the ID of the terminal which is capable of encountering the sender of the control message, the probability of encountering the sender of the control message, the time of encountering and information about own data. Then, the server sends the message to each terminal. Based on the information received, each terminal predicts positions of encountering the other terminals which cooperate with each other and share partial data. Then, each terminal prioritizes to share the partial data which is difficult to obtain based on a mobility prediction. Although the method realizes cooperative data download based on a mobility prediction, the prediction of user mobility is difficult under actual environments. Hence, the sharing efficiency of the method may decrease since the method strongly depends on the accuracy of mobility prediction.

3. Local Cooperative Data Offloading System Based on Location Information

This paper proposes a local cooperative data offloading (LoCO) system that reduces data traffic to cooperatively download and share data among local terminals called neighbors, based on their location information.

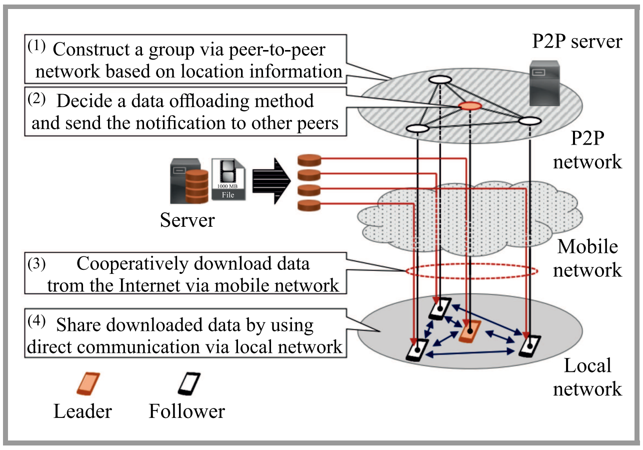

(1) Construct a group via peer-to-peer network based on location information
(2) Decide a data offloading method and send the notification to other peers
P2P server
P2P network
1000 MB File
Server
Mobile network
(3) Cooperatively download data trom the Internet via mobile network
(4) Share downloaded data by using direct communication via local network
Local network
Leader
Follower

Fig. 1. Overview of the LoCO system.

Figure 1 shows the structure of the LoCO system. First, the LoCO constructs a local group among neighbors that request the same data via a peer-to-peer network on mobile networks. Then, group members exchange control messages with each other, which include the download method and the data sharing type, to decide the offloading algorithm based on the messages. This paper also proposes several cooperative download methods and data sharing types. The LoCO system supports uniform resource locator (URL) etc. to specify the data, since it is relied upon to perform the range request function [19] which is a kind of conditional request [20] within HTTP/1.1 [21], [22] to enhance its flexibility and versatility. The function is also usable in the future because it is compatible with HTTP/2.0 [23]. After finishing downloading data, they share the data among other group members via a local communication medium, such as wireless LAN, Bluetooth, etc. After that, the terminals leave the group and the procedure ends.

3.1. Local Group Structure

This section defines the structure of a local group in the LoCO system, which consists of a single leader that manages the group and single or multiple followers. It is constructed to download the same data cooperatively. The local group is arranged in a circle centered around the leader, has a pre-determined radius and requires that the leader communicate with others via direct communication. All terminals exchange control messages except for data traffic via a peer-to-peer network on mobile networks. This paper does not focus mainly on the peer-to-peer system, since an existing peer-to-peer architecture [24], [25] specialized in sharing location information among peers, has already been proposed. The peer-to-peer architecture focuses on the construction of a local group based on location information of peers and supports the establishment of a connection among peers using session traversal utilities for NATs (STUN) [26] and traversal using relays around NAT (TURN) [27].

3.2. Local Group Construction

This section introduces the procedure of local group construction. Table 1 shows an example of peer information stored in the P2P server.

Table 1
Example of peer information in P2P server

Peer ID	Position	Content	Terminal ID	Requesting
P_1	(X_1, Y_1)	Data1	M_1	True
P_2	(X_2, Y_2)	Data2	M_2	True
P_3	(X_3, Y_3)	Data1	M_3	False
P_4	(X_4, Y_4)	Data3	M_4	True

First, when a terminal wants to obtain some data, it it searches its neighbors that also request the same data via the peer-to-peer network based on location information. Then, if there is no terminal that requested the same data, terminal i promotes a leader and registers own peer ID P_i, own location (X_i,Y_i), request data name or type, terminal ID M_i, and a Boolean value that denotes either requesting or not in the peer-to-peer control server. Peer ID P_i, which is generated at the beginning of the process of searching for neighbors on the peer-to-peer network, is used for designating the peer on the peer-to-peer network. The requested data name denotes the uniform resource identifier (URI) [28] of the requested data and so on. The terminal ID M_i, is generated when the communication among neighbors starts via the local network, and is used for designating the direct communication terminal.

If a group already exists, the terminal connects to its leader and establishes a peer-to-peer connection and then it becomes a group follower. The leader and the followers periodically update the registered information in the peer-to-peer control server in order to eliminate obsolete information.

Therefore, the LoCO system searches terminals that request the same data and enables cooperation between the group members to use the local group via a peer-to-peer network.

3.3. Local Cooperative Data Offloading Method

The LoCO system cooperatively obtains and shares data among terminals in a group. The procedure of the local cooperative offloading method among neighbors is described below.

1. Initializing process of cooperative data offloading. First, after starting the group construction of a group for content j, the leader waits a certain time T_{adv}. After that, the leader stops to request data and fixes group members for downloading cooperatively. Then, the leader obtains the number of terminals within the local group N and assigns a group member ID k $(0 \leq k \leq N-1)$ to each group member. Note that the group member ID k of 0 is assigned to the leader, and a group member ID k of 1 to $N-1$.

2. Cooperative download process via mobile network. After the initialization process, the leader $(k=0)$ sends control messages with containing the group member ID k of the receiver and the total number of members in the group N to all followers $(k = 1, 2, \ldots, N-1)$. The control message also includes the method of cooperative data download and the type of data sharing which will be used for sharing content j in the group. Upon receiving the control message, each follower k decides the offloading procedure and starts to download data based on the method and the sequence in accordance with the information of the received message. After they finish downloading data, each follower sends a notification of completing data download to the leader. When the leader receives the notification from all followers, the leader initiates the establishment of a connection via direct communication with the followers, for sharing downloaded data in the local network.

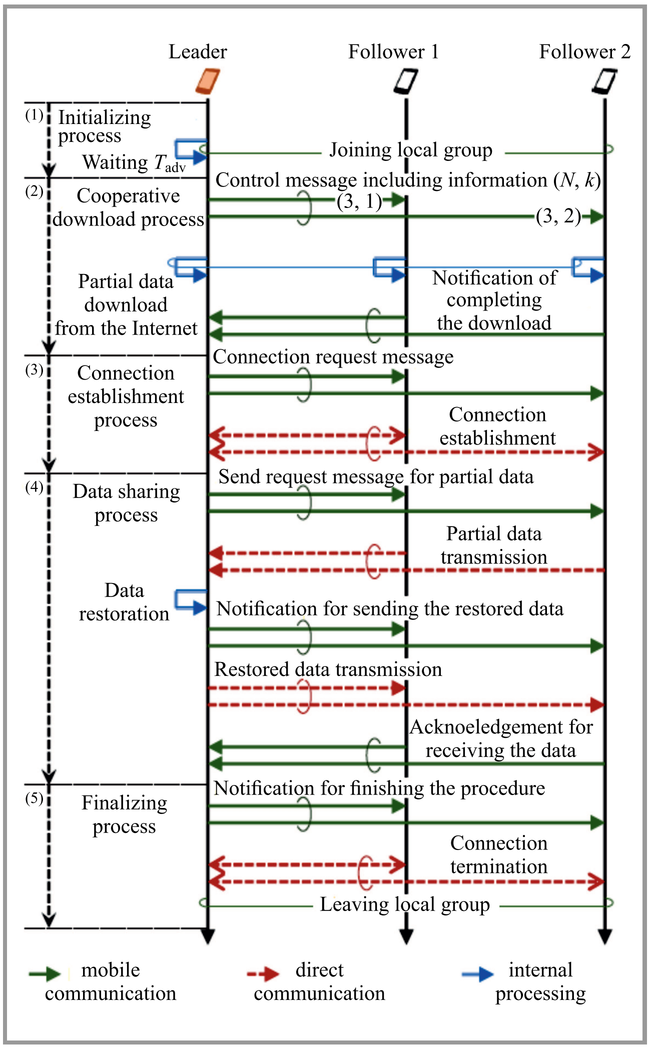

Leader
Follower 1
Follower 2
(1)
Initializing process
Waiting T_{adv}
Joining local group
Control message including information (N, k)
(2)
Cooperative download process
(3, 1)
(3, 2)
Partial data download from the Internet
Notification of completing the download
(3)
Connection request message
Connection establishment process
Connection establishment
(4)
Send request message for partial data
Data sharing process
Partial data transmission
Data restoration
Notification for sending the restored data
Restored data transmission
Acknoeledgement for receiving the data
(5)
Notification for finishing the procedure
Finalizing process
Connection termination
Leaving local group
mobile communication
direct communication
internal processing

Fig. 2. Operation sequence of the LoCO system.

3. Connection establishment process via direct communication. After completing the above process, the leader sends a connection request message to establish a local connection with all followers in the local network. Upon receiving the request message, each follower establishes the local connection using information of the leader obtained from the peer-to-peer control server in advance. The LoCO system does not depend on a specific medium. Currently, it applies mboxWi-Fi Direct [29]–[32] or Bluetooth [33] as the local direct communication medium. Therefore, the LoCO system also needs to decide the medium for local communication during the above process.

4. Data sharing process in local network. After establishment of the connection in the local network, the group initiates the sharing of the downloaded data based on the predefined sequence which is decided based on information received in advance. When a follower obtains the complete data, it sends an acknowledgement to the leader, indicating completion of data reception.

5. Finalizing process of cooperative data offloading. When the leader receives acknowledgements from all followers, the leader sends a notification informing them of the completion of cooperative data offloading. Upon receiving the notification, each follower terminates the local connection and leaves the local group.

3.4. Local Cooperative Download Method

This section proposes local cooperative download methods of the LoCO system which are (1) leader election-based cooperative download, (2) distributed cooperative download (DCD) and (3) enhanced DCD (eDCD) to be applied for downloading data from the Internet.

1. Leader election-based cooperative download. Leader election-based cooperative download (LCD) is a method where the leader downloads all data and, then, sends the downloaded data to all followers via a local network, using direct communication. Figure 3 shows an example behavior of LCD. Note that it is not necessary for the leader to be the same as the leader of the group, although the leader is defined, for convenience purposes, as a terminal responsible for the download of overall data.

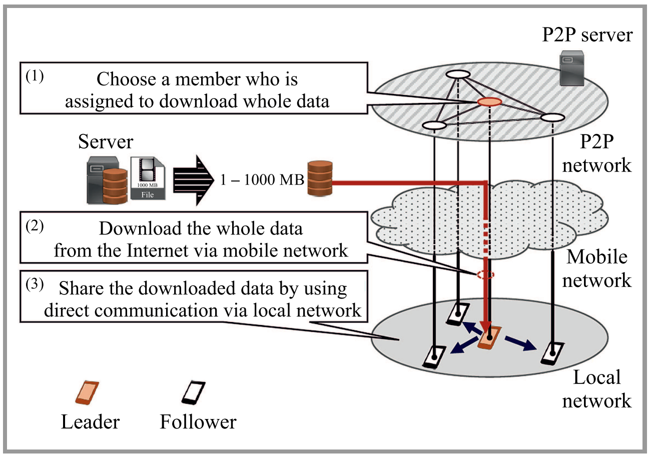

P2P server
(1) Choose a member who is assigned to download whole data
Server
P2P network
1000 MB File
1 – 1000 MB
(2) Download the whole data from the Internet via mobile network
Mobile network
(3) Share the downloaded data by using direct communication via local network
Local network
Leader
Follower

Fig. 3. Leader election-based cooperative download (LCD).

First, the terminals which request the same data construct a local group based on the construction process. Then, the group elects a responsible terminal as a leader and the leader downloads overall data from the Internet. After that, the leader sends and shares the overall data to all followers via a local network.

LCD has a simple procedure to share the data, since the leader only has to send the data to each follower. However, an unfairness in network resource consumption may occur, due to the characteristics of the leader election process. Therefore, this method is appropriate for a situation in which some content will be exchanged while changing the leader. In addition, another situation is that the leader is unrestricted or it has sufficient resources in terms of network resource consumption, such as fixed devices (e.g. digital signage, vending machine and so on) or devices connected to unlimited wireless LAN, etc.

2. Distributed cooperative download. Distributed cooperative download (DCD) is a method where all group members download partial data and, then, transfer it to others. Figure 4 shows an example behavior of DCD.

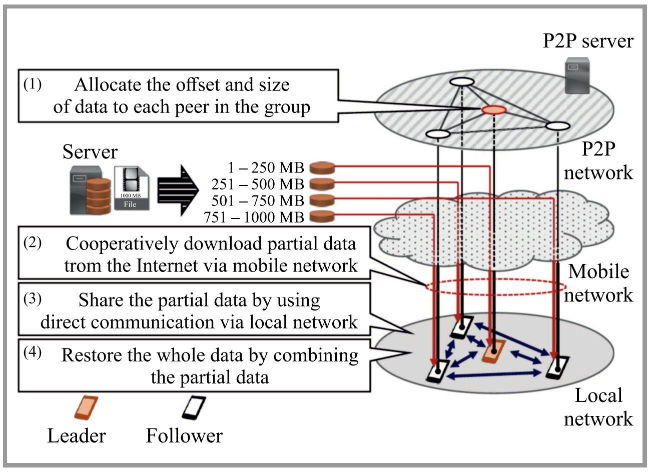

P2P server
(1) Allocate the offset and size of data to each peer in the group
Server
P2P network
1000 MB File
1 – 250 MB
251 – 500 MB
501 – 750 MB
751 – 1000 MB
(2) Cooperatively download partial data trom the Internet via mobile network
Mobile network
(3) Share the partial data by using direct communication via local network
(4) Restore the whole data by combining the partial data
Local network
Leader
Follower

Fig. 4. Distributed cooperative download (DCD).

First, the terminals which request the same data construct a local group based on the same construction process as LCD. Then, the leader assigns the range and size of the data to followers. Here, the size and range $[s_{j,k}^{(\text{head})}, s_{j,k}^{(\text{tail})}]$ are calculated based on data size s_j of a content j, the number of terminals in group N and their group member ID k from:

$$s_{j,k}^{(\text{head})} = \left\lfloor \frac{s_j \times k}{N} \right\rfloor + 1 \,, \tag{1}$$

$$s_{j,k}^{(\text{tail})} = \left\lfloor \frac{s_j \times (k+1)}{N} \right\rfloor \,. \tag{2}$$

As the above calculation shows, the leader and followers download partial data. Then, all followers send a notification to the leader and, next, all terminals share partial data via the local network. Last, they restore the whole data by combining the received partial data.

LCD achieves fairness in terms of network resource consumption in both mobile and local networks. However, due to its characteristics, DCD requires each terminal to connect with all others, like in the full mesh topology, because they need to send partial data between each other. Hence, construction of a local group may become complex, because DCD needs to establish and keep many connections. The switching and the keeping of many connections may cause higher delays.

3. Enhanced DCD. To ease the requirement, this paper also proposes an enhanced variant of DCD (eDCD) which operates differently from the normal DCD in order to reduce the complexity of the local network. eDCD mainly focuses on traffic reduction via the Internet without constructing a full mesh topology. Figure 5 shows an example behavior of eDCD.

In this variant, first, after group construction, each terminal download partial data by using the same method as the normal DCD. After downloading the partial data of all the members, all followers send their partial data to the

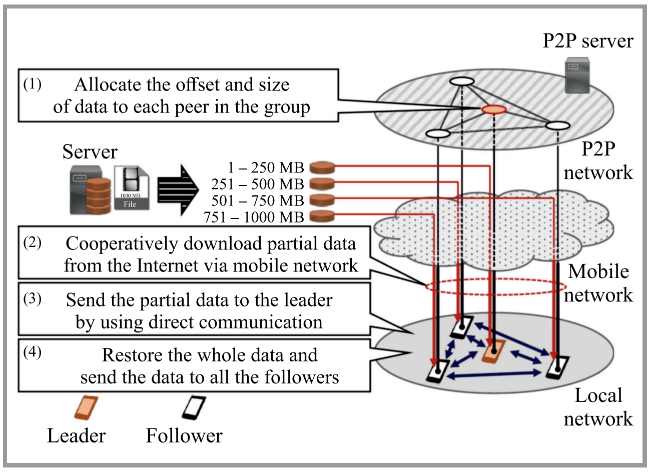

P2P server
(1) Allocate the offset and size of data to each peer in the group
Server
1000 MB File
1 – 250 MB
251 – 500 MB
501 – 750 MB
751 – 1000 MB
P2P network
(2) Cooperatively download partial data from the Internet via mobile network
(3) Send the partial data to the leader by using direct communication
(4) Restore the whole data and send the data to all the followers
Mobile network
Local network
Leader
Follower

Fig. 5. Enhanced distributed cooperative download (eDCD).

leader. After that, the leader restores the overall data by combining them. Finally, the leader sends the whole data to each follower. Therefore, in eDCD, the leader collects all partial data from all followers. As a result, each follower only establishes a connection with the leader based on a tree topology which includes the leader as the root. Namely, the eDCD reduces the number of connections without compromising the advantage, since each follower does not need to establish a connection with other followers.

3.5. Data Sharing in Local Network

This section proposes 3 ways of data sharing with the use of the LoCO system, to execute a local cooperative download. Here, we define that the sender of direct communication is called master, and the receiver of direct communication is called slave. Note that LCD executes the below sequence only once, since it is only necessary to connect the master with the slave. In contrast with LCD, DCD requires execution of the below sequence more than once while changing masters. In DCD, a master does not reconnect with a slave that has already finished the exchange of partial data. The 3 data sharing methods proposed are described below.

1. Data sharing type A. The master establishes a connection with all slaves through direct communication. After the connection is established, the master exchanges data with all slaves simultaneously. After that, they abolish the connection.

2. Data sharing type B. The master establishes a connection with all slaves through direct communication. After the connection is established, the master exchanges data with all slaves, on a one-by-one basis. After that, they abolish the connection.

3. Data sharing type C. The master establishes a connection with a slave through direct communication. After the connection is established, the master exchanges slave. After that, they abolish the connection. The master repeats the above for the unconnected terminals until they disappear.

4. Performance Evaluation

In this section, we evaluate the LoCO system based on real world experiments, by measuring the overall performance as well as comparison between cooperative download methods and data sharing types. In the experiments, we implemented the LoCO system as an application on Android OS [34] on 5 smartphones.

4.1. Overall Performance Evaluation Setup

This experiment evaluated the overall performance of the LoCO system to clarify its effectiveness. This experiment used NTT East FLET'S Hikari Next [35] as a backbone network and NTT Plala [36] as an Internet service provider, offering the downlink and uplink speed of 100 Mbps. In addition, a wireless LAN access point with IEEE 802.11g is used instead of the mobile networks to simplify the experiment. Wi-Fi Direct [29]–[32] was selected as the local direct communication method. The smartphones were fixed on a desk in a row, at 5 cm intervals. We chose the eDCD cooperative download method and the data sharing type B.

First, we boot the application on a single smartphone and the smartphone waits the requesting time T_{adv} which is set to 60 s. After that, we boot the application on the other smartphones, one-by-one, at 5 s intervals. Next, each terminal cooperatively downloads 10 MB of data via the LoCO system. Note that the data request and download via the Internet was performed based on the range request [20], which is a kind of a conditional request [19] in HTTP/1.1 [21], [22]. We also focused the evaluation on the effect on mobile and fixed networks, and thus we excluded the direct communication traffic from the result describing the total amount of traffic. We compared the LoCO system with the traditional client/server (C/S) model.

4.2. Results of Overall Performance Evaluation

Figure 6 shows total sent and received data of the traditional C/S model and the proposed system. The C/S model increases the total amount of data received in proportion to the number of terminals, since all of them need to download all data independently. In contrast to the C/S model, the total amount of data received in the case of the LoCO system is not changed, even if the number of terminals is increased. This is because the members of the LoCO group may cooperate with others, and thus data are divided into partial data based on the number of group members. Hence, they do not need to download all data. However, since LoCO requires cooperation among group members, the total amount of traffic sent is increased as the number of terminals increases, in comparison with the C/S model. Figure 7 shows the amount of sent and received data and control messages of each role on the LoCO system. Note that the result of the C/S model indicates the result of a single terminal. Each LoCO terminal reduces the received traffic in comparison to the terminal of the C/S model. As

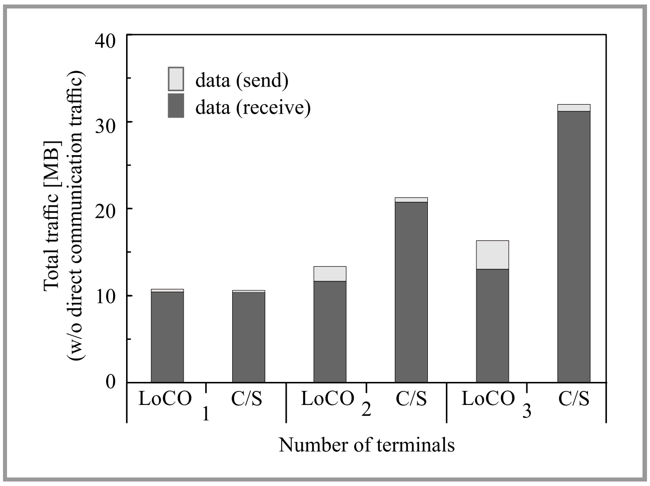


Fig. 6. Total traffic with varying number of terminals.

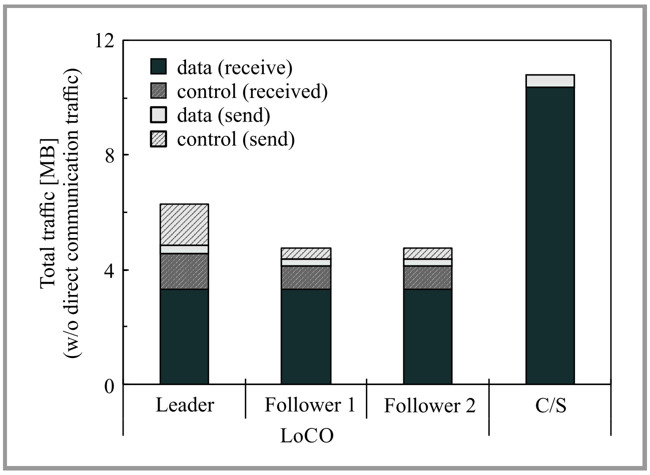


Fig. 7. Total traffic of each role in the LoCO system.

mentioned above, LoCO can cooperatively download data among the group members, and thus the amount of received data of each terminal decreases since they only download partial data. In contrast with the amount of received data, the amount of sent data of increases in LoCO in comparison to the C/S model. In particular, it can be seen that the leader has the largest amount of sent data among all group members. This is because the leader sends control messages to all followers and needs to update group information on the peer-to-peer server. In addition, LoCO imposes control traffic on each member since it requires exchange control messages via the P2P network.

4.3. Setup for Comparing Data Sharing Types and Cooperative Download Methods

The experiment evaluated the time required to complete data sharing between all members, to compare the differences between data sharing types and cooperative download methods as well. Here, we chose Bluetooth as the direct communication method [33], based on Nearby Connections API 2.0 [37]. This experiment analyzes traffic flows and enables to trace them using Wireshark [38] as a packet capturing tool.

3 to 5 terminals were used to send/receive 5 MB of data to each other, and the time required to complete the data sharing phase between all members, from the beginning to the end, was measured. Note that this experiment used throughput in a free flow scenario, as well as throughput in a congested scenario, which is the throughput pre-measured in the mobile network to eliminate the effect of deploying a fraction of the fixed network. A medium-rate congestion level was assumed as well. The throughput was measured seven times at Tokyo Big Sight, where a large-scale event was held, from 9:22 to 9:40 on November 31st 2017. Consequently, the throughput of the mobile network was determined to offer the uplink speed of 7.48 Mbps and the downlink speed of 0.8 Mbps in the congested scenario, to compare with the uplink speed and downlink speed of 11 Mbps for the normal scenario. Note that, in comparison with the cooperative download methods, data sharing type C was used, since it offers better performance than others, based on the experimental results. This experiment varies the size of data from 7.5 MB to 45 MB, with a varying number of terminals.

4.4. Results of Comparisons between Data Sharing Types and Cooperative Download Methods

Figure 8 shows the time required to complete the data sharing phase for 3 types of methods used. The results show that the difference in the time required to complete data sharing is small between type B and type C, though type A needs a longer time to complete the procedure in comparison to the remaining types. In particular, with 5 terminals, the increase observed in the case of data sharing type A is larger than in other scenarios. We investigated the reason by tracing data with the use of a packet capturing tool. Consequently, Bluetooth communication established via nearby connections is relied upon to send data using time-division multiple access (TDMA) scheduling from the master to the slave, when the terminals send data simultaneously. Therefore, if the number of terminals is increased, the delay

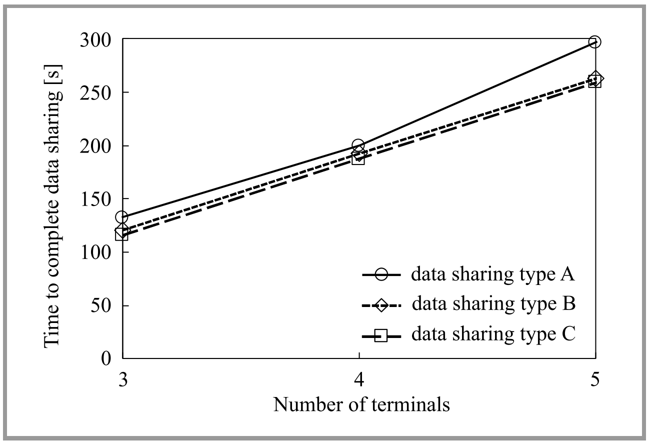


Fig. 8. Comparison between the proposed data sharing types of the LoCO system.

resulting from switching and waiting is much longer, since the switching of the slaves is more intense. In contrast with data sharing type A, the difference between type B and type C is small. Hence, they can avoid the above problem by sending data one-by-one, using their own scheduling algorithm even if some connections are established by the terminals.

Figures 9 and 10 show the time required to complete data sharing in LCD and DCD, in the normal scenario, without congestion, and in a scenario with congestion.

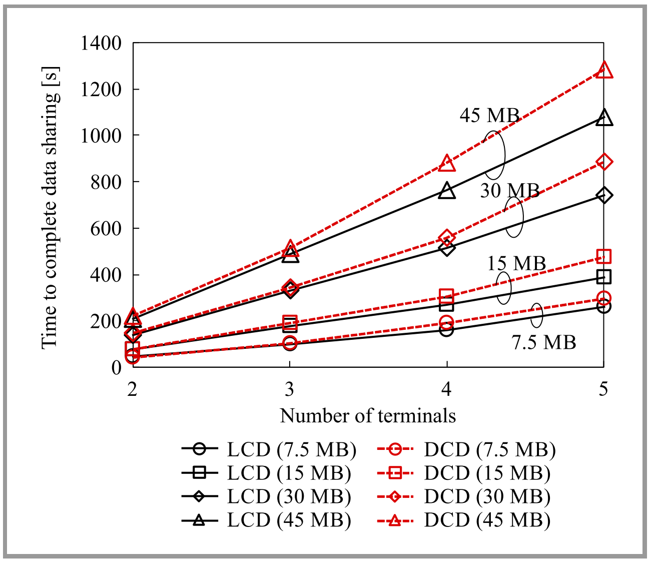


Fig. 9. Comparison between local cooperative download methods of the LoCO system in a normal scenario without congestion (uplink speed and downlink speed are set to 11 Mbps).

Figure 9 shows the time required to complete data sharing in LCD and DCD in a normal environment, where there is no congestion. The result shows that LCD has a shorter time required to complete data sharing than DCD, at least in most cases. In particular, when the data size becomes larger, the difference between LCD and DCD also becomes larger. This is because the scenario involved is a normal environment in which uplink and downlink speeds are set to 11 Mbps. In this case, the time for downloading data from the Internet is decreased in comparison to the congested scenario. As a result, the effect of the distributed download of DCD becomes relatively less evident. Therefore, LCD has the shorter time required to complete data sharing than DCD, since LCD has a simpler data sharing procedure, as it only sends all data to all followers in the local network. In addition, the result also shows that the increase in the time required to complete data sharing is larger in DCD than in LCD when the number of terminals increases. As mentioned above, this is because LCD only sends all data from the master to all slaves, in contrast to DCD. The count of connection switching tasks is $N-1$ when the number of group members is N, and thus the count of connection switching tasks increases according to the order of $\mathcal{O}(n)$. On the other hand, in DCD, all group members need to switch the connection to each other. In other words, the count of connection switching tasks is ${}_NC_2$. Hence, the difference between LCD and DCD becomes larger when the number of terminals increases, because the delays caused by switching and waiting are imposed since DCD increases the count of connection switching tasks according to the order of $\mathcal{O}(n^2)$.

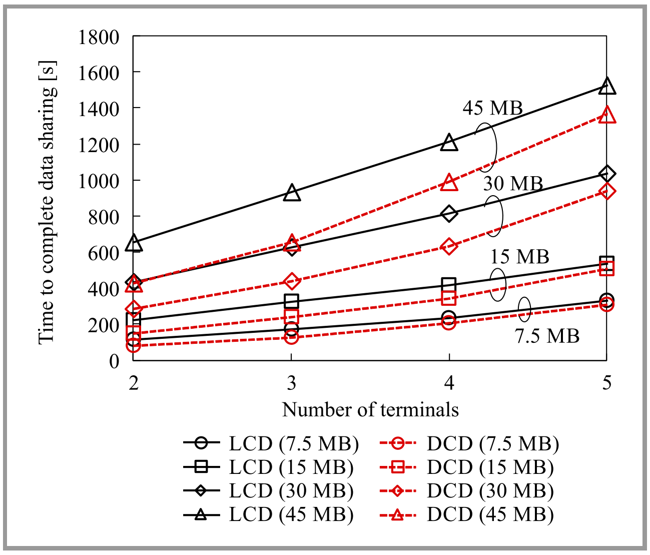


Fig. 10. Comparison between local cooperative download methods of the LoCO system in a congested scenario (uplink speed is set to 7.48 Mbps and downlink speed is set to 0.8 Mbps).

Figure 10 shows the time required to complete data sharing in LCD and DCD, in a congested scenario. The results show that DCD has a shorter time to complete data sharing than LCD, in contrast to the result of Fig. 9. This is because the size of data downloaded by each terminal from the Internet is decreased in DCD, since it downloads partial data, whereas the leader downloads all data in LCD. In DCD, each terminal may download partial data in parallel with other terminals, and thus the necessary download time is decreased in comparison to LCD. Especially, in this scenario, the effect of parallel downloads becomes much larger than in the previous case, since the backbone network is slow due to network congestion. In addition, when the size of data shared becomes larger, the time required to complete data sharing is shorter in DCD than in LCD. However, regardless of data size, although the time in LCD linearly increases as the number of terminals increases, DCD shows a higher increase ratio than LCD. Consequently, the difference between the time required to complete data sharing, observed between LCD and DCD, becomes small when the number of terminals is increased due to the characteristics of DCD mentioned above.

5. Conclusion

This paper proposed a local cooperative data offloading (LoCO) system based on location information, which realizes cooperative download and data sharing among neighbors by using peer-to-peer communication and local communication. In addition, this paper also proposed two major types of cooperative download methods and three types of data sharing sequences. Furthermore, we implemented the LoCO system on Android smartphones as an application without firmware-level modifications. In order to enable the system to be used on smartphones, experiments were conducted in real world environments to evaluate the overall performance of LoCO and to establish the differences between the cooperative download methods and individual data sharing types. It has been shown that LoCO reduces data traffic by using local cooperation between neighbors when they download the same data.

In the future, in order to precisely assess the scalability of LoCO, its performance should be evaluated in a large-scale experiment, because performance of the cooperative data offloading process is affected by the number of group terminals.

The cooperative download methods remain primitive for the time being, as this paper only confirms their fundamental characteristics in real world environments. However, real world environments, the LoCO system should take into consideration various conditions, such as communication quality degradation caused by radio interference, terminal mobility and so on. Therefore, cooperative download methods taking the above conditions into consideration should be studied as future work.

Acknowledgements

This work was supported by TAF Research Grant and JSPS KAKENHI Grant Numbers 17K12680 and 17K06441.

References

[1] "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 201–2021" [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.pdf

[2] "The Ericsson Mobility Report" [Online]. Available: https://www.ericsson.com/en/mobility-report

[3] F. Rebecchi *et al.*, "Data offloading techniques in cellular networks: A survey," *IEEE Commun. Surveys & Tutor.*, vol. 17, no. 2, pp. 580–603, 2014 (doi: 10.1109/COMST.2014.2369742).

[4] "docomo Wi-Fi" [Online]. Available: https://www.nttdocomo.co.jp/service/wifi/docomo_wifi [in Japanese]

[5] X. Kang, Y.-K. Chia, and S. Sun, "Mobile data offloading through a third-party WiFi access point: An operator's perspective", in *Proc. IEEE Globecom Worksh. GC Wkshps 2013*, Atlanta, GA, USA, 2013, pp. 696–701 (doi: 10.1109/GLOCOMW.2013.6825069).

[6] E. Bulut and B. K. Szymanski, "WiFi access point deployment for efficient mobile data offloading", in *Proc. 1st ACM Int. Worksh. on Pract. Issues and Appl. in Next Gener Wirel. Netw. PINGEN 2012*, Istanbul, Turkey, 2012, pp. 45–50 (doi: 10.1145/2348714.2348723).

[7] E. Bulut and B. K. Szymanski, "WiFi access point deployment for efficient mobile data offloading", *ACM SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 17, no. 1, pp. 71–78, 2013 (doi: 10.1145/2502935.2502948).

[8] 3GPP, "Architecture enhancements for non-3GPP accesses", 3GPP TS 23.402, March 2018 [Online]. Available: https://www.3gpp.org/DynaReport/23402.htm

[9] 3GPP, "Access to the 3GPP evolved packet core (EPC) via non-3GPP access networks", 3GPP TS 24.302, March 2018 [Online]. Available: https://www.3gpp.org/dynareport/24302.htm

[10] 3GPP, "Access network discovery and selection function (ANDSF) management object (MO)", 3GPP TS 24.312, June 2018 [Online]. Available: https://www.3gpp.org/dynareport/24312.htm

[11] "Hotspot 2.0 (release 2) technical specification", Wi-Fi Alliance, Dec. 2016.

[12] J. Márquez-Barja, C. T. Calafate, J.-C. Cano, and P. Manzoni, "Review: An overview of vertical handover techniques: Algorithms, protocols and tools", *Comp. Commun.*, vol. 34, no. 8, pp. 985–997, 2011 (doi: 10.1016/j.comcom.2010.11.010).

[13] U. Lee *et al.*, "P2P content distribution to mobile Bluetooth users", *IEEE Trans. on Veh. Technol.*, vol. 59, no. 1, pp. 356–367, 2010 (doi: 10.1109/TVT.2009.2030893).

[14] S.-S. Kang and M. W. Mutka, "A mobile peer-to-peer approach for multimedia content sharing using 3G/WLAN dual mode channels", *J. Wireless Commun. Mobile Comput.*, vol. 5, no. 6, pp. 633–645, 2005 (doi: 10.1002/wcm.332).

[15] S. Sharafeddine, K. Jahed, N. Abbas, E. Yaacoub, and Z. Dawy, "Exploiting multiple wireless interfaces in smartphones for traffic offloading", in *Proc. 1st Int. Black Sea Conf. on Commun. and Netw. BlackSeaCom 2013*, Batumi, Georgia, 2013, pp. 142–146 (doi: 10.1109/BlackSeaCom.2013.6623398).

[16] A. Le *et al.*, "MicroCast: Cooperative video streaming using cellular and local connections", *IEEE/ACM Trans. on Netw.*, vol. 24, no. 5, pp. 2983–2999, 2016 (doi: 10.1109/TNET.2015.2501349).

[17] H. Hanano, Y. Murata, N. Shibata, K. Yasumoto, and M. Ito, "A cooperative download method for low cost video ads dissemination through WiFi-cellular hybrid network", *IPSJ J.*, vol. 51, no. 2, pp. 440–452, 2010 [Online]. Available: https://core.ac.uk/download/pdf/75905394.pdf [in Japanese].

[18] Y. Takamatsu, W. Sun, K. Yasumoto, Y. Yamauchi, and M. Ito, "Energy-aware cooperative download method for mobile phones utilizing street pass communication", *IPSJ J.*, vol. 53, no. 2, pp. 783–794, 2012 [in Japanese].

[19] R. Fielding, Y. Lafon, and J. Reschke, Eds., "Hypertext transfer protocol (HTTP/1.1): Range requests", IETF RFC7233, June 2014 (doi: 10.17487/RFC7233).

[20] R. Fielding and J. Reschke, Eds., "Hypertext transfer protocol (HTTP/1.1): Conditional requests", IETF RFC7232, June 2014 (doi: 10.17487/RFC7232).

[21] R. Fielding and J. Reschke, Eds., "Hypertext transfer protocol (HTTP/1.1): Message syntax and routing", IETF RFC7230, June 2014 (doi: 10.17487/RFC7230).

[22] R. Fielding and J. Reschke, Eds., "Hypertext transfer protocol (HTTP/1.1): Semantics and content", IETF RFC7231, June 2014 (doi: 10.17487/RFC7231).

[23] M. Belshe, R. Peon, and M. Thomson, Eds., "Hypertext transfer protocol version 2 (HTTP/2)", IETF RFC7540, May 2015 (doi: 10.17487/RFC7540).

[24] Y. Shimomura and T. Miyoshi, "Location-based peer-to-peer communication system", *IEICE Tech. Rep.*, vol. 117, no. 114, pp. 77–80 2017 [in Japanese] [Online]. Available: https://www.ieice.org/ken/paper/201707078bvr/eng

[25] T. Miyoshi, Y. Shimomura, and O. Fourmaux, "G-LocOn: A P2P-based communication framework for geo-location oriented networks", in *IEICE Inform. and Commun. Technol. Forum ICTF 2018*, Graz, Austria, 2018.

[26] J. Rosenberg, R. Mahy, P Matthews, and D. Wing, "Session traversal utilities for NAT (STUN)", IETF RFC5389, Oct. 2008 (doi: 10.17487/RFC5389).

[27] R. Mahy, P. Matthews, and J. Rosenberg, "Traversal using relays around NAT (TURN): Relay extensions to session traversal utilities for NAT (STUN)", IETF RFC5766, April 2010 (doi: 10.17487/RFC5766).

[28] T. Berners-Lee, R. Fielding, and L. Masinter, "Uniform resource identifier (URI): Generic syntax", IETF RFC3986, Jan. 2005 (doi: 10.17487/RFC3986).

[29] "Wi-Fi Direct" [Online]. Available: http://www.wi-fi.org/discover-wi-fi/wi-fi-direct

[30] "Wi-Fi peer-to-peer (P2P) technical specification", Wi-Fi Alliance, July 2016.

[31] "Wi-Fi peer-to-peer Services (P2Ps) technical specification", Wi-Fi Alliance, July 2015.

[32] "UPnP file transfer service technical specification", Wi-Fi Alliance, June 2015.

[33] Bluetooth [Online]. Available: https://www.bluetooth.com

[34] Android [Online]. Available: https://www.android.com

[35] NTT East FLET'S Hikari [Online]. Available: https://flets.com/english

[36] NTT Plala Inc. [Online]. Available: http://www.nttplala.com/english

[37] Nearby Connections API [Online]. Available: https://developers.google.com/nearby/connections/overview

[38] Wireshark [Online]. Available: https://www.wireshark.org

Kazuma Asano received his B.E. degree in Electronic Information Systems from the Shibaura Institute of Technology, Tokyo, Japan, in 2018. He is presently a master's course student at the Graduate School of Systems and Information Engineering, University of Tsukuba, Ibaraki, Japan. His research interests include mobile data offloading and computer vision.

E-mail: bp14004@shibaura-it.ac.jp
College of Systems Engineering and Science
Shibaura Institute of Technology
Saitama, Japan

Satoshi Arai received his B.E. degree in Electronic Information Systems from the Shibaura Institute of Technology, Tokyo, Japan, in 2018. He is presently an engineer at Nippan Computer Technology Inc., Tokyo, Japan. His research interest is in mobile data offloading.

E-mail: bp14006@shibaura-it.ac.jp
College of Systems Engineering and Science
Shibaura Institute of Technology
Saitama, Japan

Taku Yamazaki – for biography, see this issue, p. 12.

Yusuke Shimomura – for biography, see this issue, p. 66.

Takumi Miyoshi – for biography, see this issue, p. 22.

Paper

Rectangular Dielectric Resonator Antenna with Single Band Rejection Characteristics

Mohamed Debab and Zoubir Mahdjoub

Laboratory of Electromagnetism, Photonics and Optronics (LEPO), Djillali liabes University of Sidi Bel Abbès, Sidi Bel Abbès, Algeria

https://doi.org/10.26636/jtit.2019.124718

Abstract—In this paper, a rectangular dielectric resonator antenna (DRA) suitable for wideband applications is presented and a band notch of WLAN (5.15–5.75) GHz is proposed. The DRA is mainly composed of a 20×20 mm rectangular dielectric resonator, coated with metal on the top surface, and a circular monopole excitation patch with an air gap insert. A coaxial line feed is used to excite the circular, planar monopole. An open-ended quarter wavelength C-shaped slot is embedded in the circular patch to create the notched band. The simulated results demonstrate that the proposed design produces an impedance bandwidth of more than 80%, ranging from 3.10 to 7.25 GHz for a reflection coefficient of less than -10 dB and with a band rejection at 5.50 GHz. Band notch characteristics, VSWR, and radiation patterns are studied using the HFSS high-frequency simulator and CST Studio software.

Keywords—band-stop function, C-shaped slot, dielectric resonator antenna (DRA), planar monopole.

1. Introduction

Dielectric resonator antennas (DRAs) are widely used due to their remarkable characteristics, such as different excitation mechanisms, small size and high permittivity. Other inherent advantages of DRAs include: low dissipation loss at high frequency, wide bandwidths and high radiation efficiency due to the absence of conductors and surface wave losses. Many investigations were focused on its bandwidth and input impedance [1]–[7]. Such parameters may easily be varied by changing the antenna's specifications, such as the dielectric constant of the resonator material, the dimensions and feed mechanisms. Special geometric configurations of DRAs may also enhance bandwidth, e.g. P-shapes, conical, cylindrical and others [8]–[10].

In the past few years, hybrid dielectric resonator antennas have received a great deal of attention due to the wideband operation that is possible without increasing antenna volume. For example, paper [11] introduced multi-segment DRAs to enhance wideband coupling between a microstrip line and a DRA, [12] proposed a hybrid-fed DRA with a stepped patch and an intermediate substrate to obtain bandwidth between 7.5 and 12.5 GHz. In [13], a DRA was designed with an added monopole patch so that the antenna can simultaneously act as a radiator and a loading element, to produce an ultra-wide bandwidth (UWB). UWB DRAs with band stop performance have been proposed in [14], [15] and they were also designed to minimize interference between the UWB and narrowband systems, such as WiMAX and WLAN. A coplanar-fed UWB DRA with dual band-notched characteristics (WiMAX and WLAN) was created by introducing two slots in the radiation patch [16]. The notched bands are mainly implemented by adding stubs around the radiator or a feed line and etching slots onto the patch. The lengths of the etched slots or additional stubs are about a quarter wavelength or half wavelength, corresponding to the designed notch, using U-shaped [17], C-shaped [18], π-shaped [19], Y-shaped [20] or L-shaped slots [21].

In this paper, a compact wideband DRA with single band-notched characteristics (WLAN band) is presented, which uses a rectangular dielectric resonator (DR), coated with metal on the top surface, and a circular monopole excitation patch together with an air gap inserting technique. The notched frequency is realized by etching a C-shaped slot of a quarter wavelength onto the radiation patch. The tuning of the notched center frequencies is done by changing the length of the slot. The proposed antenna achieves an impedance bandwidth of 3.10 to 7.25 GHz, with a return loss being lower than -10 dB, and presents a decrement gain at approximately 5.60 GHz. The design of the antenna was first simulated using the frequency domain An-soft high-frequency structure simulator (HFSS), and was then confirmed with the time domain CST Studio microwave simulator.

2. Antenna Design

The configuration of the proposed DRA is shown in Fig. 1. It has physical dimensions of 20×20 mm and is centrally placed above a finite ground plane with the size of 50×50 mm. The proposed DR is depicted by L_D, W_D, and $h - h_1$. The DR is designed using microwave dielectric

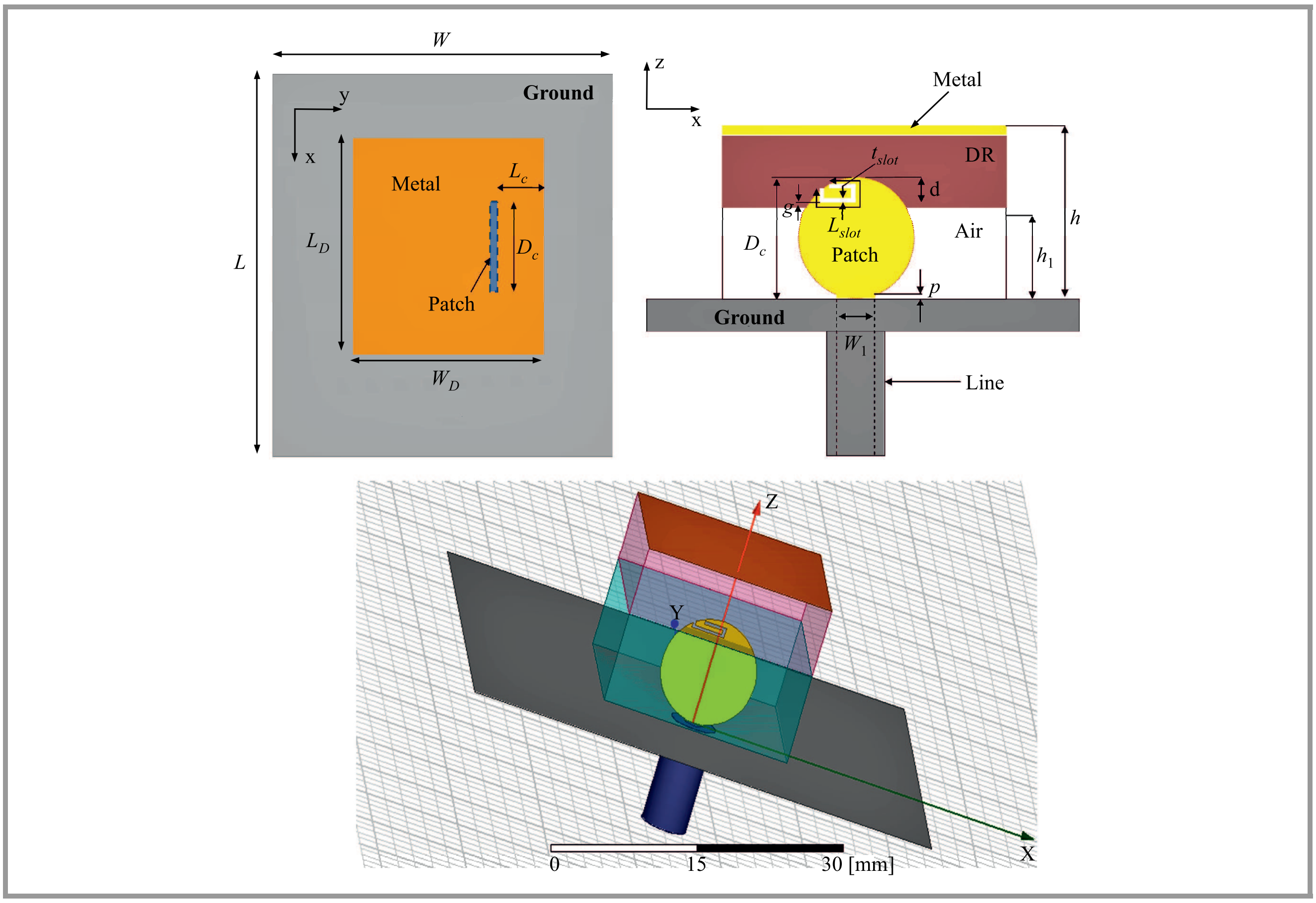


Fig. 1. Geometry of the proposed antenna.

Rogers (RO3006) material with a relative permittivity of $\varepsilon_{rD} = 6.15$ and dielectric loss tangent of 0.0025.

Table 1
Optimal parameters of the proposed antenna

Parameter	Value [mm]	Parameter	Value [mm]
W	50	L_c	3
L	50	d	1.5
L_D	20	L_{slot}	7.2
W_D	20	t_{slot}	0.3
h_1	8.5	p	0.12
h	15	W_1	2
ε_{rD}	6.15	g	0
D_c	10		

The circle patch antenna penetrates into the DR and is connected to a 50 Ω coaxial line. The thickness of the air gap inserted between the DR and the ground plane is denoted by h_1. The exciting patch has a top width of $W_1 = 2$ mm and the width of the gap between the patch and the ground plane is $p = 0.12$ mm. A C-shaped slot of width $t_s = 0.3$ mm is etched onto the patch. The optimized parameters of the antenna are listed in Table 1.

2.1. Basic Antenna Design without C-shaped Slot

First of all, the design approach is to simulate the proposed DRA without a C-shaped slot by varying some parameters; a parametric study is then performed to see the effect on the reflection coefficients. The HFSS software was used for the parametric analysis.

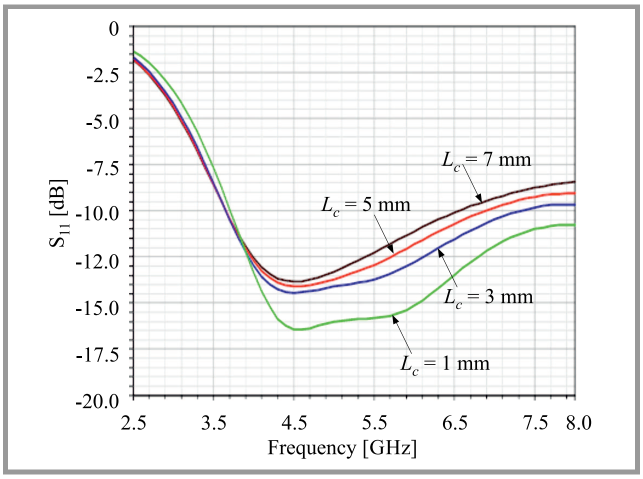


Fig. 2. Simulated S_{11} of the basic antenna (without metal coating) for different values of L_c. (For color pictures visit https://doi.org/10.26636/jtit.2019.124718)

Figure 2 shows the simulated S_{11} without metal coating, when the position of the patch L_c alters from 1 to 7 mm, with other parameters remaining fixed. It is clear that for S_{11} less than -10 dB, the lower edge frequency of the bandwidth is about 3.6 GHz and the height edge frequency increases. When $L_c = 3$ mm, the antenna offers a height edge frequency with the bandwidth of 7.25 GHz, and the broad impedance bandwidth of 67% for S_{11} less than -10 dB, giving the 3.60 to 7.25 GHz frequency band. The air gap between the DR and the ground plane (with the thickness h_1) plays an important role in the bandwidth enhancement. Figure 3 describes the effects of different values of h_1. It may be seen that by introducing an air gap, the lower edge frequency decreases at 3.6 GHz when $h_1 = 8.5$ mm.

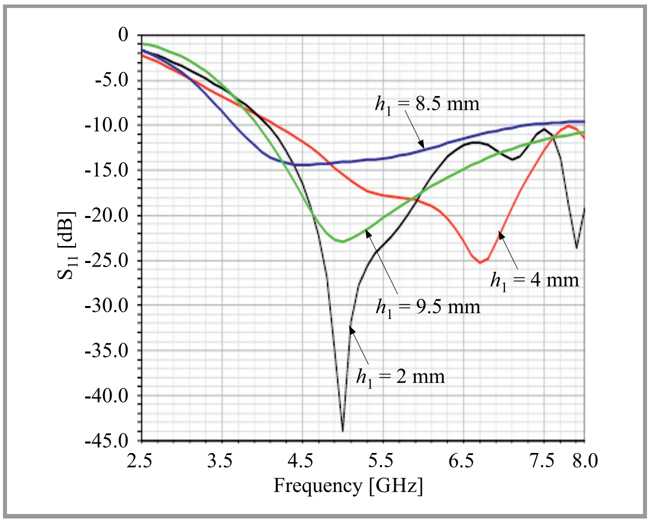


Fig. 3. Simulated S_{11} of the basic antenna (without metal coating) for different values of h_1.

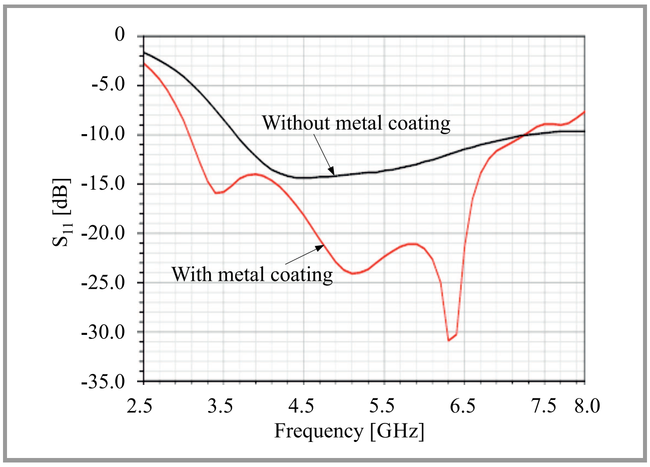


Fig. 4. Effect of the metal coating on the impedance matching characteristic.

Figure 4 illustrates the proposed antenna with and without metal coating. When the structure is not coated, the antenna works in the range of 3.60 to 7.25 GHz, with a 67% impedance bandwidth (for reflection coefficients S_{11} lower than -10 dB). When it is coated, the lower band shifts to 3.10 GHz and the antenna has a sharp resonance dip of S_{11} -31 dB at 6.30 GHz with an 80% impedance bandwidth, for S_{11} lower than -10 dB, which is the highest when compared to the antenna without metal coating. The permittivity of the dielectric is much higher than that of the air. The dielectric-air interface can be approximated as a perfect magnetic conductor (PMC) boundary. The metallic foil on the dielectric resonator is treated as a perfect electrical conductor (PEC). Hence, the structure forms a cavity with PMC and PEC on different portions of the DR, filled with a high-permittivity dielectric.

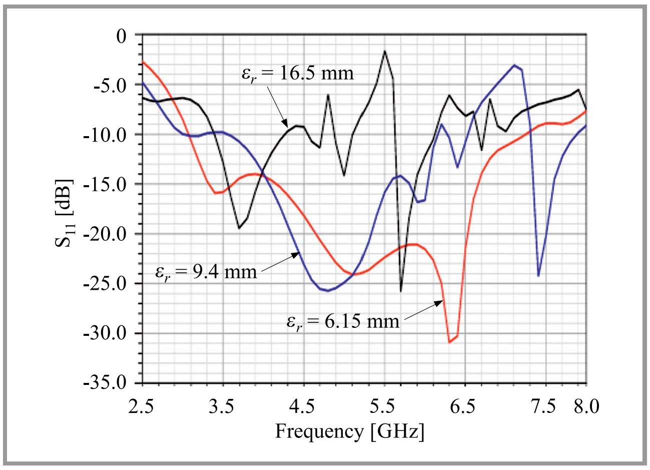


Fig. 5. Simulated reflection coefficient with different values of ε_{rD} with a metal coating.

It is well known that as the dielectric constant is increased, the wavelength in the DR is decreased, which results in a lower resonant frequency. Figure 5 shows the effect that DR permittivity ε_{rD} exerts on resonant frequencies. Increasing the permittivity leads to an increase of the Q factor, thus reducing the bandwidth of the resonant modes. Note that the resonant frequency is greatly affected by the dielectric constant. Therefore, permittivity of $\varepsilon_{rD} = 6.15$ is used to design the proposed DRA.

2.2. C-shaped Slot Analysis

The central frequency of the notch band function was designed to adjust the length of the slot. The length of the slot is about a quarter of the wavelength corresponding to the resonant frequency:

$$L_{slot} \approx \frac{\lambda_g}{4} = \frac{\lambda_0}{4\sqrt{\varepsilon_{eff}}} = \frac{c}{4 f_{notch}\sqrt{\varepsilon_{eff}}}\,, \tag{1}$$

$$\varepsilon_{eff} = \frac{\varepsilon_r + 1}{2}\,, \tag{2}$$

where λ_0 is the free space wavelength, f_{notch} is the central frequency of the notch band and c and ε_{eff} are the speed of light and the approximated effective dielectric constant, respectively.

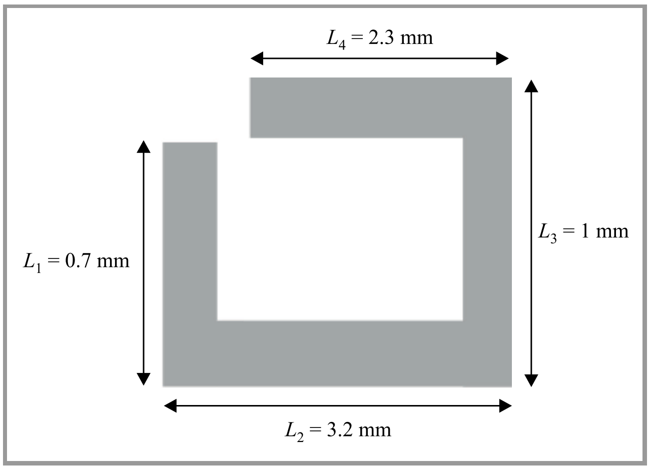


Fig. 6. C-shaped slot dimensions.

Table 2
Simulations versus theoretical predictions
for a band-notched antenna

L_1 [mm]	L_2 [mm]	L_3 [mm]	L_4 [mm]	L_{slot} [mm]	Predicted [GHz]	Simulated [GHz]
0.7	3.2	1	1.8	6.7	5.8	5.9
0.7	3.2	1	2.3	7.2	5.41	5.49
0.7	3.1	1	2.8	7.6	4.5	4.6
0.7	4.6	1	3.3	9.6	4.03	4.3

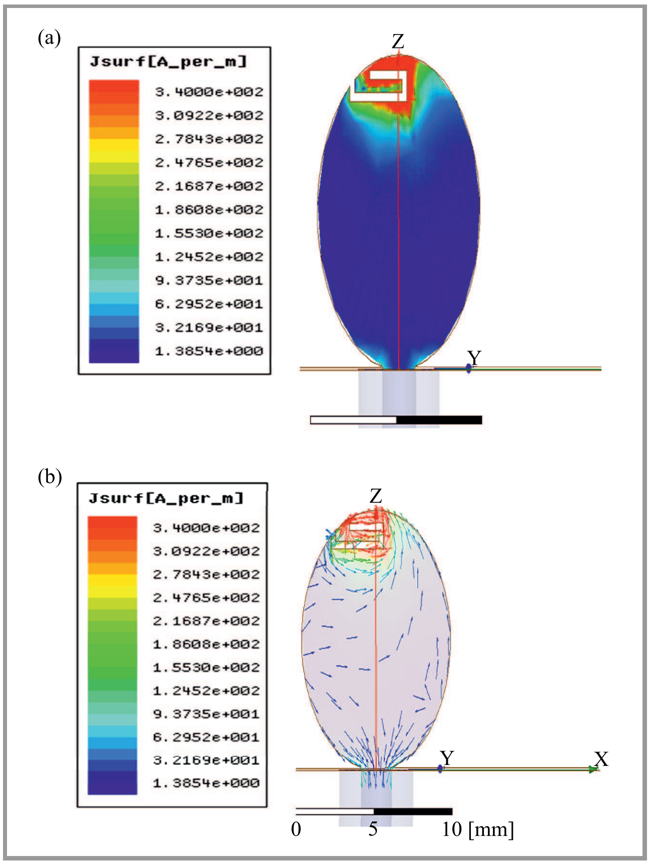


Fig. 7. Current distribution at 5.5 GHz.

The dimensions of the C-shaped slot for generating a relativity wide notch band for WLAN are shown in Fig. 6. The length of the slot can be deduced by:

$$L_{slot} \approx \frac{\lambda_g}{4} = L_1 + L_2 + L_3 + L_4 = 7.2 \text{ mm} . \qquad (3)$$

When the L_s length simulation values are compared to the predictions shown in Table 2, it is found that only a few differences exist.

To understand the phenomena behind notch band performance, the simulated current distributions on WLAN band notched center frequencies were analyzed on the proposed antenna, as shown in Fig. 7. It can be observed that the current is concentrated on the edge of the slot (Fig. 7a), and that current paths around the straight slots are oriented in opposite directions (Fig. 7b). When the antenna is working at the center notched band at 5.5 GHz, the outer slot behaves as a separator.

The length of L_{slot} is varied from 6.7 to 9.6 mm. The simulated VSWR is shown in Fig. 8. It is observed that when the length of the slot is increased, the band notch shifts towards a lower frequency and the bandwidth of the notch band is increased. This is because the slot length

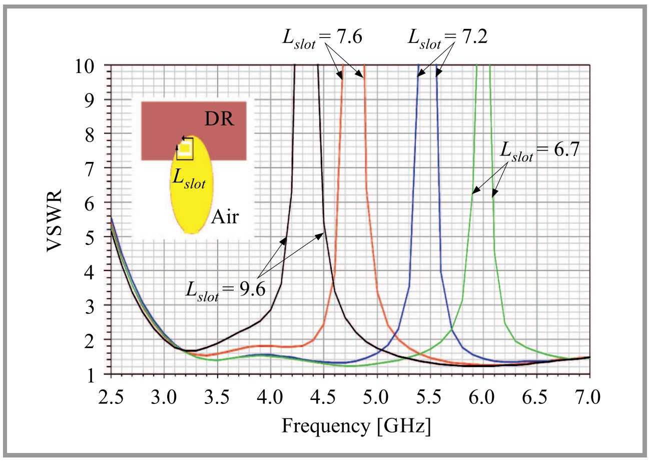


Fig. 8. VSWR characteristics of the single notch band for various L_{slot} ($h_1 = 8.5$ mm).

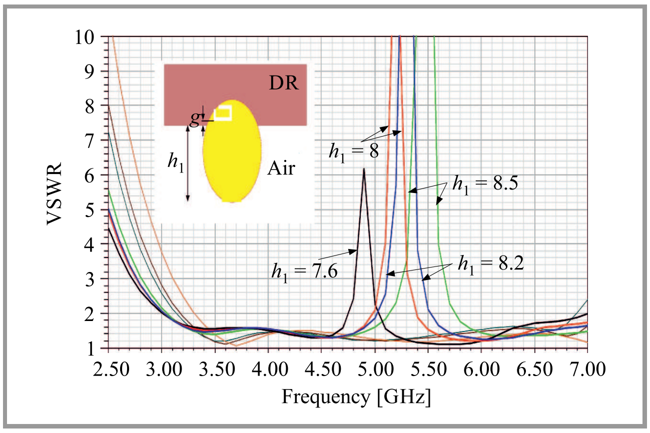


Fig. 9. VSWR characteristics of the single notch band for various h_1 ($L_{slot} = 7.2$ mm).

and notch frequency are inversely proportional to each other, as specified in Eq. (1). Interfering WLAN frequencies are within the band of 5.15 to 5.75 GHz and, hence, optimized L_{slot} is obtained at 7.2 mm for the center frequency of the WLAN band.

Gap g between the C-shaped slot and the air gap plays a crucial role in deciding the rejection band. As the gap increases from $g = 0$ ($h_1 = 8.5$ mm) to 0.9 mm ($h_1 = 7.6$ mm), the notched band shifts to the lower frequency spectrum, as shown in Fig. 9. For our requirement of rejection within the band 5.15 to 5.75 GHz, the optimized value is obtained as $h_1 = 8.5$ mm. It is observed that the notch bandwidth decreases when h_1 decreases, but with a lower peak rejection ratio.

3. Results and Discussion

The simulated VSWR plot of the proposed antenna is given in Fig. 10. It is clear that the band notch has been attained (5.15 to 5.75 GHz) and the results indicate a wide impedance bandwidth from 3.10 to 7.25 GHz. The comparison plot between the two different numerical analytical techniques, CST and HFSS, shows a similarity in verifying the performance of the antenna.

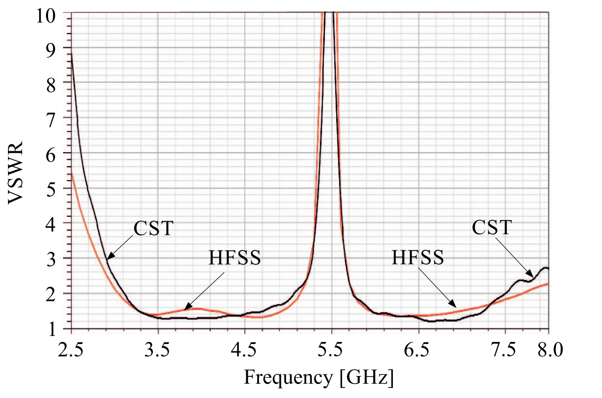


Fig. 10. Simulated VSWR using HFSS and CST software.

Figure 11 shows the simulated radiation in the E plane (x-z) and H plane (x-y) at frequencies of 3.5, 4.5, 5.5 and 6.5 GHz. The nature of H plane radiation patterns is omnidirectional, while the E plane radiation patterns are directional, which is mainly due to the effects of the metal coating. In both cases, the simulated results from the two software packages were found to be in close agreement. The antenna meets the directional requirement of UWB terminals.

The real gain comparison for the proposed DRA (with and without the C-shaped slot antenna) is shown in Fig. 12. Stable gain is observed over the entire UWB frequency range, except for band notches because the radiation at the notched band frequencies is attenuated. The real gain variation is 4.6 to 6.8 dBi. The decrease in the value of gain for the WLAN band is 9.0 dBi. As the ultra-band technol-

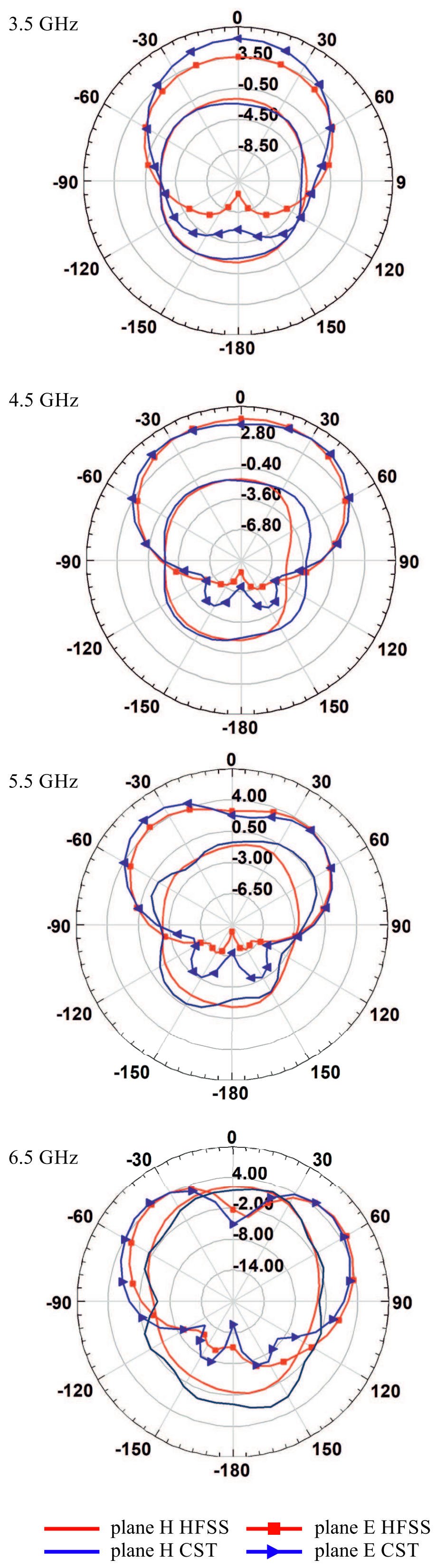


Fig. 11. HFSS and CST simulated directivity patterns in the E plane (x-z) and H plane (x-y) for the proposed antenna at 3.5, 4.5, 5.5, and 6.5 GHz.

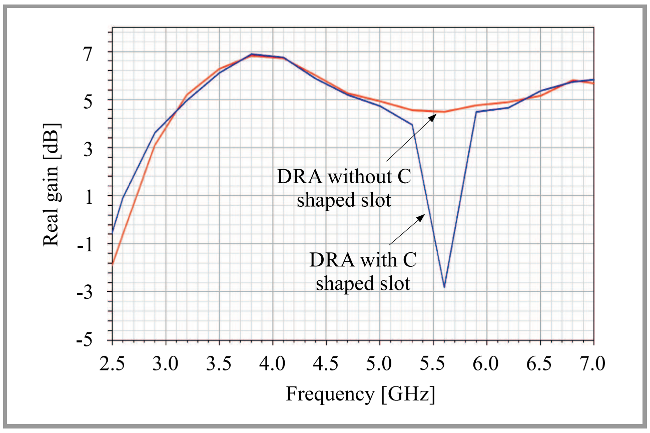


Fig. 12. Real gain versus frequency plot with and without C-shaped slot.

ogy works at a lower power level, the effect of the ultra-wideband radiation at the notched band is too weak to affect the WLAN communication system, which uses higher power levels.

4. Conclusion

The results of the simulation work conducted with the use of HFSS and CST software show that the proposed DRA provides a wide impedance bandwidth of approximately 80%, offering the range of 3.1 to 7.25 GHz, while providing one notched band operation at 5.5 GHz. This antenna is very simple in structure and has a very low overall height of $0.14\lambda_{\min}$ at its lowest operation frequency and it is able to work in the WiMAX system (3.2–3.8 GHz). This DRA is easy to fabricate and is capable of removing interference from the ultra-wideband system in the WLAN band. The impact of changes in dimensions and the position of the C-shaped slot on the band-notch characteristics of the proposed antenna was analyzed as well. It was observed that the notched band can be adjusted by changing the thickness of DRA. The air gap, the metal coating on the top and the position of the patch are important for improving DRA bandwidth. Furthermore, the proposed antenna demonstrated a good omnidirectional radiation pattern, an acceptable gain in operating frequencies and may be a good candidate for wireless applications.

References

[1] S. Keyrouz, and D. Caratelli, "Dielectric resonator antennas: basic concepts, design guidelines", *Int. J. on Antennas and Propag.*, vol. 2016, Article ID 6075680 (doi: 10.1155/2016/6075680).

[2] K. M. Luk and K. W. Leung, *Dielectric Resonator Antennas*. Hertfordshire, UK: Research Studies Press, 2003 (ISBN 9780863802638).

[3] R. N. Simons and R. Q. Lee, "Effect of parasitic dielectric resonator on CPW aperture-coupled dielectric resonator antenna", *IEE Proc. H, (Microw. Antennas and Propag.)*, vol. 140, no. 5, pp. 336–338, 1993 (doi: 10.1049/ip-h-2.1993.0052).

[4] M. S. Al Salameh, Y. M. M. Antar, and G. Seguin, "Coplanar-waveguide-fed slot-coupled rectangular dielectric resonator antenna", *IEEE Trans. on Antennas and Propag.*, vol. 50, no. 10, pp. 1415–1419, 2002 (doi: 10.1109/TAP.2002.802097).

[5] T. H. Chang and J. F. Kiang, "Broadband dielectric resonator antenna with metal coating", *IEEE Trans. on Antennas and Propagation*, vol. 55, no. 5, pp. 1254–1259, 2007 (doi: 10.1109/TAP.2007.895582).

[6] Q. Rao, T. A. Denidni, A. R. Sebak, and R. H. Johnston, "Compact independent dual-band hybrid resonator antenna with multifunctional beams", *IEEE Microw. and Wirel. Compon. Lett.*, vol. 5, pp. 239–242, 2006 (doi: 10.1109/LAWP.2006.875886).

[7] K. W. Leung and K. K. So, "Frequency-tunable designs of the linearly and circularly polarized dielectric resonator antenna using a parasitic slot", *IEEE Trans. on Antennas and Propag.*, vol. 53, no. 1, pp. 572–578, 2005 (doi: 10.1109/TAP.2004.838762).

[8] A. A. Kishk, Y. Yin, and A. W. Glisson, "Conical dielectric resonator antennas for wide-band applications", *IEEE Trans. on Antennas and Propag.*, vol. 50, no. 4, pp. 469–474, 2002 (doi: 10.1109/TAP.2002.1003382).

[9] B. N. Taralkar and A. R. Wadhekar, "Fractal dielectric resonator antenna for wideband applications", *Adv. Res. in Elec. and Electron. Engin.*, vol. 2, no. 5, pp. 1–3, 2015 [Online]. Available: https://www.krishisanskriti.org/vol_image/24Sep201509093401%20Bajrang%20N%20%20Taralkar.pdf

[10] Z. Chen and H. Wong, "Wideband glass and liquid cylindrical dielectric resonator antenna for pattern reconfigurable design", *IEEE Trans. on Antennas and Propag.*, vol. 65, no. 5, pp. 2157–2164, 2017 (doi: 10.1109/TAP.2017.2676767).

[11] A. Petosa, N. Simons, R. Siushansian, A. Ittipiboon, and M. Cuhaci, "Design and analysis of multi segment dielectric resonator antennas", *IEEE Trans. on Antennas and Propag.*, vol. 48, pp. 738–742, 2000 (doi: 10.1109/8.855492).

[12] Y. Coulibaly, T. A. Denidni, and H. Boutayeb, "Broadband microstrip fed dielectric resonator antenna for x-band applications", *IEEE Antennas and Wirel. Propag. Lett.*, vol. 7, pp. 341–345, 2008 (doi: 10.1109/LAWP.2008.921326).

[13] M. Lapierre, Y. M. M. Antar, A. Ittipiboon, and A. Petosa, "Ultra wideband monopole dielectric resonator antenna", *IEEE Microw. and Wirel. Compon. Lett.*, vol. 15, no. 1, pp.7–9, 2005 (doi: 10.1109/LMWC.2004.840952).

[14] M. Abedian, S. K. A. Rahim, Sh. Danesh, M. Khalily, and S. M. Noghabaei, "Ultrawideband dielectric resonator antenna with WLAN band rejection at 5.8 GHz", *IEEE Microw. and Wirel. Compon. Lett.*, vol. 12, pp. 1523–1526, 2013 (doi: 10.1109/LAWP.2013.2291271).

[15] Y. F. Wang, T. A. Denidni, Q. S. Zeng, and G. Wei, "Band-notched UWB rectangular dielectric resonator antenna" *Electron. Lett.*, vol. 50, no. 7, pp. 483–484, 2014 (doi: 10.1049/el.2014.0188).

[16] T. A. Denidni and Z. Weng, "Hybrid ultrawideband dielectric resonator antenna and band-notched designs", *IET Microw. Antennas Propag.*, vol. 5, no. 4, pp. 450–458, 2011 (doi: 10.1049/iet-map.2009.0425).

[17] Y. J. Cho, K. H. Kim, D. H. Choi, S. S. Lee, and S. Park, "A miniature UWB planar monopole antenna with 5GHz band rejection filter and the time domain characteristics", *IEEE Trans. on Antennas and Propag.*, vol. 54, no. 5, pp. 1453–1460, 2006 (doi: 10.1109/TAP.2006.874354).

[18] A. Syed and R. W. Aldhaheri, "A very compact and low profile UWB planar antenna with WLAN band rejection", *The Scient. World J.*, vol. 2016 (doi: 10.1155/2016/3560938).

[19] Y. Li, W. Li, and Q. Ye, "A reconfigurable triple-notch-band antenna integrated with defected microstrip structure band stop filter for ultra wideband cognitive radio applications", *Int. J. of Antennas and Propag.*, vol. 2013, no. 7, Article ID 472645 (doi: 10.1155/2013/472645).

[20] W. C. Liu and C. F. Hsu, "Dual-band CPW-fed Y-shaped monopole antenna for PCS/WLAN application", *Electron. Lett.*, vol. 41, no. 17, pp. 390–391, 2005 (doi: 10.1049/el:20057887).

[21] N. D. Trang, D. H. Lee, and H. C. Park, "Compact printed CPW-fed monopole ultra-wideband antenna with triple subband notched characteristics", *Electron. Lett.*, vol. 46, no. 17, pp. 1177–1179, 2010 (doi: 10.1049/el.2010.1140).

Mohamed Debab received his B.Sc. degree in Electronics from the Electronics Institute of the University of Djillali Liabes, Sidi Bel Abbes, Algeria, in 1998. Then he received an M.Sc. from the Department of Electronics, University of Djillali Liabes, Sidi Bel Abbes, Algeria, 2005. Currently, he is working as an Assistant Professor at the Department of Electronics, University of Hassiba Ben Bouali Chlef Algeria. His research interests focus on design and analysis of coplanar and dielectric antennas.

https://orcid.org/0000-0002-1779-0323
E-mail: debab_telecoms2005@hotmail.fr
Laboratory of Electromagnetism, Photonics
and Optronics (LEPO)
Djillali liabes University of Sidi Bel Abbès
22000 Sidi Bel Abbès, Algeria

Zoubir Mahdjoub received his B.Sc. degree in Electronics from the Electronics Institute of USTO of Oran, Algeria, in 1982, a Diploma of Advanced Studies, from the National Polytechnic Institute of Grenoble, France, in 1983, and a Ph.D. degree from the University of Claude Bernard de Lyon I, France, in 1987. Between 1988 and 1991, he was the President of the Scientific Council of the Electronics Institute, University of Djillali Liabes, Sidi Bel Abbes, Algeria. Since 1988, he has been involved in conducting research on microwaves, telecommunications and photonics. Between 1998 and 2006, he was the head of the Electronics Department, University of Djillali Liabes. He is now a full professor and a vice dean for post-graduation programs at the Electrical Engineering Faculty of the same University.

E-mail: mahdjoubz@yahoo.com
Laboratory of Electromagnetism, Photonics
and Optronics (LEPO)
Djillali liabes University of Sidi Bel Abbès
22000 Sidi Bel Abbès, Algeria

Paper

Product of Three Random Variables and its Application in Relay Telecommunication Systems in the Presence of Multipath Fading

Dragana Krstic[1], Petar Nikolic[2], Danijela Aleksic[3], Sinisa Minic[4],
Dragan Vuckovic[5], and Mihajlo Stefanovic[1]

[1] *Faculty of Electronic Engineering, University of Niš, Niš, Serbia*
[2] *TigarTyres, Pirot, Serbia*
[3] *College of Applied Technical Sciences Niš, Serbia*
[4] *Teachers' Training Faculty, Prizren-Leposavic, University of Pristina, Kosovska Mitrovica, Serbia*
[5] *Faculty of Economics and Engineering Management, University Business Academy, Novi Sad, Serbia*

https://doi.org/10.26636/jtit.2019.130018

Abstract—In this paper, the product of three random variables (RVs) will be considered. Distribution of the product of independent random variables is very important in many applied problems, including wireless relay telecommunication systems. A few of such products of three random variables are observed in this work: the level crossing rate (LCR) of the product of a Nakagami-*m* random variable, a Rician random variable and a Rayleigh random variable, and of the products of two Rician RVs and one Nakagami-*m* RV is calculated in closed forms and presented graphically. The LCR formula may be later used for derivation of average fade duration (AFD) of a wireless relay communication radio system with three sections, working in the multipath fading channel. The impact of fading parameters and multipath fading power on the LCR is analyzed based on the graphs presented.

Keywords—level crossing rate, Nakagami-m fading, Rayleigh fading, relay telecommunication systems, Rician fading.

1. Introduction

Statistical characteristics of products and ratios of random variables (RVs) are essential in analyzing the performance of contemporary wireless telecommunications systems, as well as in solving numerous applied problems. The products of RVs are encountered naturally in such applications as: channel modeling, multihop wireless relaying systems, cascaded fading channels, MIMO keyhole systems [1], quantum physics, signal processing, tensor sensing problem, the rate offset of the hybrid automatic repeat request (H-ARQ) transmission, and even in biological and physical sciences, econometrics, classification, ranking and selection [2].

Because of that, in recent years, the products and ratios of random processes are investigated in the literature by many researchers [1]–[6]. In the 1960s, Donahue, Springer, Thompson and Lomnicki started with derivations concerning the distributions of the products of two RVs [7]–[10]. At the beginning of this century, interest in this area has increased again [11]—[13]. The latest works, with significant application in wireless communications systems, are [14]–[17].

Computational algorithms for derivating the distribution of the product of two RVs are given in [9]. The product and the ratio of two independent, Student's t distributed RVs, are observed in [10]. The derivation of the probability density function (PDF) of the product of two independent, non-identical, and triangularly distributed RVs, by using integral calculation, is presented in [11].

The problem of characterizing products of independent RVs is investigated for normal RVs, products of their absolute values, and products of their squares in [13]. Power-log series expansions of cumulative distribution functions (CDF), based on the theory of Fox H functions, is computed. It is numerically shown that CDF is well approximated by the lowest orders of this expansion for small arguments. The moment generating functions (MGF) in terms of Meijer G functions are also computed for two non-negative RVs. In that paper, the fading amplitudes of cascaded fading channels have the distribution of the product of Nakagami-*m* RVs, as in [5] and [6], and of the product of Rician RVs, as in [18] and [19].

Signal envelope variations, called fading, are results of reflections, refractions, diffraction and scattering. They can be described by several distributions. So, Rayleigh [20] and Nakagami-*m* [21] distributions are used when no dominant component is present. Signal envelope variation is modeled by Rician distribution when a line-of-sight (LOS) dominant component exists in the channel [22].

Level crossing rate (LCR) of a product of two Nakagami-*m* random processes is analyzed in [6]. Then, the average fade duration (AFD) of a wireless relay communications system consisting of two sections in a Nakagami-*m* short term fading channel is obtained. The performance of the product of arbitrary and independent RVs with a general α–μ distribution is given in [15]. The closed-form expressions for PDF, CDF and moments are obtained and the calculation process used to obtain the amount of fading (AoF) and outage probability (OP) in cascaded channels is presented.

An analysis of the performance of the product of two independent and non-identically distributed κ–μ RVs is presented in [17], where analytical formulations for PDF, CDF and MGF are obtained. By using these formulations, closed-form expressions for higher order moments, AoF and channel quality estimation index are derived, as are analytical formulas for OP, average channel capacity, average symbol error probability (SEP) and average bit error probability. This applies to different fading scenarios, such as double Rayleigh, double Rician, double Nakagami-*m*, κ–μ/Nakagami-*m*, and Rician/Nakagami-*m*, identified as special cases.

The cascaded keyhole channels may be modeled using the product of individual channels [23]. Further, independent and identically distributed (i.i.d.) double fading channels in a line-of-sight (LOS) environment, typical for keyhole Multiple input multiple output (MIMO) systems, are analyzed in [24]. LOS double fading, i.e. double Rician fading in MIMO channels, is investigated and the density function of the signal-to-noise ratio (SNR) is discussed. After obtaining the exact expression in a form with infinite series, an approximation formula of SNR density is presented by using the Nakagami-*m* approximation of Rician distribution.

In [25], the product of Nakagami-*m* RV, Rician RV and Rayleigh RV is analyzed. LCR of this product is calculated. The results obtained may also be used for the evaluation the AFD of a relay wireless communications system with three sections in the presence of Nakagami-*m* fading in the first section, Rician fading in the second section and Rayleigh fading in the third section. LCR of the product of three independent Rician RVs is observed in [26].

In this article, except for the results from [25], the product of two Rician RVs and one Nakagami-*m* RV is processed. The result can be applied for derivation, in a closed form, of the second order performance of a wireless relay communications system with three sections operating in Rician and Nakagami-*m* fading environments. The formulas are validated by numerical results and impact of the individual parameters is analyzed.

This work is composed of four sections. Section 1 serves as an introduction and describes previous works in the area. In two next sections, the product of three random variables is given and an expression for LCR in a closed form is performed for two different sets of RVs. The influence of parameters is shown via several graphics for both sets. The work ends with conclusions presented in Section 4.

2. Derivation of LCR of Product of Nakagami-*m*, Rician and Rayleigh Random Variables

Here, we examine the scenario involving a wireless relay communications system with three sections. The signal envelope at the output of the relay communications system with three sections is the product of envelopes at the individual sections. In the first example observed, the Nakagami-*m* signal envelope is at the first section, the Rician signal envelope at the second section and the Rayleigh signal envelope at the third section.

As a rule, it is first necessary to calculate the probability density function. By using PDF, bit error probability (BEP) can be evaluated, and by using CDF, outage probability can be obtained [27], [28]. OP and BEP are the first order performance measure of the wireless communications system. OP may be calculated as probability that the signal envelope is below the threshold [27], [29]. Level crossing rate is the second order statistic measure of the wireless communications system and is associated with envelope fading, as well as with average fade duration. LCR shows how often the envelope crosses a specified level and may be calculated as the number of crossings at this defined level. AFD shows how long the envelope remains below a specified level and can be evaluated as the ratio of OP and LCR. These two quantities are second order statistics because they are affected not only by the scattering in environment, but also by the speed of mobile stations. Here, PDF of the product of Nakagami-*m* RV, Rician RV and Rayleigh RV will be obtained by using the transformation method. Further, using this PDF, CDF and moments can be evaluated, as can be the level crossing rate.

2.1. Distribution of Random Variables

Nakagami-*m* random variable x_1 follows the distribution defined in [21]:

$$p_{x_1}(x_1)=\frac{2}{\Gamma(m)}\left(\frac{m}{\Omega_1}\right)^m x_1^{2m-1}e^{-\frac{m}{\Omega_1}x_1^2},\ m\geq\frac{1}{2},\ x_1\geq 0\,, \quad (1)$$

where $\Gamma(.)$ is a gamma function. This distribution has two parameters. The first parameter controls spread. Actually, $\Omega_1=E[x_1^2]$ is the average power of the multipath scattering field. m is the fading depth parameter or the shape factor of the Nakagami distribution [30]. For RV x_1 the shape parameter is:

$$m=\frac{\Omega_1^2}{E\left\{\left[x_1^2-\Omega_1^2\right]^2\right\}}\ .$$

It describes the fading degree of the propagation environment caused by the interference of scattering and multipath phenomena. So, the severity of fading is indicated by the Nakagami parameter m.

Random variable x_2 follows Rician distribution [22]:

$$p_{x_2}(x_2) = \frac{2(\kappa+1)}{\Omega_2} \sum_{i_1=0}^{\infty} \left(\frac{\kappa(\kappa+1)}{\Omega_2} \right)^{i_1} \times \frac{1}{(i_1!)^2} x_2^{2i_1+1} \, \mathrm{e}^{-\frac{\kappa+1}{\Omega_2} x_2^2}, \quad x_2 \geq 0 \ , \tag{2}$$

where Ω_2 is the average received power for x_2 and κ is the Rician factor. Rician factor κ is defined as a ratio of dominant component's power and the scattering components' powers. This type of fading, called Rician fading, is very often observed in microcellular and mobile satellite applications [27].

Rayleigh distribution can be easily derived from Rician distribution for Rician factor $\kappa = 0$. For $\kappa = \infty$ we have no fading, i.e. a channel with no multipath and only a LOS component. The fading parameter κ is therefore a measure of the severity of fading: a small κ implies severe fading, a large κ implies more mild fading [31].

Two-dimensional isotropic scattering, where the arriving waves arrive to the receiver from all directions, with equal probability, is a scattering model that is often used for the communication channel in a macrocellular system. For this type of scattering environment, the received envelope is Rayleigh distributed at any time, and is said to be Rayleigh fading.

Random variable x_3 follows Rayleigh distribution [31]:

$$p_{x_3}(x_3) = \frac{x_3}{\Omega_3} \mathrm{e}^{-\frac{x_3^3}{\Omega_3}}, \quad x_3 \geq 0 \ , \tag{3}$$

where Ω_3 is the average received signal power of signal x_3, i.e. the received power based alone on path loss and shadowing alone [31].

2.2. Product of Three Random Variables

The product of three random variables x_1, x_2 and x_3 is:

$$x = x_1 x_2 x_3 \ . \tag{4}$$

Then, it is valid that:

$$x_1 = \frac{x}{x_2 x_3} \ . \tag{5}$$

The first time derivative of x is:

$$\dot{x} = \dot{x}_1 x_2 x_3 + x_1 \dot{x}_2 x_3 + x_1 x_2 \dot{x}_3 \ . \tag{6}$$

The first time derivative of x_i has Gaussian distribution:

$$p_{\dot{x}_i}(\dot{x}_i) = \frac{1}{\sqrt{2\pi\dot{\sigma}_i^2}} \, \mathrm{e}^{-\frac{\dot{x}_i^2}{2\dot{\sigma}_i^2}}, \quad -\infty < \dot{x}_1 < \infty \ ,$$

where $\dot{\sigma}_1^2 = \frac{\pi^2 f_m^2 \Omega_i}{m}$ and f_m being maximal Doppler frequency. The processes x_i and $\dot{x}_i$ are considered to be independent, as Rice demonstrated in [32].

This is a very interesting result which shows that, in the Nakagami case and in the Rayleigh and Rice cases, x_i and $\dot{x}_i$ are mutually independent random variables [33], i.e., it is valid that [32]:

$$p_{x_i \dot{x}_i}(x_i \dot{x}_i) = p_{x_i}(x_i) p_{\dot{x}_i}(\dot{x}_i) \ .$$

Moreover, the probability density function of the time derivative of the Nakagami envelope is also Gaussian distributed as are the time derivatives of both Rayleigh and Rice envelopes [34].

So, all random variables $\dot{x}_1$, $\dot{x}_2$, and $\dot{x}_3$ have Gaussian distribution. A linear combination of Gaussian RVs is a Gaussian RV. The mean signal level of $\dot{x}$ is:

$$\overline{\dot{x}} = \overline{\dot{x}}_1 x_2 x_3 + x_1 \overline{\dot{x}}_2 x_3 + x_1 x_2 \overline{\dot{x}}_3 = 0 \ , \tag{7}$$

because:

$$\overline{\dot{x}}_1 = \overline{\dot{x}}_2 = \overline{\dot{x}}_3 = 0 \ . \tag{8}$$

The variance of $\dot{x}$ is:

$$\sigma_{\dot{x}}^2 = x_2^2 x_3^2 \sigma_{\dot{x}_1} + x_1^2 x_3^2 \sigma_{\dot{x}_2} + x_1^2 x_2^2 \sigma_{\dot{x}_3} \ , \tag{9}$$

where:

$$\begin{aligned} \sigma_{\dot{x}_1} &= \pi^2 f_m^2 \frac{\Omega_1}{m} \ , \\ \sigma_{\dot{x}_2} &= \pi^2 f_m^2 \frac{\Omega_2}{\kappa+1} \ , \\ \sigma_{\dot{x}_3} &= \pi^2 f_m^2 \Omega_3 \ . \end{aligned} \tag{10}$$

After substituting, the expression for variance becomes:

$$\begin{aligned} \sigma_{\dot{x}}^2 &= \pi^2 f_m^2 \left(x_2^2 x_3^2 \frac{\Omega_1}{m} + x_1^2 x_3^2 \frac{\Omega_2}{\kappa+1} + x_1^2 x_2^2 \Omega_3 \right) \\ &= \pi^2 f_m^2 x_2^2 x_3^2 \frac{\Omega_1}{m} \left(1 + \frac{x^2}{x_2^4 x_3^2} \frac{\Omega_2}{\Omega_1} \frac{m}{\kappa+1} + \frac{x^2}{x_2^2 x_3^4} \frac{\Omega_3}{\Omega_1} m \right) . \end{aligned} \tag{11}$$

The joint probability density function of x, $\dot{x}$, x_2 and x_3 is:

$$p_{x\dot{x}x_2x_3}(x\dot{x}x_2x_3) = p_{\dot{x}}(\dot{x}/xx_2x_3) p_x(x/x_2x_3) p_{x_2}(x_2) p_{x_3}(x_3), \tag{12}$$

where

$$p_x(x/x_2x_3) = \left| \frac{dx_1}{dx} \right| p_{x_1}\left(\frac{x}{x_2 x_3} \right) , \tag{13}$$

$$\frac{dx_1}{dx} = \frac{1}{x_2 x_3} \ . \tag{14}$$

The joint probability density function of x and $\dot{x}$ is:

$$p_{x\dot{x}}(x\dot{x}) = \int_0^\infty \mathrm{d}x_2 \int_0^\infty \mathrm{d}x_3 \, p_{\dot{x}}(\dot{x}/xx_2x_3) \frac{1}{x_2 x_3} \times p_{x_1}\left(\frac{x}{x_2 x_3} \right) p_{x_2}(x_2) p_{x_3}(x_3) \ . \tag{15}$$

2.3. LCR of Product of Three Random Variables

Level crossing rate of x in a fading environment is [35]:

$$N_x = \int_0^\infty \mathrm{d}\dot{x}\dot{x}p_{x\dot{x}}(x\dot{x}) \ . \tag{16}$$

For our case LCR is:

$$N_x = \int_0^\infty \mathrm{d}x_2 \int_0^\infty \mathrm{d}x_3 \frac{1}{x_2x_3} p_{x_1}\left(\frac{x}{x_2x_3}\right) p_{x_2}(x_2) p_{x_3}(x_3)$$
$$\times \int_0^\infty \mathrm{d}\dot{x}\dot{x}p_{\dot{x}}(\dot{x}/xx_2x_3) = \int_0^\infty \mathrm{d}x_2 \int_0^\infty \mathrm{d}x_3 \frac{1}{x_2x_3} p_{x_1}\left(\frac{x}{x_2x_3}\right) p_{x_2}(x_2)$$
$$\times p_{x_3}(x_3)\frac{1}{\sqrt{2\pi}}\sigma_{\dot{x}} = \int_0^\infty \mathrm{d}x_2 \int_0^\infty \mathrm{d}x_3 \frac{1}{x_2x_3} p_{x_1}\left(\frac{x}{x_2x_3}\right) p_{x_2}(x_2)$$
$$\times p_{x_3}(x_3)\frac{1}{\sqrt{2\pi}}\pi f_m x_2 x_2 \frac{\Omega^{\frac{1}{2}}}{m^{\frac{1}{2}}}\left(1 + \frac{x^2}{x_2^4x_3^2}\frac{\Omega_2}{\Omega_1}\frac{m}{\kappa+1}\right.$$
$$\left. + \frac{x^2}{x_2^2x_3^4}\frac{\Omega_3}{\Omega_1}m\right)^{\frac{1}{2}} = \frac{1}{\sqrt{2\pi}}\pi f_m \frac{\Omega_1^{\frac{1}{2}}}{m^{\frac{1}{2}}} \int_0^\infty \mathrm{d}x_2 \int_0^\infty \mathrm{d}x_3$$
$$\times p_{x_1}\left(\frac{x}{x_2x_3}\right) p_{x_2}(x_2) p_{x_3}(x_3)\left(1 + \frac{x^2}{x_2^4x_3^2}\frac{\Omega_2}{\Omega_1}\frac{m}{\kappa+1}\right.$$
$$\left. + \frac{x^2}{x_2^2x_3^4}\frac{\Omega_3}{\Omega_1}m\right)^{\frac{1}{2}} = \frac{1}{\sqrt{2\pi}}\pi f_m \frac{\Omega_1^{\frac{1}{2}}}{m^{\frac{1}{2}}}\frac{2}{\Gamma(m)}\left(\frac{m}{\Omega_1}\right)^m$$
$$\times x^{2m-1}\frac{2(\kappa+1)}{\Omega_2}\sum_{i_1=0}^{\infty}\left(\frac{\kappa(\kappa+1)}{\Omega_2}\right)^{i_1}\frac{1}{(i_1!)^2}\frac{2}{\Omega_3}$$
$$\times \int_0^\infty \mathrm{d}x_2 \int_0^\infty \mathrm{d}x_3 x_2^{-2m+1+2i_1+1} x_3^{-2m+1+1} \mathrm{e}^{-\frac{m}{\Omega_1}\frac{x^2}{x_2^2x_3^2} - \frac{\kappa+1}{\Omega_2}x_2^2 - \frac{1}{\Omega_3}x_3^2}$$
$$\times \left(1 + \frac{x^2}{x_2^4x_3^2}\frac{\Omega_2}{\Omega_1}\frac{m}{\kappa+1} + \frac{x^2}{x_2^2x_3^4}\frac{\Omega_3}{\Omega_1}m\right)^{\frac{1}{2}} \ . \tag{17}$$

The previous two-fold integral may be solved using the Laplace approximation theorem for the solution the two-fold integrals [36], [37]:

$$\int_0^\infty \mathrm{d}x_2 \int_0^\infty \mathrm{d}x_3 (x_2,x_3)\,\mathrm{e}^{\lambda f(x_2,x_3)} = \frac{\pi g(x_{20},x_{30})}{\lambda B(x_{20},x_{30})}\mathrm{e}^{\lambda f(x_{20},x_{30})}, \tag{18}$$

where B is the matrix:

$$B(x_{20},x_{30}) = \begin{vmatrix} \frac{\partial^2 f(x_{20},x_{30})}{\partial x_{20}^2} & \frac{\partial^2 f(x_{20},x_{30})}{\partial x_{20}\partial x_{30}} \\ \frac{\partial^2 f(x_{20},x_{30})}{\partial x_{20}\partial x_{30}} & \frac{\partial^2 f(x_{20},x_{30})}{\partial x_{30}^2} \end{vmatrix} , \tag{19}$$

and x_{20} and x_{30} are solution of the equations:

$$\frac{\partial f(x_{20},x_{30})}{\partial x_{20}} = 0, \quad \frac{\partial f(x_{20},x_{30})}{\partial x_{30}} = 0 \ . \tag{20}$$

For considered case, it is:

$$g(x_2,x_3) = x_2^{-2m+2i_1+2} x_3^{-2m+2}$$
$$\times \left(1 + \frac{x^2}{x_2^4x_3^2}\frac{\Omega_2}{\Omega_1}\frac{m}{\kappa+1} + \frac{x^2}{x_2^2x_3^4}\frac{\Omega_3}{\Omega_1}m\right)^{\frac{1}{2}}, \tag{21}$$

$$f(x_2,x_3) = -\frac{m}{\Omega_1}\frac{x^2}{x_2^2x_3^2} - \frac{(\kappa+1)}{\Omega_2}x_2^2 - \frac{1}{\Omega_3}x_3^2 \ , \tag{22}$$

$$\frac{\partial f(x_2,x_3)}{\partial x_2} = \frac{2m}{\Omega_1}\frac{x^2}{x_2^3x_3^2} - \frac{2(\kappa+1)}{\Omega_2}x_2 \ , \tag{23}$$

$$\frac{\partial f(x_2,x_3)}{\partial x_3} = \frac{2m}{\Omega_1}\frac{x^2}{x_2^2x_3^3} - \frac{2}{\Omega_3}x_3 \ , \tag{24}$$

The solutions of the next two equations are x_{20} and x_{30}:

$$\frac{2m}{\Omega_1}\frac{x^2}{x_2^3x_3^2} - \frac{2(\kappa+1)}{\Omega_2}x_2 = 0 \ , \tag{25}$$

$$\frac{2m}{\Omega_1}\frac{x^2}{x_2^2x_3^3} - \frac{2}{\Omega_3}x_3 = 0 \ . \tag{26}$$

They should be introduced in Eq. (18) for solving two-fold integral from Eq. (17). In this manner LCR of the product of Nakagami-*m*, Rician and Rayleigh random variables will be obtained in a closed form.

2.4. Numerical Examples and Discussion

The level crossing rate of the product of Nakagami-*m* random variable, Rician RV and Rayleigh RV is shown in the next few figures versus resulting signal x for different values of fading parameters and signal powers.

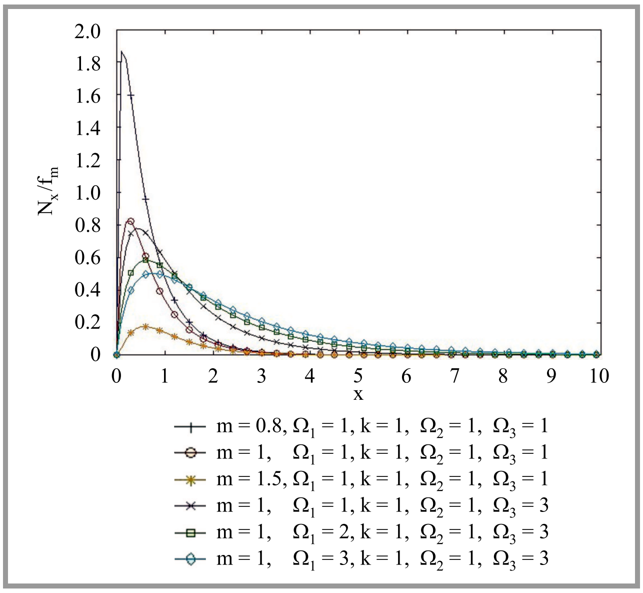


Fig. 1. LCR normalized by f_m depending on signal envelope x for various values of parameters m and Ω_1.

Dependence of the LCR, normalized by f_m, on the resulting signal x, for various values of parameters m and Ω_1 is presented in Fig. 1. It is possible to notice that LCR increases for lower values of resulting signal and decreases for greater values of the resulting signal. All curves reach the maximum and start to decline. Lower values of the resulting signal have a greater impact on LCR. LCR increases for low values of Nakagami-m small scale fading parameter m. The impact of resulting x on LCR is larger for smaller magnitudes of parameter m. LCR is larger for smaller values of m.

From this picture, the influence of power Ω_1 can also be observed. For low values of x, LCR increases with the reduction of power Ω_1, but for bigger values of x, LCR increases along with the growth of power Ω_1.

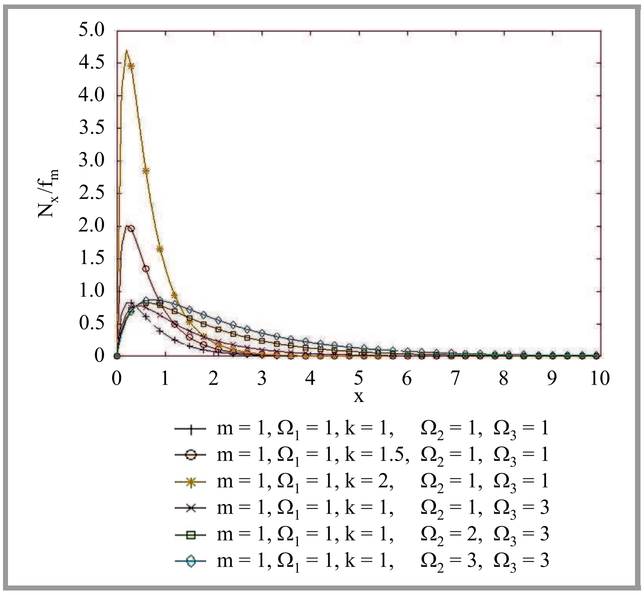


Fig. 2. LCR normalized by f_m for different parameters κ and Ω_2.

Figure 2 shows the influence of the other two parameters: Rician factor κ and signal power Ω_2. LCR becomes bigger as the Rician factor κ grows. The influence of x on LCR is greater for lower values of Rician factor κ. The impact of Nakagami-m fading parameter m on LCR is higher for bigger values of Rician factor κ. From this figure, one can also see that LCR is larger for greater values of power Ω_2.

In Fig. 3 the impact of power Ω_3 is shown. Based on the image, one may remark that LCR is higher for bigger values of Ω_3 and low values of x. For higher values of x, LCR is greater for smaller Ω_3. The small resulting signal x exerts a greater impact on LCR.

The results obtained may be used to evaluate LCR of the product of Nakagami-m and two Rayleigh RVs, LCR of Rician and two Rayleigh RVs, and LCR of the product of three Rayleigh (3* Rayleigh) RVs. This can be achieved because Nakagami-m and Rician distributions are of the general variety. For the same reason, LCR of the product of three independent Rician RVs from [26] can be used for determination of LCR of the product of three Rayleigh RVs, or the LCR of the product of two Rician RVs and Rayleigh RV, or LCR of the product of Rician RV and two Rayleigh RVs, because Rayleigh distribution may be easily derived from Rician distribution for Rician factor $\kappa = 0$.

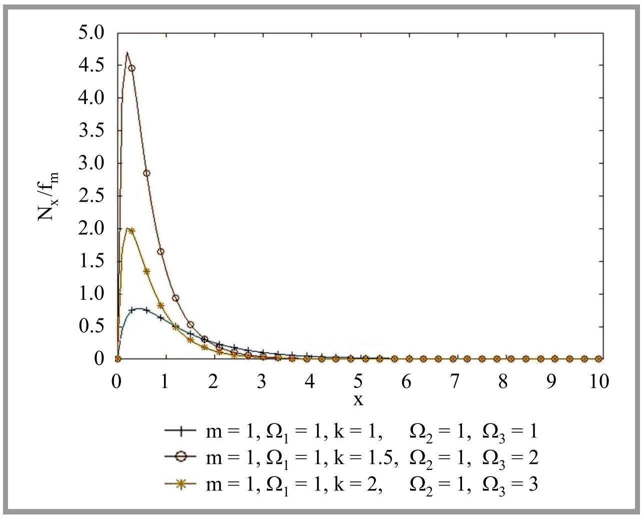


Fig. 3. LCR normalized by f_m for various values of Ω_3.

If Nakagami fading severity parameter $m = \frac{1}{2}$, Nakagami distribution is reduced to unilateral (one-sided) Gaussian distribution. For $m = 1$, Nakagami distribution reduces to Rayleigh distribution, and for $m > 1$, Nakagami distribution is reduced to Rician distribution. The ratio between Rician factor κ and parameter m is [27], [38]:

$$\kappa = \frac{\sqrt{m^2 - m}}{m - \sqrt{m^2 - m}}, \quad m > 1 \ .$$

On the other hand, for [31]:

$$m = \frac{(\kappa + 1)^2}{2(\kappa + 1)} \ ,$$

the distribution in Eq. (1) is approximately Rician fading with parameter κ. For $m = \infty$ we get an additive white Gaussian noise (AWGN) channel without fading. We see that as m increases, fading decreases.

As the Nakagami distribution does not contain a Bessel function, it can get the close form solution more convenient than Rician distribution [30].

Thus, the Nakagami distribution may model Rayleigh distribution and Rician distribution, with certain restrictions [21]. Note that some empirical measurements support values of the m parameter being equal to less than one, in which case the Nakagami fading causes a more severe performance degradation than Rayleigh fading.

3. LCR of Product of Two Rician and Nakagami-*m* Random Variables

In the second example, presented in this section, Rician fading exists in the first two sections and Nakagami-m fading is present in the third section. These results are applicable in analyzing the performance of multi-hop relay wireless telecommunications systems when the signal level is much higher than the noise level. In such a case, the noise level can be ignored. For that matter, the output signal is a product of as many random variables as there are sections in the relay system [38].

3.1. Distribution of the Second Set of Random Variables

Let random variables x_4 and x_5 have Rician distribution [22]:

$$p_{x_4}(x_4) = \frac{2(\kappa_1+1)}{\Omega_1}\sum_{i_2=0}^{\infty}\left(\frac{\kappa_1(\kappa_1+1)}{\Omega_1}\right)^{i_2} \times \frac{1}{(i_2!)^2}x_4^{2i_2+1}\,\mathrm{e}^{-\frac{\kappa_1+1}{\Omega_1}x_4^2}\,, \quad x_4 \geq 0\,, \tag{27}$$

$$p_{x_5}(x_5) = \frac{2(\kappa_2+1)}{\Omega_2}\sum_{i_3=0}^{\infty}\left(\frac{\kappa_2(\kappa_2+1)}{\Omega_2}\right)^{i_3} \times \frac{1}{(i_3!)^2}x_5^{2i_3+1}\,\mathrm{e}^{-\frac{\kappa_2+1}{\Omega_2}x_5^2}\,, \quad x_5 \geq 0, \tag{28}$$

and let random variable x_6 have Nakagami-m distribution [29]:

$$p_{x_6}(x_6)=\frac{2}{\Gamma(m_3)}\left(\frac{m_3}{\Omega_3}\right)^{m_3}x_3^{2m_3-1}\mathrm{e}^{-\frac{m_3}{\Omega_3}x_6^2},\; x_4 \geq 0,\, x_6 \geq 0, \tag{29}$$

where Ω_i, $i = 1, 2, 3$, are powers of RVs x_i, $i = 4, 5, 6$, κ_1 and κ_2 are Rician factors for variables x_4 and x_5, and m_3 is Nakagami-m fading severity parameter of RV x_6.

3.2. Product of Three Random Variables

Here, the random variable x is defined as a product of x_i, $i = 4, 5, 6$:

$$x = \prod_{i=4}^{6} x_i\;. \tag{30}$$

The first time derivative of x is:

$$\dot{x} = \dot{x}_4x_5x_6 + x_4\dot{x}_5x_6 + x_4x_5\dot{x}_6\;, \tag{31}$$

the average value of x is:

$$\overline{\dot{x}} = \overline{\dot{x}}_4x_5x_6 + x_4\overline{\dot{x}}_5x_6 + x_4x_5\overline{\dot{x}}_6 = 0\;, \tag{32}$$

because [29]

$$\overline{\dot{x}}_4 = \overline{\dot{x}}_5 = \overline{\dot{x}}_6 = 0\;. \tag{33}$$

The variance of $\dot{x}$ is given by:

$$\sigma_{\dot{x}}^2 = x_5^2x_6^2\sigma_{\dot{x}_4}^2 + x_4^2x_6^2\sigma_{\dot{x}_5}^2 + x_4^2x_5^2\sigma_{\dot{x}_6}^2\;, \tag{34}$$

with:

$$\sigma_{\dot{x}_4}^2 = \pi f_m^2\frac{\Omega_1}{\kappa_1+1}\;, \tag{35}$$

$$\sigma_{\dot{x}_5}^2 = \pi f_m^2\frac{\Omega_2}{\kappa_2+1}\;, \tag{36}$$

$$\sigma_{\dot{x}_6}^2 = \pi f_m^2\frac{\Omega_3}{m_3}\;. \tag{37}$$

After transformation of Eqs. (35)–(37) into Eq. (34), the variance is:

$$\begin{aligned}\sigma_{\dot{x}}^2 &= \pi^2 f_m^2\left(x_5^2x_6^2\frac{\Omega_1}{\kappa_1+1} + \frac{x^2}{x_5^2}\frac{\Omega_2}{\kappa_2+1} + \frac{x^2}{x_6^2}\frac{\Omega_3}{m_3}\right)\\ &= \pi^2 f_m^2x_5^2x_6^2\frac{\Omega_1}{\kappa_1+1}\\ &\times\left(1 + \frac{x^2}{x_5^4x_6^2}\frac{\Omega_2}{\kappa_2+1}\frac{\kappa_1+1}{\Omega_1} + \frac{x^2}{x_5^2x_6^4}\frac{\Omega_3}{m_3}\frac{\kappa_1+1}{\Omega_1}\right).\end{aligned} \tag{38}$$

Joint PDF of x, $\dot{x}$, x_5 and x_6 is:

$$p_{x\dot{x}x_5x_6}(x\dot{x}x_5x_6) = p_{\dot{x}}\left(\dot{x}/xx_5x_6\right) \times p_x\left(x/x_5x_6\right)p_{x_5}(x_5)p_{x_6}(x_6)\,, \tag{39}$$

and joint PDF of x and $\dot{x}$:

$$\begin{aligned}p_{x\dot{x}}(x\dot{x}) &= \int_0^\infty \mathrm{d}x_5\int_0^\infty \mathrm{d}x_6p_{x\dot{x}x_5x_6}(x\dot{x}x_5x_6)\\ &=\int_0^\infty \mathrm{d}x_5\int_0^\infty \mathrm{d}x_6p_{\dot{x}/xx_5x_6}\left(\dot{x}/xx_5x_6\right)p_x\left(x/x_5x_6\right)p_{x_5}(x_5)p_{x_6}(x_6),\end{aligned} \tag{40}$$

with:

$$p_x\left(x/x_5x_6\right) = \left|\frac{\mathrm{d}x_4}{\mathrm{d}x}\right|p_{x_4}\left(\frac{x}{x_5x_6}\right)\,, \tag{41}$$

$$\frac{\mathrm{d}x_4}{\mathrm{d}x} = \frac{1}{x_5x_6}\;. \tag{42}$$

The expression for $p_{x\dot{x}}(x\dot{x})$ in Eq. (40), after some replacements is:

$$p_{x\dot{x}}(x\dot{x}) = \int_0^\infty \mathrm{d}x_5\int_0^\infty \mathrm{d}x_6p_{\dot{x}}\left(\dot{x}/xx_5x_6\right) \times \frac{1}{x_5x_6}p_{x_4}\left(\frac{x}{x_5x_6}\right)p_{x_5}(x_5)p_{x_6}(x_6)\;. \tag{43}$$

3.3. LCR of Product of the Second Set of Random Variables

Level crossing rate of x is defined by Eq. (16) [39]. LCR of product x from Eq. (30), with $p_{x\dot{x}}(x\dot{x})$ from Eq. (43), is:

$$N_x = \int_0^\infty \mathrm{d}x_5\int_0^\infty \mathrm{d}x_6\left(\int_0^\infty \mathrm{d}\dot{x}\dot{x}p_{\dot{x}}\left(\dot{x}/xx_5x_6\right) \times \frac{1}{x_5x_6}p_{x_4}\left(\frac{x}{x_5x_6}\right)p_{x_5}(x_5)p_{x_6}(x_6)\right)\;. \tag{44}$$

After introducing Eqs. (27)–(29) and Eq. (38) into Eq. (44), we obtain LCR as:

$$
\begin{aligned}
N_x = & \frac{1}{\sqrt{2\pi}}\pi f_m \frac{\Omega_1^{\frac{1}{2}}}{(\kappa_1+1)^{\frac{1}{2}}} x^{2i_1+1} \frac{2(\kappa_1+1)}{\Omega_1} \sum_{i_2=0}^{\infty} \left(\frac{\kappa_1(\kappa_1+1)}{\Omega_1}\right)^{i_2} \\
& \times \frac{1}{(i_2!)^2} \frac{2(\kappa_2+1)}{\Omega_2} \sum_{i_3=0}^{\infty} \left(\frac{\kappa_2(\kappa_2+1)}{\Omega_2}\right)^{i_3} \frac{1}{(i_3!)^2} \frac{2}{\Gamma(m_3)} \\
& \times \left(\frac{m_3}{\Omega_3}\right)^{m_3} \int_0^{\infty} \mathrm{d}x_5 \int_0^{\infty} \mathrm{d}x_6 x_5^{-2i_2-1+2i_3+1} x_6^{-2i_2-1+2m_3-1} \\
& \times \mathrm{e}^{-\frac{\kappa_1+1}{\Omega_1}\frac{x^2}{x_5^2 x_6^2} - \frac{\kappa_2+1}{\Omega_2} x_5^2 - \frac{m_3}{\Omega_3} x_6^2} \\
& \times \left(1 + \frac{x^2}{x_5^4 x_6^2} \frac{\Omega_2}{\kappa_2+1} \frac{\kappa_1+1}{\Omega_1} + \frac{x^2}{x_5^2 x_6^4} \frac{\Omega_3}{m_3} \frac{\kappa_1+1}{\Omega_1}\right)^{\frac{1}{2}} . \qquad (45)
\end{aligned}
$$

Now we need to use the Laplace approximation theorem for the solution of double integrals, defined in Eqs. (18)–(20), for x_5 and x_6, and x_{50} and x_{60} as solutions [39], to solve last integrals in Eq. (45).

For this case the following is valid:

$$
\begin{aligned}
g(x_5,x_6) = & x_5^{-2i_2+2i_3} x_6^{-2i_2+2m_3-2} \\
& \times \left(1 + \frac{x^2}{x_5^4 x_6^2} \frac{\Omega_2}{\kappa_2+1} \frac{\kappa_1+1}{\Omega_1} + \frac{x^2}{x_5^2 x_6^4} \frac{\Omega_3}{m_3} \frac{\kappa_1+1}{\Omega_1}\right)^{\frac{1}{2}} , \qquad (46)
\end{aligned}
$$

$$
f(x_5,x_6) = -\frac{\kappa_1+1}{\Omega_1} \frac{x^2}{x_5^2 x_6^2} - \frac{\kappa_2+1}{\Omega_2} x_5^2 - \frac{m_3}{\Omega_3} x_6^2 , \qquad (47)
$$

$$
\frac{\partial f(x_5,x_6)}{\partial x_6} = \frac{2(\kappa_1+1)}{\Omega_1} \frac{x^2}{x_5^3 x_6^2} - \frac{2(\kappa_2+1)}{\Omega_2} x_5 , \qquad (48)
$$

$$
\frac{\partial f(x_5,x_6)}{\partial x_6} = \frac{2(\kappa_1+1)}{\Omega_1} \frac{x^2}{x_5^2 x_6^3} - \frac{2m_3}{\Omega_3} x_6 . \qquad (49)
$$

3.4. Numerical Examples and Discussion

Level crossing rate of the product of two Rician random variables and a Nakagami-m random variable is calculated and shown in the next few figures. The influence that Rician factors κ_1 and κ_2, Nakagami-m fading severity parameter m_3, Rician multipath fading powers Ω_1 and Ω_2, and Nakagami-m fading power Ω_3 exert on LCR is discussed.

LCR, normalized by f_m, depending on the signal envelope x, is presented in Fig. 4, for different values of Rician factor κ_1 and Rician fading power Ω_1. It is obvious from the picture that LCR achieves the maximum for small values of signal envelope x, and starts to decrease for higher values of x. So, it is evident that the impact of the signal envelope on LCR is bigger for small values of the signal envelope. It is also visible from this figure that LCR increases along with the increase in Rician factor κ_1 and in power Ω_1. It is known that system performance is better for smaller values of LCR.

In Fig. 5, the normalized LCR is shown versus signal envelope for different values of parameters κ_1 and Ω_2. It is possible to see from this figure that when κ_1 grows, LCR increases as well, but the increase is insignificant. On the other hand, with the increase in Ω_2, LCR increases visibly, the curves become wider and the maximums move towards higher values of the signal envelope x.

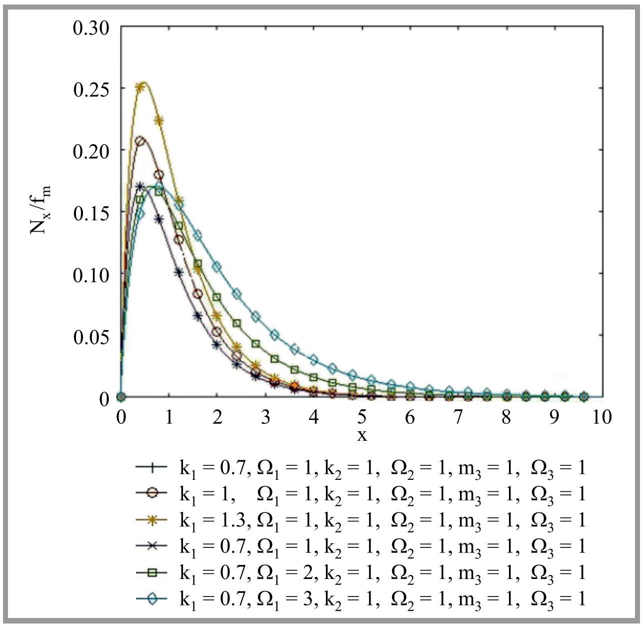


Fig. 4. LCR normalized by f_m versus signal envelope x for various values of parameters κ_1 and Ω_1.

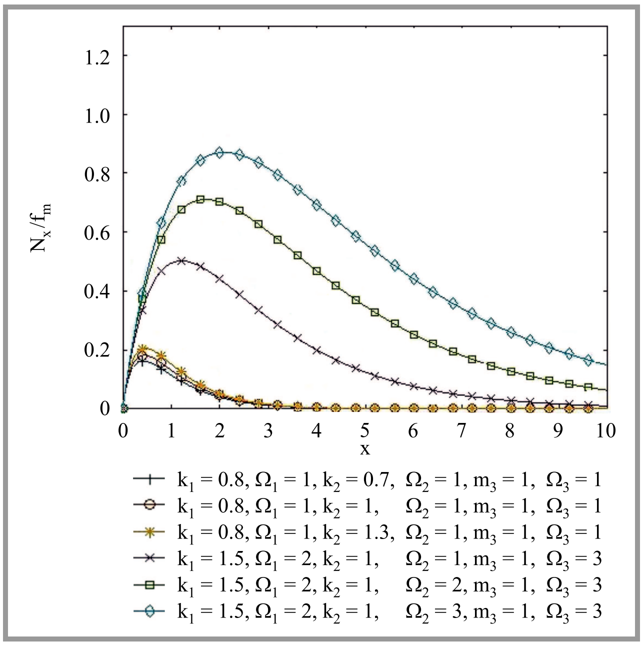


Fig. 5. LCR normalized by f_m versus signal envelope x for several values of parameters κ_2 and Ω_2.

The last figure, Fig. 6, presents the LCR, normalized by f_m, depending on signal envelope for various values of Nakagami-m fading severity parameter m_3, and Nakagami-m fading power Ω_3. It can be noticed that LCR grows with an increase in fading power Ω_3 and with a reduction in the Nakagami-m fading parameter m_3.

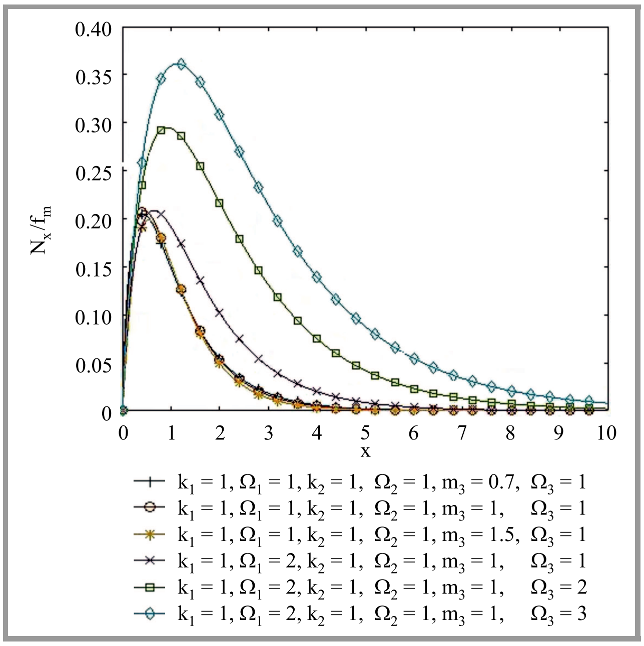


Fig. 6. LCR normalized by f_m versus signal envelope for different values of parameters m_3 and Ω_3.

4. Conclusion

The product of RVs is applied in multiple relay channels in the presence of composite fading. In this work, we focused on a wireless relay communications channel with three sections, where the product of three RVs describes the amplitude at the output of the cascaded fading channel with three sections. A closed form LCR has been calculated for that channel. The formula obtained has been checked for different values of fading and power parameters.

The results are valuable for scientists and system designers dealing with fading models for different wireless channels. It is possible to verify the proposed distribution of the products of other fading amplitudes by measuring parameters in real wireless relay channels in the presence of multipath fading, and also due to the fact that Nakagami-*m* and Rician distributions are of the general variety. By entering adequate values of fading parameters, other fading distributions in the individual sections may be obtained.

5. Acknowledgments

This work was partially supported under projects TR-33035 and III-44006 of the Ministry of Education, Science and Technological Development of Republic of Serbia.

References

[1] Y. Chen, G. K. Karagiannidis, Hao Lu, and Ning Cao, "Novel approximations to the statistics of products of independent random variables and their applications in wireless communications", *IEEE Trans. on Veh. Technol.*, vol. 61, no. 2, 2012, pp. 443–454 (doi: 10.1109/TVT.2011.2178441).

[2] S. Nadarajaha and D. K. Dey, "On the product and ratio of t random variables", *Appl. Mathem. Lett.*, vol. 19, no. 1, pp. 45–55, 2006 (doi: 10.1016/j.aml.2005.01.004).

[3] E. Mekić , N. Sekulović, M. Bandjur, M. Stefanović, and P. Spalević, "The distribution of ratio of random variable and product of two random variables and its application in performance analysis of multi-hop relaying communications over fading channels", *Przegląd Elektrotechniczny* (*Electrical Review*), vol. 88, no. 7a, pp. 133–137, 2012.

[4] D. Krstic, M. Stefanovic, V. Milenkovic, and Dj. Bandjur, "Level crossing rate of ratio of product of two α-k-μ random variables and α-k-μ random variable", *WSEAS Trans. on Commun.*, vol. 13, no. 1, pp. 622–630, 2014.

[5] D. Krstic, I. Romdhani, M. B. Y. Masadeh, S. Minic, G. Petkovic, and P. Milacic, "Level crossing rate of ratio of product of two k-u random variables and Nakagami-m random variable", *IEEE Int. Conf. on Comp. and Inform. Technol.; Ubiquitous Comput. and Commun.; Depend., Autonom. and Sec. Computi.; Perv. Intellig. and Comput.*, Liverpool, UK, 2015 (doi: 10.1109/CIT/IUCC/DASC/PICOM.2015.244).

[6] N. Zlatanov, Z. Hadzi-Velkov, and G. K. Karagiannidis, "Level crossing rate and average fade duration of the double Nakagami-*m* random process and application in MIMO keyhole fading channels", IEEE Commun. Lett., vol. 12, no. 11, pp. 822–824, 2008 (doi: 10.1109/LCOMM.2008.081058).

[7] J. D. Donahue, "Products and quotients of random variables and their applications", ARL 64-115, Aerospace Research Laboratories, Wright-Patterson Air Force Base, Ohio, The Martin Company, Denver, Colorado, July 1964 [Online]. Available: https://apps.dtic.mil/dtic/tr/fulltext/u2/603667.pdf

[8] M. D. Springer and W. E. Thompson, "The distribution of products of independent random variables", *SIAM J. on Appl. Mathem.*, vol. 14, no. 3, pp. 511–526, 1966 (doi: 10.1137/0114046).

[9] M. D. Springer and W. E. Thompson, "The distribution of products of beta, gamma and Gaussian random variables", *SIAM J. on Appl. Mathem.*, vol. 18, no 4, pp. 721–737, 1970 (doi: 10.1137/0118065).

[10] Z. A. Lomnicki, "On the distribution of products of random variables", *J. of the Royal Statist. Soc. Series B (Methodological)*, vol. 29, no. 3, pp. 513–524, 1967.

[11] A. G. Glen, L. M. Leemis, and J. H. Drew, "Computing the distribution of the product of two continuous random variables", *Comput. Statist. and Data Anal.*, vol. 44, no. 3, pp. 451–464, 2004 (doi: 10.1016/S0167-9473(02)00234-7).

[12] T. S. Glickman and F. Xu, "The distribution of the product of two triangular random variables", *Statist. & Probab. Lett.*, vol. 78, no. 16, pp. 2821–2826, 2008 (doi: 0.1016/j.spl.2008.03.031).

[13] G. K. Karagiannidis, N. C. Sagias, and P. T. Mathiopoulos, "N*Nakagami: a novel stochastic model for cascaded fading channels", *IEEE Trans. Commun.*, vol. 55, no. 8, pp. 1453–1458, 2007 (doi: 10.1109/TCOMM.2007.902497).

[14] Z. Zheng, L. Wei, J. Hamalainen, and O. Tirkkonen, "Approximation to distribution of product of random variables using orthogonal polynomials for lognormal density", *IEEE Commun. Lett.*, vol. 16, no. 12, pp. 2028–2031, 2012 (doi: 10.1109/LCOMM.2012.101712.122141).

[15] E. J. Leonardo and M. D. Yacoub, "Statistics of the product of arbitrary α-μ variates with applications", in *Proc. 25th Int. Symp. on Pers., Indoor and Mob. Radio Commun. PIMRC 2014*, Washington, DC, USA, 2014, pp. 73–76 (doi: 10.1109/PIMRC.2014.7136135).

[16] Z. Stojanac, D. Suess, and M. Kliesch, "On products of Gaussian random variables", arXiv:1711.10516 [math.PR], 2018.

[17] N. Bhargav *et al.*, "On the product of two κ-μ random variables and its application to double and composite fading channels", *IEEE Trans. on Wirel. Commun.*, vol. 17, no. 4, pp. 2457–2470, 2018 (doi: 10.1109/TWC.2018.2796562).

[18] D. H. Pavlovic *et al.*, "Statistics for ratios of Rayleigh, Rician, Nakagami-*m*, and Weibull distributed random variables", *Mathem. Problems in Engin.*, vol. 2013, Article ID 252804 (doi: 10.1155/2013/252804).

[19] M. Shakil and B. M. Golam Kibria, "On the product of Maxwell and Rice random variables", *J. of Modern Appl. Statist. Meth.*, vol. 6, no. 1, Article 19, pp. 212–218, 2007 (doi: 10.22237/jmasm/1177993080) [Online]. Available: http://digitalcommons.wayne.edu/jmasm/vol6/iss1/19

[20] K. Pearson, "The problem of the random walk", *Nature*, vol. 72, p. 318, 1905 (doi: 10.1038/072294b0).

[21] M. Nakagami, "The *m*-distribution: A general formula of intensity distribution of rapid fading", in *Statistical Methods in Radio Wave Propagation: Proceedings of a Symposium held June 18–20, 1958*, W. C. Hoffman, Ed. New York: Pergamon Press, 1960, pp. 3–36 (doi: 10.1016/b978-0-08-009306-2.50005-4).

[22] S. O. Rice, "Mathematical analysis of random noise", *Bell Syst. Technic. J.*, vol. 24, no. 1, pp. 46–156, 1945 (doi: 10.1002/j.1538-7305.1945.tb00453.x).

[23] D. Chizhik, G. J. Foschini, M. J. Gans, and R. A. Valenzuela, "Keyholes, correlations, and capacities of multielement transmit and receive antennas", *IEEE Trans. on Wirel. Commun.*, vol. 1, no. 2, pp. 361–368, 2002 (doi: 10.1109/7693.994830).

[24] T. Taniguchi, Y. Karasawa, and M. Tsuruta, "An analysis method of double fading MIMO channels including LOS environments", in *Proc. of the IEEE 19th Int. Symp. Pers., Indoor Mob. Radio Commun.*, Cannes, France, 2008, pp. 1–5 (doi: 10.1109/PIMRC.2008.4699512).

[25] D. Krstic, M. Stefanovic, and P. Nikolić, "Level crossing rate of product of Nakagami-*m* random variable, Rician random variable and Rayleigh random variable", ICTF 2018, IEICE Information and Communication Technology Forum, July 2018, Graz, Austria.

[26] D. Krstic, M. Stefanovic, M. M. B. Yaseen, S. Aljawarneh, and P. Nikolić, "Statistics of the product of three Rician random processes with application", in *Proc. 1st Int. Conf. on Data Sci., E-learn. and Inform. Syst. DATA'18*, Madrid, Spain, 2018 (doi: 10.1145/3279996.3280015).

[27] G. L. Stüber, *Principles of Mobile Communication*, 2nd ed. Norwell, MA, USA: Kluwer, 2001 (ISBN: 0792379985).

[28] P. M. Shankar, *Fading and Shadowing in Wireless Systems*. New York Dordrecht Heidelberg London: Springer, 2012 (doi: 10.1007/978-1-4614-0367-8).

[29] S. Panic, M. Stefanovic, J. Anastasov, and P. Spalevic *Fading and Interference Mitigation in Wireless Communications*. Boca Raton, USA: CRC Press, 2013 (ISBN: 9781466508415).

[30] D. Shen, Y. Cui, A. Zhang, Y. Yang, and K. Wu, "A simple simulation method for Nakagami fading channel", in *Proc. Int. Conf. on Microw. and Millim. Wave Technol.*, Chengdu, China, 2010 (doi:10.1109/icmmt.2010.5525262).

[31] A. Goldsmith, *Wireless Communications*. Stanford University, 2004 (ISBN: 9780521837163).

[32] S. O. Rice, "Statistical properties of a sine wave plus random noise", *Bell Syst. Tech. J.*, vol. 27, no. 1, pp. 109–157, 1948 (doi:10.1002/j.1538-7305.1948.tb01334.x).

[33] A. Mitić, D. Milović, M. Jakovljević, A. Panajotović, "Second order statistics of signal in Nakagami – lognormal fading channels with selection combining", in XIII Telekomunikacioni forum TELFOR 2005, Beograd, Serbia, 2005 [in Serbian].

[34] M. D. Yacoub, J. E. V. Bautista, and L. Guerra de Rezende Guedes, "On higher order statistics of the Nakagami-*m* distribution", *IEEE Trans. on Veh. Technol.*, vol. 48, no. 3, pp. 790–794, 1999 (doi: 10.1109/25.764995).

[35] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Upper Saddle River, N.J.: Prentice Hall, 2002 (ISBN: 0130422320).

[36] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*. National Bureau of Standards, 1964, reprinted Dover Publications, 1965 (ISBN: 9780486612720).

[37] J. L. Lopez and P. J. Pagola, "A simplification of the Laplace method for double integrals. Application to the second Appell function*", *Electron. Trans. on Num. Analysis*, vol. 30, pp. 224–236, 2008 (ISSN: 1068-9613).

[38] N. C. Karmakar, Ed., *Handbook of Smart Antennas for RFID Systems*. Wiley, 2010 (doi: 10.1002/9780470872178, ISSN: 1068-9613).

[39] Z. Cao and Y. D. Yao, "Definition and derivation of level crossing rate and average fade duration in an interference-limited environment", in *Proc. IEEE 54th Veh. Technol. Conf. VTC Fall 2001*, Atlantic City, N.J., USA, 2001 (doi: 10.1109/VTC.2001.956470).

Dragana S. Krstic received her B.Sc., M.Sc. and Ph.D. degrees in Electrical Engineering from the Faculty of Electronic Engineering, University of Niš, Serbia in 1990, 1998. and 2006, respectively. She has been working at the Faculty of Electronic Engineering, University of Niš, since 1990. Her fields of interest include telecommunications theory, optical communication systems, as well as wireless, mobile and satellite communication systems. As an author or co-author, she wrote about 260 scientific research papers, of which about 60 have been printed in international journals, several in national journals, and close to 140 have been referred to international symposia and conferences. Dr. Krstic has held more plenary and keynote lectures, panels and tutorials, by invitation, at international conferences and some faculties. She is also a member of several international journals and a reviewer for many. Also, she is a'member of the technical program committees for nearly a hundred conferences and a reviewer for about 120 conferences.

https://orcid.org/0000-0002-2579-3911
E-mail: dragana.krstic@elfak.ni.ac.rs
Faculty of Electronic Engineering
University of Niš
Niš, Serbia

Petar B. Nikolić graduated from the Faculty of Electronic Engineering, University of Niš, Serbia, and received his M.Sc. and Ph.D. degrees in 2008 and 2016, respectively. He is working for a company of Tigar Tyres, Pirot. His main research interest is connected with the wireless communication systems. He has written or co-authored a considerable number of papers, has published in renowned journals and conferences proceedings.

E-mail: nikpetar@gmail.com
TigarTyres
Pirot, Serbia

Danijela A. Aleksić received the B.Sc. degree in Electronics and Telecommunications Engineering from the Faculty of Electronic Engineering, University of Niš, Serbia in 2001, and M.Sc. degree in the field of electromagnetics at the Faculty of Technical Sciences in Cacak, Serbia, in 2010. Her research interests are statistical models of wave propagation in urban environments and diversity techniques for reducing the influence of fading on the performance of wireless systems. She is working at College of Applied Technical Sciences, Niš since 2002. She was a lab technician, then a teaching assistant. Currently, she works as a lecturer. Previous field of interest are sensors and transducers. She has written or co-authored of some papers, published in eminent journals and conferences proceedings.

E-mail: danijela.aleksic@vtsnis.edu.rs
College of Applied Technical Sciences
Niš, Serbia

Dragan Z. Vučković received the M.Sc. degree in the field of safety and health from the Faculty of Occupational Safety, University of Niš, Serbia in 2004, and B.Sc. degree in the field of electrical engineering and computing at the College of Applied Technical Sciences, Serbia in 2006. His research interest are storage of electronic waste and waste management on railways. He worked at Joint Stock Company for freigh trailway transport “Srbija Cargo” a.d. since 1998. He was a leading engineer for safety and health. Currently he works as a chief coordinator for work and fire protection. He has written or co-authored of some papers, published in eminent journals and conferences proceedings. He is currently on doctoral studies at the Faculty of Economics and Engineering Management, Department of Ecology, University Business Academy, Novi Sad, Serbia.

E-mail: dvdvucko@gmail.com
Faculty of Economics and Engineering Managment
University Business Academy
Novi Sad, Serbia

Mihajlo C. Stefanović received the B.Sc., M.Sc. and Ph.D. degrees in Electrical Engineering from the Faculty of Electronic Engineering, Department of Telecommunications, University of Nis, Serbia, in 1971, 1976 and 1979, respectively. His primary research interests are statistical communication theory, optical and wireless communications. He has written or coauthored a great number of journal publications. He has written five monographs, too. He was a mentor to hundreds of graduates, for dozens of master’s theses and doctoral dissertations, and many times a member of commissions for the defense of such works. Now, Dr. Stefanović is a retired professor at the Faculty of Electronic Engineering in Niš.

https://orcid.org/0000-0002-8063-8981
E-mail: mihajlo.stefanovic@elfak.ni.ac.rs
Faculty of Electronic Engineering
University of Niš
Niš, Serbia

Paper

Enhancement of Ground-to-Aircraft Communication Using Audio Watermarking

Przemysław Dymarski

Faculty of Electronics and Information Technology, Warsaw University of Technology, Warsaw, Poland

https://doi.org/10.26636/jtit.2019.128418

Abstract—This paper presents research on improving the intelligibility of spoken messages transmitted to aircraft from a ground station. The proposed solution is based on the selective calling (SELCAL) system and the audio watermarking technique. The most important elements of a spoken message (commands, numerical values) are transmitted as a watermark embedded in the speech signal and are displayed to the cockpit crew. The synchronization signal is embedded in SELCAL duo-tones. The proposed system is resistant to resampling and channel noise (at SNR $>$ 25 dB).

Keywords*—*audio watermarking, aviation radio services, SELCAL.

1. Introduction

Voice communication between ground and aircraft stations is based on analog DSB–AM modulation and relies on the 117.975–137.000 MHz band. In order to ensure proper understanding of the messages, special phraseology standardized by International Civil Aviation Organization (ICAO) is used [1], [2]. It consists of a series of keywords (e.g. acknowledge, affirm, cleared, confirm, over, report, roger), requires the use of a special spelling system, both with regard to letters (A – alpha, B – bravo, C – Charlie, D – delta, etc.) and digits (4 – fower, 9 – niner), pronunciation of numbers (each digit is uttered separately, but such words as “thousand”, “hundred” and “decimal” are allowed). Special scenarios are used to increase intelligibility: “read back” – repeat this message back to me exactly as received, “say again” – repeat the entire transmission or a portion of your last transmission, “speak slower” – reduce your rate of speech, “words twice” – every word, or group of words, in this message will be pronounced twice. Nevertheless, some messages are still misunderstood, particularly by pilots having problems with English. Graphical representation of the most important elements of the message (e.g. numerical flight parameter values, such as flight level, heading, runaway number) would facilitate comprehension of messages sent by the ground station. This requires the transmission of digital information accompanying voice messages.

How to transmit such digital information? The Aircraft Communication Addressing and Reporting System (ACARS) is a tool that is commonly used for the transmission of short burst data (SBD) using VHF or satellite links [3]. A transmission speed of 2400 bps is sufficient to send weather reports and additional information concerning the flight. However, it is not a real time communication link, as delivery of messages is delayed by about 5–20 seconds if a satellite link and the SBD protocol are used [3]. The Controller Pilot Data Link Communications (CPDLC) system [4] is more suitable for Air Traffic Control (ATC) applications. It is used for non-time-critical communications between aircraft and ground. Similarly to ACARS, a digital VHF radio link is used that is independent of the analog legacy system. CPDLC is implemented in some airports in the USA and Europe (e.g. Maastricht). It is useful in relieving congestion of the analog speech communications system, but it will not replace it, because of its latency. Therefore, digital data should be transmitted along with voice message, using the same link.

Two solutions may be applied here: transmission of the data burst before or after the voice message, or embedding data in the speech signal using watermarking techniques. The sending data bursts may be disturbing to cockpit crews of other aircraft. Ground stations use the same channel to establish voice communications with a number of aircraft, and flight crews continuously monitor the frequency awaiting radio communications targeted specifically for their flight. Therefore, the use of audio watermarking techniques would be a better solution. Due to the short duration of a typical voice message (several seconds) and a low bit rate of the watermark transmission (in this case: 20 bps), only small data packs may be transmitted. The watermark transmission proposed will make it possible to send short digital messages, such as FL100 (flight level one zero zero), HEAD080 (heading zero eight zero) or RUN27 (runway two seven).

The watermark transmission system proposed may be easily integrated with the commonly used selective calling (SELCAL) system [5], [6]. A traditional voice callout (e.g. “LOT 245”) is replaced with a special SELCAL code consisting of 4 tones and attributed to a specific aircraft using this system. In fact, two duo-tones are transmitted, each with the duration of about 1 second. The aircraft crew relying on SELCAL does not have to maintain a listening watch. The reception of a proper code activates the cockpit notification system (a lamp, a bell or a chime). Then, a cockpit crew member responds with a full radio call sign (e.g. "LOT 245") a communication with the ground station begins by uttering the "go ahead" message. SELCAL is quite popular – 10920 codes have been assigned. At present, the system is fully saturated (duplicate codes start to appear), and its further extension is planned [6]. In the proposed watermark transmission system, SELCAL pulses are used for synchronization of digital transmissions. The task is accomplished by adding a third tone to both duo-tones. Its frequency is lower than the frequencies of all SELCAL tones, so the SELCAL system itself remains unaffected.

A solution that is an alternative to SELCAL was proposed in [7]. A 24-bit aircraft identifier was to be transmitted as a watermark embedded in the speech signal. However, this idea has not been implemented in practice.

The problem of misunderstandings in controller – pilot exchanges is the subject of extensive research, with some solutions including automatic speech recognition [8]. In this paper, a simple solution is proposed, fulfilling the following requirements:

- compatibility with the existing analog voice communications system. Watermarks should not degrade the quality of transmitted speech signals transmitted;
- compatibility with SELCAL system. SELCAL codes should be detected and no other non-speech signals should appear;
- no surplus tasks for cockpit crews, except for reading the information displayed. Messages typed by cockpit crews resulted in latency in the CPDLC system [4]. In the proposed system, digital data is sent from the ground to the aircraft only. The task of typing the data accompanying the voice message is the responsibility of the controller;
- safety and reliability: inconsistent digital data (parity check failure, atypical syntax or semantics) is not displayed;
- resistance to channel noise generated at the AM receiver's output;
- resistance to resampling. Due to the analog nature of the transmission, the watermarked signal should be resampled on the receiver side, but the sampling frequency differs, by several dozen Hz, from that used on the transmitter side. This leads to desynchronization of the watermark transmission.

Watermarking algorithms described in this paper are partially based on a solution proposed for steganography in VoIP transmissions [9]. However, they are thoroughly modified to deal with short messages and with the resampling of watermarked speech.

The remaining sections of this paper are organized as follows. In Section 2 the syntax of digital data and the manner in which it is embedded in the accompanying voice message are presented. In Section 3 the watermark synthesis and detection algorithms are described. In Section 4 synchronization problems are discussed, particularly those concerned with sampling frequency offset estimation and correction. Section 5 is devoted to testing resistance to channel noise and resampling. A short summary concludes in Section 6.

2. Embedding Digital Messages in their Spoken Counterparts

The watermark transmission system proposed should enhance the intelligibility of most important keywords and parameters. It should be noted that a watermark transmission is relatively slow, so only abbreviated forms of ground-

Table 1
Examples of messages transmitted in an abbreviated form

Spoken	Transmitted as a watermark	Displayed to the pilot
Flight level niner five	FL95	Flight level 95
Heading one one zero	HD110	Heading 110
Wind two zero zero degrees two five knots	WIND200D25K	Wind 200 deg 25 knots
Cloud base two thousand two hundred	CB2200	Cloud base 2200
Visibility seven hundred	VIS700	Visibility 700
Runway visual range six hundred	RVR600	Runway visual range 600
Altimeter setting one thousand	QNH1000	QNH 1000
Report level	RLEV	Report level
Climb flight level seven zero	CL70	Climb 70
Descend flight level six zero	DS60	Descend 60
Cleared for take off	CLETAO	Cleared for take off
Cancel take off	CANCTAO	Cancel take off
Runway two seven	RUN27	Runway 27

to-aircraft messages may be embedded in the speech signal. The watermark bit rate used in the system proposed equals 20 bits per second, so up to 2.5 ASCII codes may be transmitted in one second. Generally, such as rate is sufficient and the duration of the watermark is not longer than that of the corresponding spoken message. Some examples of spoken messages along with their abbreviated and displayed counterparts are given in Table 1.

The encoding of abbreviated messages in a bit stream is presented in Fig. 1. It starts with a preamble consisting of 8 bits – the 01010101 pattern was selected due to its favorable synchronization properties. 7-bit ASCII codes are used to encode the message. They are extended to 8 bits due to parity checks. At the end, another series of at least 8 bits appears, according to the same pattern. This supplement continues to be generated until the end of the spoken message.

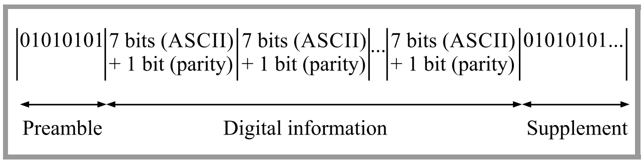

01010101
7 bits (ASCII) + 1 bit (parity)
7 bits (ASCII) + 1 bit (parity)
...
7 bits (ASCII) + 1 bit (parity)
01010101...
Preamble
Digital information
Supplement

Fig. 1. Bit stream representing the abbreviated message.

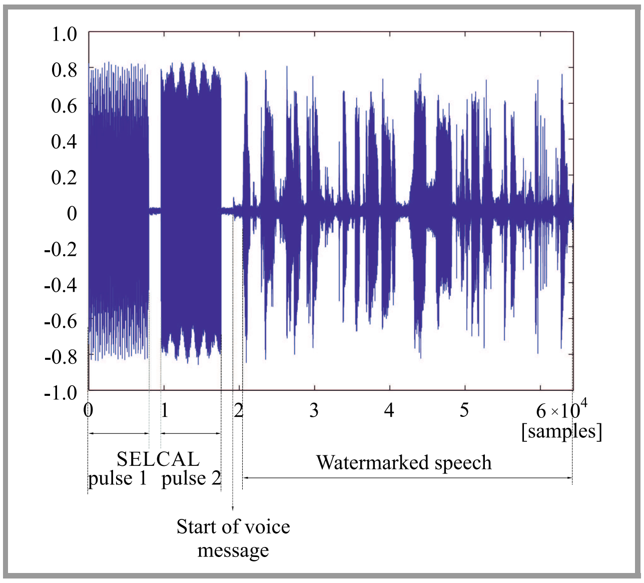


Fig. 2. Waveform of a watermarked voice message with SELCAL pulses – short bit stream.

The method of embedding this bit stream in the spoken message is explained in Fig. 2. The transmission starts with two SELCAL pulses with the duration of 1 s, each containing two tones. These tones form a selective call code, identifying a given aircraft. The basic SELCAL system uses 16 tone frequencies, from 312.6 Hz to 1479.1 Hz. Speech signal transmissions may start much later, but in Fig. 2, the time interval between SELCAL pulses and the voice message is shortened for simulation purposes. The watermark containing digital information commences when speech amplitude becomes greater than a preset threshold. It continues until the end of the voice message. After the information bits have been sent, the 010101... sequence is transmitted until the end of the voice message.

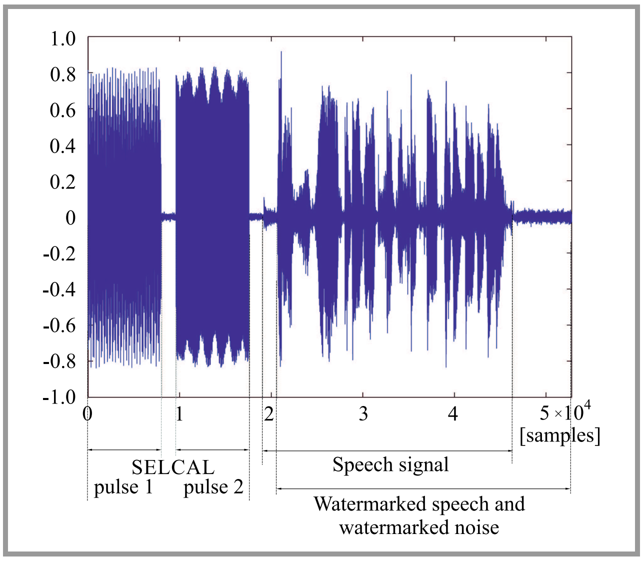


Fig. 3. Waveform of a watermarked voice message with SELCAL pulses – long bit stream.

If digital information is longer than the spoken message, then low amplitude noise is appended to the speech signal. The watermark is embedded in the speech and accompanying noise. This is shown in Fig. 3. If the energy of the speech signal drops to a preset threshold (e.g. intervals between spoken words), noise is added in order to maintain the required level of the watermarked signal.

3. Watermark Embedding and Detection

Watermark embedding and decoding algorithms proposed for ground-to-aircraft communications are partially based on steganographic algorithms intended for VoIP communications systems [9]. Watermarking in the frequency domain is applied, i.e. spectral analysis is performed to detect the watermark. This approach was also applied in watermarking of wideband audio [10]. Frequency domain watermarking yields the transmission system robust to imprecise synchronization [10], [11].

For embedding one bit of digital information, two windows with the length of $N = 200$ samples are used at the sampling frequency of 8000 Hz. This yields the bit rate equal to 20 bps. In each window the amplitude and phase spectrum of the windowed audio signal $\overline{x}$ (speech or appended noise) are calculated:

$$\begin{aligned} \overline{X} &= DFT\,[\text{Hanning}(\overline{x})]\ , \\ |\overline{X}| &= ABS(\overline{X}) \qquad (1) \\ \overline{\Phi} &= \overline{X}./|\overline{X}|\ , \end{aligned}$$

where the absolute value calculations (ABS) and divisions (./) are element-wise operations, $|\overline{X}|$ and $\overline{\Phi}$ are vectors.

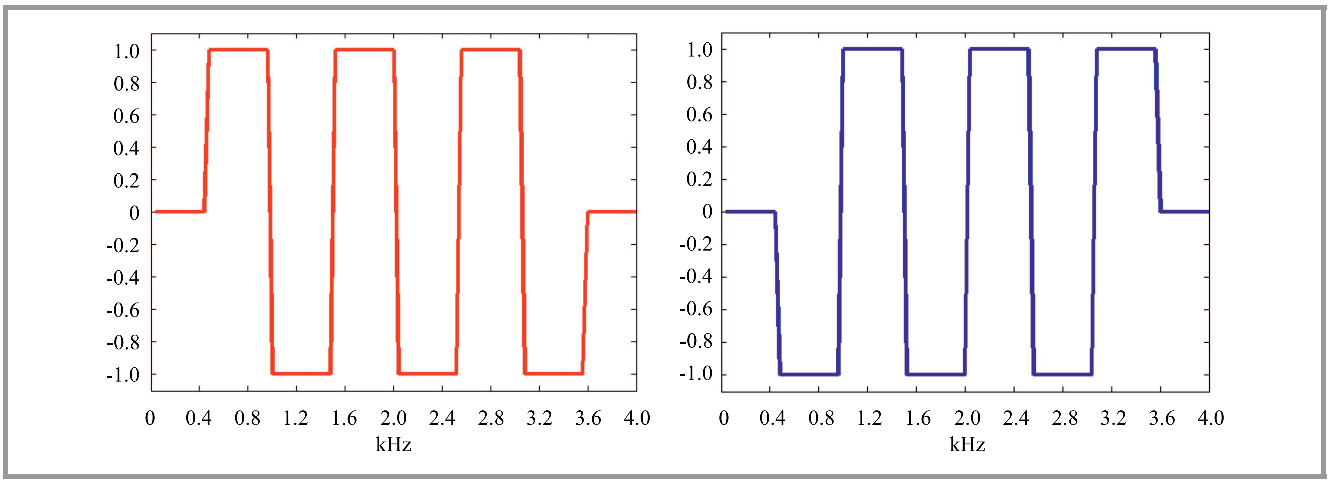


Fig. 4. Two patterns (data transmission symbols w and $-w$) used for spectrum modulation: 1 – increase, (-1) – decrease, 0 – no change.

Next, the selected spectral components (here, within the frequency range of 0.5–3.5 kHz) are modulated – their amplitudes are increased or decreased. Low frequencies (0–0.5 kHz) and high frequencies (3.5–4 kHz) are not modified, because of potential attenuation in the DSB-AM transmission. Modulation is performed in 6 sub-bands using two opposite polarity patterns (Fig. 4).

To reduce the influence of strong spectral peaks (formants) of speech signal on watermark detection, a distinctive type of differential coding is applied. Two windows (subframes) are used to transmit a single bit. For a logical "1", the w pattern (red pattern in Fig. 4) is used in the first window, and the $(-w)$ pattern, i.e. the blue pattern in Fig. 4, is used in the second window. For a logical "0" - the situation is reversed. This increases the difference between modified spectrums, but strong spectral components of the speech signal are attenuated.

In order to maintain good quality of watermarked speech, modification of spectral components should not exceed the masking threshold. A simplified algorithm used to compute the masking curve is applied. It is based on the perceptual filtering concept that is widely used in speech compression. The masking curve is the frequency response of the attenuated IIR predictive filter:

$$M(z) = \frac{\alpha}{1+\sum_{i=1}^{10} a_i \gamma^i z^{-i}} , \tag{2}$$

where a_1, a_2, …, a_{10} are prediction coefficients calculated for $N = 200$ samples of the speech signal, $\gamma = 0.95$ is the attenuation coefficient, α is the offset, influencing watermark strength and quality of watermarked speech. An example of a masking curve is shown in Fig. 5. Modifications of speech spectrum amplitudes should not exceed the masking threshold. These modifications (amplifications or attenuations of spectral components) are performed only in the frequency range of 0.5–3.5 kHz (Fig. 4). Moreover, due to the simplified method of calculating the masking curve, amplifications are restricted to triple values of speech spectrum amplitudes, and attenuations to 0.3 of these amplitudes.

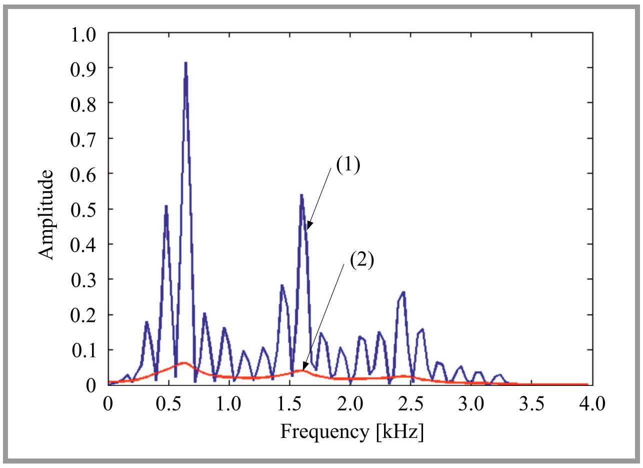


Fig. 5. Amplitude spectrum of a window of speech signal (1) and the estimated masking curve (2).

The spectrum of the watermark is obtained by subtracting the original amplitude spectrum $|\overline{X}|$ from the modified amplitude spectrum $|\overline{X}|^*$ and by applying the phase spectrum $\overline{\Phi}$ of windowed speech from Eq. (1):

$$\overline{V} = \left(|\overline{X}|^* - |\overline{X}|\right) .* \overline{\Phi} , \tag{3}$$

where $(.*)$ denotes element-wise multiplications.

Then, the time domain of the watermark is calculated using inverse DFT. In order to suppress discontinuities at the edges of the windows, a trapezoidal window is applied in the time domain:

$$\overline{v} = \text{Trapezoid}\left[\text{IDFT}(\overline{V})\right] , \tag{4}$$

Then, the watermark is added to speech or noise (if noise is appended to a short speech phrase):

$$\overline{y} = \overline{x} + \overline{v} , \tag{5}$$

Reception of the hidden bit is based on correlation, as previously proposed for speech and audio watermarking – Fig. 6 [9], [10].

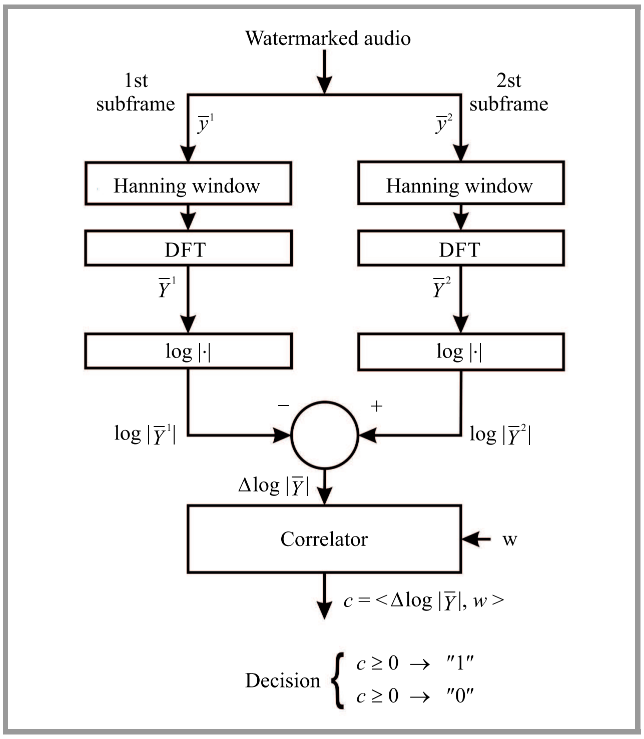


Fig. 6. Reception of a single symbol $\overline{y}$, $<\cdot>$ denotes correlation (inner product).

The use of logarithms in the frequency domain requires some explanation. Let us assume that in both windows (subframes, each one with the duration of $N = 200$ samples) the amplitude spectrum of speech is almost the same: $|\overline{X}|^1 \approx |\overline{X}|^2 \approx |\overline{X}|$. The masking curve $\overline{M}$ (frequency response of predictive filter $M(z)$) is a smoothed and attenuated copy of the signal spectrum $|\overline{X}|$. If $\gamma \rightarrow 1$ and the number of prediction coefficients are high, then $\overline{M} \rightarrow \alpha|\overline{X}|$. Watermarking consists in adding or subtracting components of $\overline{M}$ to/from components of $|\overline{X}|$:

$$|\overline{Y}| = |\overline{X}| \pm \overline{M} \approx |\overline{X}| \pm \alpha|\overline{X}| = |\overline{X}|(1 \pm \alpha)\ . \qquad (6)$$

The subtraction of watermarked spectrums of two subframes yields:

$$\Delta|\overline{Y}| = |\overline{Y}|^1 - |\overline{Y}|^2 \approx |\overline{X}|(1 \pm \alpha) - |\overline{X}|(1 \mp \alpha) = \pm 2\alpha|\overline{X}|\ . \qquad (7)$$

Due to the great dynamic range of speech spectrum, the $\Delta|\overline{Y}|$ function is weakly correlated with the pattern w and correlation receiver yields frequent errors. Moreover, only a small part of the signal spectrum influences the decision-making process – Fig. 7.

Using the log spectrum for correlation computations, yields:

$$\log|\overline{Y}| \approx \log\left[|\overline{X}|(1 \pm \alpha)\right] = \log|\overline{X}| + \log(1 \pm \alpha)\ . \qquad (8)$$

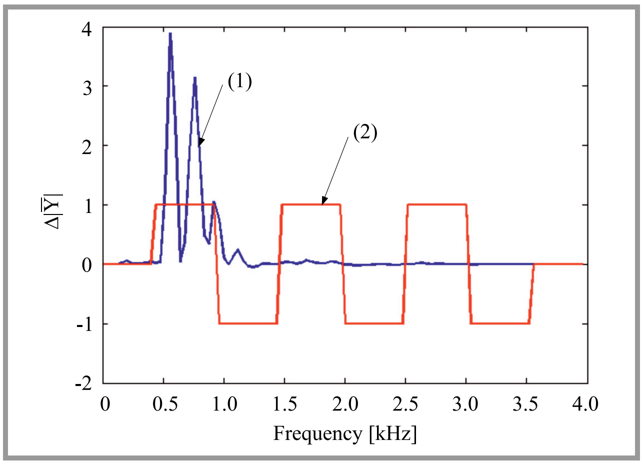


Fig. 7. Correlation computation relying on a linear spectrum (log operation skipped): $\Delta|\overline{Y}|$ – blue (1), w – red (2).

The subtraction of log spectrums of two subframes yields:

$$\Delta\log|\overline{Y}| = \log|\overline{Y}|^1 - \log|\overline{Y}|^2 \approx \log(1 \pm \alpha) - \log(1 \mp \alpha)\ . \qquad (9)$$

It needs to be noted that there is no influence of the speech spectrum $|\overline{X}|$ on the decision algorithm and that $\Delta\log|\overline{Y}|$ should be flat within each subband. In a real situation, it is not exactly like that, because the condition $\overline{M} \rightarrow \alpha|\overline{X}|$ is not fulfilled. Nevertheless, $\Delta\log|\overline{Y}|$ is strongly correlated with the w or $(-w)$ pattern, depending on the logical value of the bit transmitted – Fig. 8.

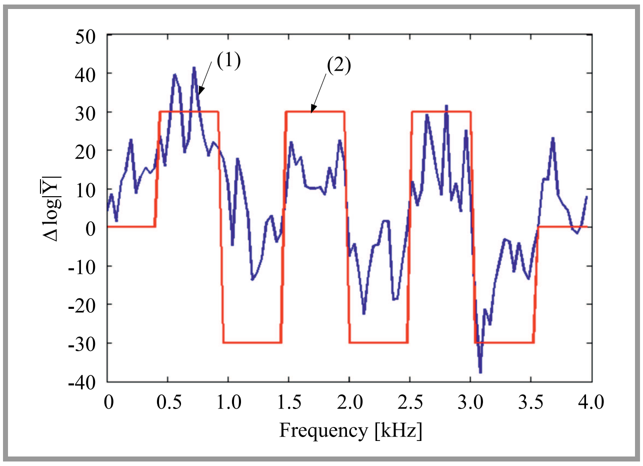


Fig. 8. Correlation computation using log spectrum: $\Delta\log|\overline{Y}|$ – blue (1), w – red (2).

This allows to detect the bit stream presented in Fig. 1. The decoder starts at the beginning of the ground-to-aircraft transmission or after the SELCAL pulses. The digital watermark transmission starts later, so many random bits may be detected before the preamble. Nevertheless, the entire bit stream received is analyzed for the positions of ASCII codes. There are eight possible segmentation methods: starting from the first, second, ..., then eighth bit received. Every time bytes are extracted, the parity test is performed and the number of failures is noted. The lowest value of parity errors indicates a proper segmentation manner. Then the preamble is found and the digital message is decoded. All bytes carrying ASCII codes should fulfill the requirements of the parity test. If not, the whole message is classified as uncertain and is not displayed.

Additional tests may be performed, based on prior knowledge of syntax and semantics of the transmitted messages. For example, messages presented in Table 1 contain capital letters and numbers only. Detection of other characters indicates a transmission error. Such a message will not be displayed.

4. Bit Synchronization and Sampling Frequency Offset Correction

The watermark reception algorithm described in the previous section requires bit synchronization. Time intervals lasting 50 ms (400 samples) should be localized in the time domain. The synchronization algorithm is based on the observation that the correlation $c = <\Delta \log|\overline{Y}|,\ w>$ (Fig. 8) attains the maximum absolute value if both subframes are correctly localized. Therefore, the reception algorithm (Fig. 6) is executed many times with a small shift (here, 10 samples). Each time, the absolute value of correlation is noted (Fig. 9). Note the maximum values every $40 \times 10 = 400$ samples. They correspond to the true positions of windows containing watermarked bits. In the middle of each window the watermarking pattern is changed (from w to $(-w)$ or vice versa) and the difference of log spectrums $\Delta \log|\overline{Y}|$ is maximized. If the same logical value is transmitted in neighboring windows, the watermarking pattern is changed at the edge. That is why additional maximum values in between the true ones are observed. If the bit sequence of 010101… is transmitted, no additional maximum values are observed. That is why these sequences are used as the preamble and the supplement for the transmitted data.

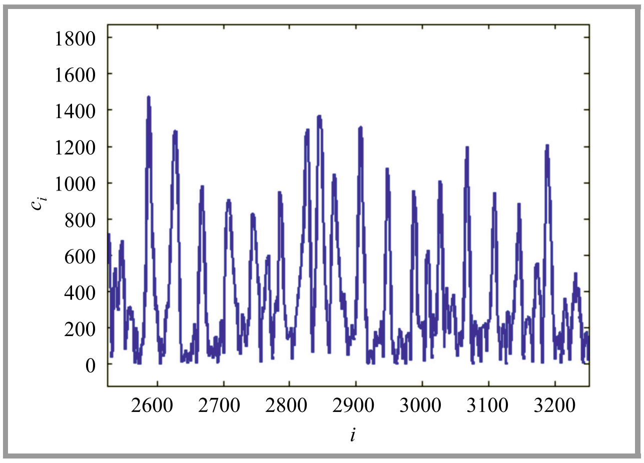


Fig. 9. Series of correlations $|c_i|$ calculated with a shift of 10 samples.

To identify the positions of bit transmission windows, the correlations (Fig. 9) are summed up with the shift equal to window duration (40 times ten samples). This is performed 40 times, starting from different positions:

$$\begin{array}{ll} C_1 & = |c_1| + |c_{41}| + |c_{81}| + \ldots \\ C_2 & = |c_2| + |c_{42}| + |c_{82}| + \ldots \\ \ldots & \\ C_{40} & = |c_{40}| + |c_{80}| + |c_{120}| + \ldots \end{array} . \qquad (10)$$

An example of these sums of correlations is presented in Fig. 10. The maximum value indicates the position of data transmitting windows, the secondary maximum is also visible, pointing to the middle of the windows.

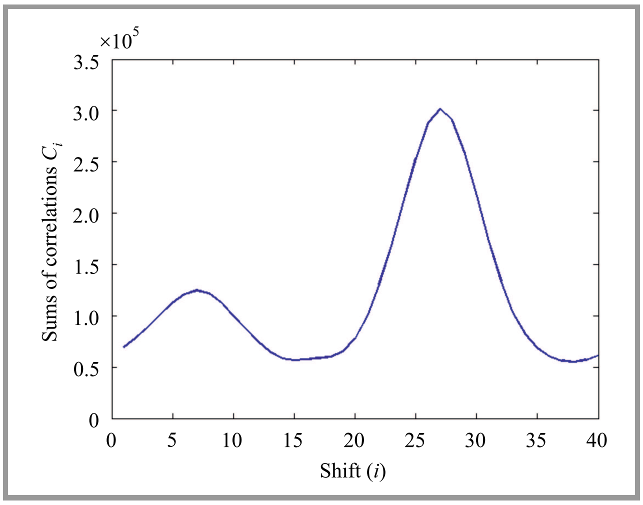


Fig. 10. Sums of correlations $C_1, C_2, \ldots, C_{40}$.

Satisfactory performance of the bit synchronization algorithm is obtained if each window contains exactly $2N = 400$ samples. Unfortunately, it is not the case because the watermarked signal is transmitted using an analog DSB-AM communication link and is then resampled at the receiver side. Sampling frequencies used for watermark embedding and detection are not synchronized and a difference of some tens of Hz may be expected. Thus the sampling frequency offset should be estimated and the number of samples in a window (real number T) should be calculated. Then, the true position of the first window is found by maximizing the modified sums of correlations:

$$\begin{array}{l} i_{\max} = \arg\max(C_i)\ , \\ C_i = |c_i| + |c_{i+\text{round}(\frac{T}{10})}| + c_{i+\text{round}(\frac{2T}{10})}| + \ldots\ , \end{array} \qquad (11)$$

where round denotes rounding to the nearest integer. Thus the first window starts at $10 i_{\max}$ and the others at $10 i_{\max} + \text{round}(nT)$.

Now it begs the question of how to estimate the sampling frequency f_s' at the receiver and the number of samples in the window T.

The first approach is based on a series of correlations $|c_i|$ (Fig. 9). Its quasi-period $\frac{T}{10}$ may be estimated with Fourier analysis. In Fig. 11 the absolute values of DFT coefficients of a correlations series $|c_i|$ are shown. The position of the first harmonic indicates the inverse of the quasi period. In order to increase resolution, zeros were appended to the correlations series. Therefore, big values of DFT lags appear in Fig. 11.

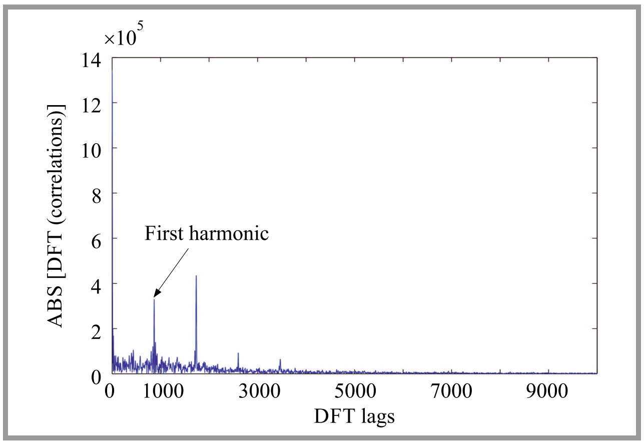


Fig. 11. Fourier analysis of a series of correlations.

The second approach to sampling frequency offset estimation consists in transmitting tones. This technique is widely used in OFDM systems [12]. It has been also applied in audio watermarking systems [13], [14]. This approach would be particularly interesting if the SELCAL system is used. SELCAL pulses consist of two tones, so there is no problem if the third tone is added and used for sampling frequency estimation. Its frequency should be out of the band used for SELCAL tones (312.6–1479.1 Hz). Thus, the frequency of $f_p = 8000/28 \approx 285.71$ Hz is selected. One period contains exactly 28 samples. The amplitude of this tone is 6 dB below that of SELCAL tones. Two SELCAL pulses are used for sampling frequency estimation.

This additional tone, $A\cos(2\pi f_p t + \varphi_0)$, is synthesized at the transmitter side as a series of samples $A\cos(2\pi f_p \frac{n}{f_s} + \varphi_0)$, where $f_s = 8000$ Hz. Its period is $P = \frac{f_s}{f_p} = 28$ samples. At the receiver side, this tone is sampled at the sampling frequency of f'_s: $A\cos(2\pi f_p \frac{n}{f'_s} + \varphi_0)$. For estimation of f'_s the maximum likelihood estimator may be used, maximizing absolute value of the correlation of the received tone with $\mathrm{e}^{\mathrm{j}2\pi f_p \frac{n}{f'_s}}$ [15]. This estimator is optimal in the Cramer-Rao sense, but it requires many correlation calculations for all tested values of f'_s. The algorithm used in [12]–[14] is suboptimal but less complex, because the correlation is computed only once, in windows of short duration (here, in windows containing $P = 28$ samples). For correlation computation, one period of the complex signal sampled at $f_s = 8000$ Hz is used: $\mathrm{e}^{\mathrm{j}2\pi f_p \frac{n}{f_s}}$, $n = 0, 1, \ldots, P-1$. For the k-th window this correlation equals:

$$r(k) = A \sum_{n-(k-1)P}^{kP-1} \cos\left(2\pi f_p \frac{n}{f'_s} + \varphi_0\right) \mathrm{e}^{\mathrm{j}2\pi n \frac{f_p}{f_s}} . \quad (12)$$

If $f'_s = f_s$ and noise and the other distortions are absent, then the complex correlations are equal. If $f'_s \neq f_s$, then the phase shift $\Delta\varphi$ appears at the end of each window and is accumulated. At the end of the first window the phase shift equals:

$$\Delta\varphi = 2\pi f_p \frac{P}{f_s} + \varphi_0 - 2\pi f_p \frac{P}{f'_s} - \varphi_0 = 2\pi - 2\pi \frac{f_s}{f'_s} . \quad (13)$$

Then it is cumulated: $\Delta\varphi(k) = k\Delta\varphi$.

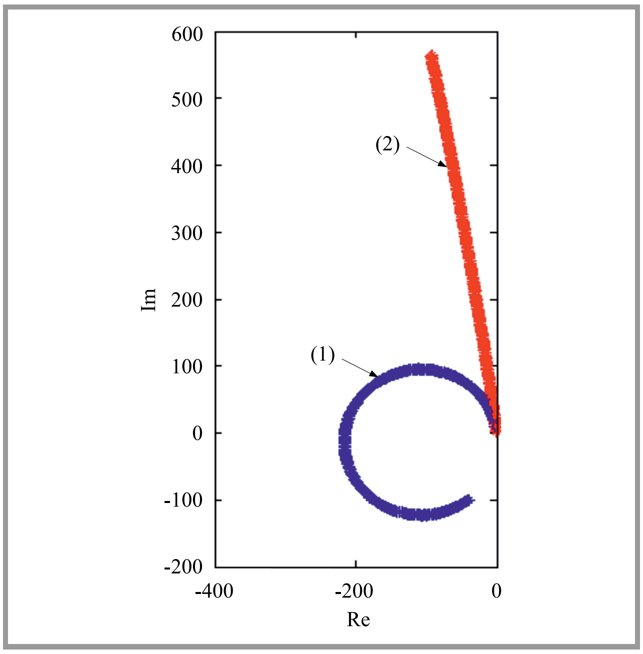


Fig. 12. Cumulated complex correlations before (blue – 1) and after (red – 2) correction.

In Fig. 12 sums of complex correlations $R(K) = \sum_{k=1}^{K} r(k)$ are shown ($f'_s - f_s = 25$ Hz). An increasing phase shift may be observed. Compensation of the phase shift makes all correlations equal and the corresponding sum yields its maximum absolute value: $R'(K) = \sum_{k=1}^{K} r(k)\mathrm{e}^{-\mathrm{j}\Delta\varphi(k)}$. This suggests an algorithm for phase shift estimation:

$$\Delta\phi = \arg\max_{v} \left| \sum_{k=1}^{K_{\max}} r(k)\,\mathrm{e}^{-\mathrm{j}kv} \right| . \quad (14)$$

Having $\Delta\varphi$ we may calculate the sampling frequency $f'_s = \frac{2\pi f_s}{2\pi - \Delta\varphi}$ and the number of samples within a bit transmitting window: $T = \frac{2Nf'_s}{f_s}$.

5. Testing

Tests were performed with Matlab, using seven phrases of duration between 3 and 10 s, recorded during a listening watch at the Warsaw Chopin Airport. Only ground-to-airplane communications were recorded. Phrases were of good quality, noise level was more than 30 dB below the speech level. Two SELCAL pulses were inserted before the speech phrase (Figs. 2–3).

To simulate an analog communication channel, pseudorandom noise was added, at SNR = 5–30 dB. At the receiver the incoming signal was resampled at the sampling frequency of 7970–8030 Hz.

The bit stream included a 8-bit preamble, 8 bytes (7-bit ASCII codes with 1 bit for parity control) and a supplement containing at least 8 bits (Fig. 1). Simulations were repeated 3–10 times to improve the accuracy of bit error rate (BER) estimation.

The quality of watermarked speech was evaluated using the PESQ algorithm [16]. The mean opinion score (MOS) and listening quality MOS (MOS-LQO) were measured before the addition of channel noise. The results are shown in Fig. 13. Speech quality depends on watermark strength α – Eq. (2). Watermark attenuation of 3 dB ($\alpha = 0.707$) yields a MOS improvement of about 0.2. Mean MOS-LQO value equals 3.82 for a stronger watermark ($\alpha = 1$) and 4.03 for a weaker watermark ($\alpha = 0.707$). In both cases speech quality is judged as good.

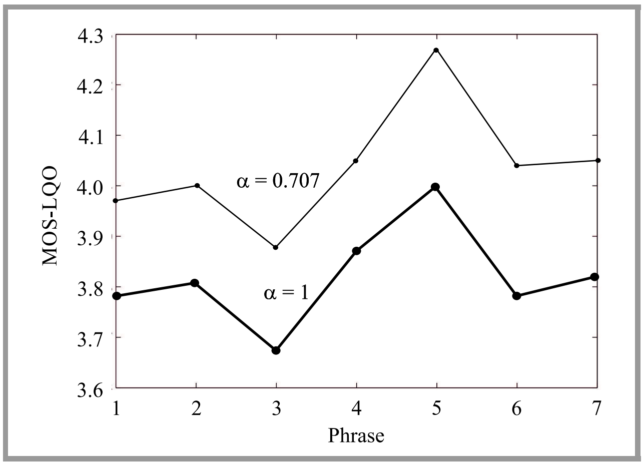


Fig. 13. MOS-LQO for 7 watermarked speech phrases.

Preliminary tests of the bit detection algorithm (Fig. 6) were performed to check the robustness of this algorithm and its resistance to resampling and window shift. Without sampling frequency offset correction, the transmission and the reception of windows cannot be aligned and BER increases. Due to the short duration of watermarked speech, tolerance to sampling frequency offset of up to 10 Hz is obtained (Fig. 14). Therefore, the sampling frequency estimation error should not exceed 10 Hz.

Then, the robustness to window shift was tested. The sampling frequency at the receiver was set to 8 kHz and bit synchronization was blocked. The increase in BER started at a shift value equal to 30 samples (Fig. 15). The bit synchronization algorithm proposed localizes windows with the position error of up to 10 samples (Figs. 9–10), which seems to be sufficient.

Then, the comparison of two sampling frequency estimation algorithms was made. One phrase of speech signal with

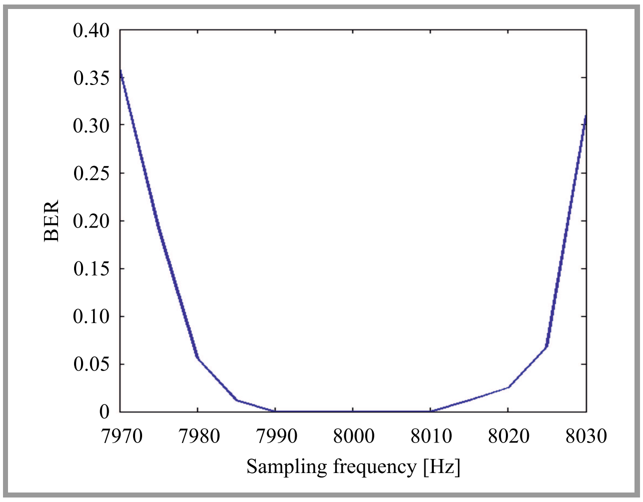


Fig. 14. BER as a function of sampling frequency (without frequency offset correction).

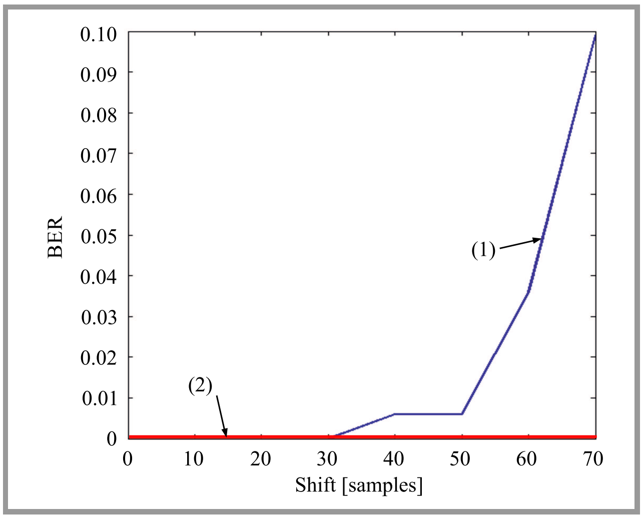


Fig. 15. BER as a function of window shift: without bit synchronization (blue – 1) and with bit synchronization (red – 2).

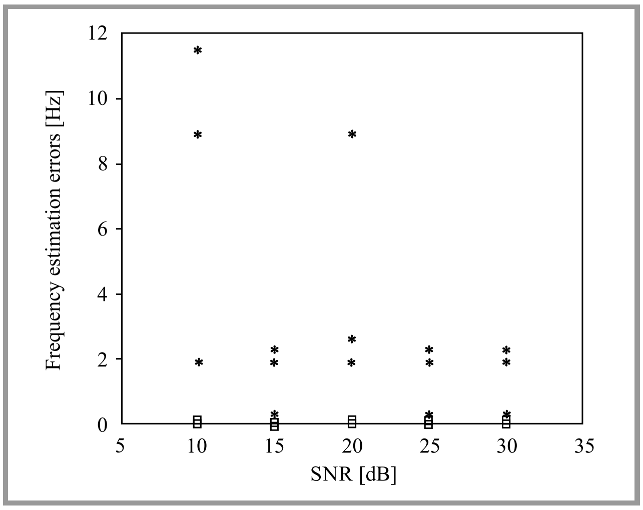


Fig. 16. Observation of sampling frequency estimation errors: DFT of a series of correlations (∗) and adding a tone to SELCAL signals (□).

SELCAL pulses was generated (Fig. 2), then channel noise was added (SNR from 10 to 30 dB). Three sampling frequency values were tested: 8000 Hz (no change of sampling frequency), 7975 Hz (sampling frequency offset −25 Hz) and 8025 Hz (sampling frequency offset +25 Hz). Errors of sampling frequency estimation are shown in Fig. 16. Frequency estimation based on DFT of a series of correlations (Fig. 9, Fig. 11) was less accurate than frequency estimation based on an additional tone added to SECAL pulses (Fig. 12). The bit detection algorithm is robust to a sampling frequency mismatch of up to 10 Hz (Fig. 14), so both algorithms may be applied.

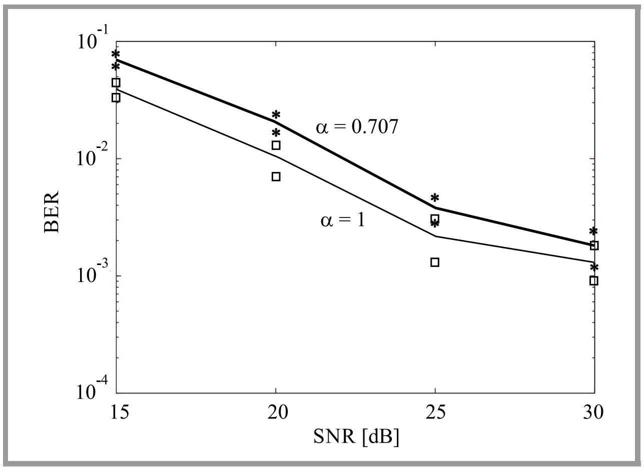


Fig. 17. BER as a function of SNR at the output of the AM receiver (confidence intervals marked with asterisks and squares).

Finally, robustness to channel noise was tested using 7 speech phrases, 4 SNR values and 2 watermark strength coefficients: $\alpha = 1$ (full strength) and $\alpha = 0.707$ (watermark attenuation of 3 dB). Each simulation was repeated 10 times using different noise waveforms, in order to reduce confidence intervals. The results (Fig. 17) show that BER approaches 0.001 at SNR = 30 dB. A typical message does not exceed 100 bits, so it can be received without any error at a probability greater than 0.9.

6. Conclusions

The audio watermarking system proposed may be helpful in increasing comprehension of voice commands transmitted from ground to aircraft using an analog communication link. Digital information is embedded in the speech signal and may be displayed in the cockpit. A relatively low bit rate of 20 bps is sufficient to encode keywords and flight parameters. The algorithms proposed meet the requirements specified in the introduction, namely:

- compatibility with existing analog voice communications systems. Digital information embedded in the speech signal does not degrade its quality. MOS values measured with the PESQ algorithm [16] show a good speech quality. The mean MOS-LQO value equals 3.82 for a stronger watermark and 4.03 for a weaker watermark (attenuation of 3 dB). No other non-speech signals appear, like in a modem-based approach [17];
- compatibility with SELCAL system. SELCAL pulses contain duo-tones of frequencies from 312.6 to 1479.1 Hz. In the proposed system the third tone is added outside of this range, at frequency of $f_p \approx$ 285.71 Hz. It is used for sampling frequency estimation does not affect detection of duotones;
- no surplus charge for cockpit crew and low latency. Digital messages are transmitted from the ground to the aircraft, the cockpit crew is only required to observe a display. The decoding of the message is commenced immediately after reception of the bitstream. A decoder programmed in Matlab and run on a typical laptop was operating at less than half of real time. A voice message lasts several seconds, so the displayed message should appear a few seconds after the spoken phrase;
- safety and reliability: only error-free messages are displayed. Parity check is used for error detection and syntax of the detected commands is verified (Table 1). In the case of doubts, digital content is not displayed and the cockpit crew should rely on the voice message alone, as in the standard case;
- robustness to channel noise. In most cases the distance between the transmitter and the receiver is short, because ground-to-airplane messages are used during takeoff or landing phases. Therefore, the AM signal is strong and SNR is about 30 dB. In these conditions BER of the watermark transmission approaches 0.001 (Fig. 17) and more than 90% of typical messages are received without errors. This may be improved if EEC are used, at the cost of a lower bit rate. The proposed system may be used at low SNR values, but below 20 dB BER becomes too high and the quality of voice messages deteriorates considerably;
- robustness to resampling. The sampling frequency used at the receiver differs by some tens of Hz from the sampling frequency at the transmitter. Therefore, the sampling frequency should be estimated at the receiver and it should be used in the bit synchronization algorithm. A sampling frequency estimation algorithm based on tone embedding was applied [12]–[14]. This tone is added to SELCAL pulses. The bit synchronization algorithm is robust to a sampling frequency offset of up to 30 Hz, which is sufficient in practice.

The problem of ground-to-airplane messaging may be solved by transmitting a burst of data before or after the speech phrase [17]. This guarantees good robustness to channel noise but the data transmitting signal is audible as a short burst of noise. This is not convenient for crews on a listening watch.

References

[1] “ICAO Standard Phraseology – A Quick Reference Guide for Commercial Air Transport Pilots”, Eurocontrol [Online]. Available: https://www.skybrary.aero/bookshelf/books/115.pdf

[2] “Annex 10 Radiotelephony Procedures”, Aeronautical Mobile Communications Panel (Working Group C), Anchorage, Alaska, 2001.

[3] “Introduction to ACARS Messaging Services”, International Communications Group, Application Note ICS-200-01, Apr. 2006.

[4] “Controller Pilot Data Link Communications (CPDLC)”, SKYbrary, Flight Safety Foundation, Dec. 2017 [Online]. Available: https://www.skybrary.aero/index.php/Controller_Pilot_Data_Link_Communications_(CPDLC)

[5] “Selective Calling (SELCAL) Users Guide”, Aviation Spectrum Resources, 61742 REV C, Dec. 30, 2013.

[6] “Selective Calling (SELCAL) Code Pool Expansion”, Aeronautical Mobile Communications Panel (Working Group – Maintenance), Bucharest, Romania 30 May – 1 June 2012.

[7] K. Hofbauer, H. Hering, and G. Kubin, “Speech watermarking for the VHF radio channel”, in *Proc. EUROCONTROL Innovative Res. Worksh. INO 2005*, Brétigny-sur-Orge, France, 2005, pp. 215–220 [Online]. Available: https://pure.tugraz.at/ws/portalfiles/portal/2721152/Hofbauer_INO_2005.pdf

[8] “SCOPE – Safety of Controller-Pilot Dialogue”, THALES Research & Technology, IntuiLab and Institut de Recherche en Informatique de Toulouse, 2008 [Online]. Available: https://www.eurocontrol.int/eec/public/standard_page/proj_CARE_INO_II_SCOPE.html

[9] P. Dymarski and R. Markiewicz ,“Steganografia akustyczna w tle sygnału mowy”, *Przegl. Telekomun. i Wiadom. Telekomun.*, no. 8/9, pp. 665–669, 2017 (doi: 10.15199/59.2017.8-9.5) [in Polish].

[10] P. Dymarski and R. Markiewicz, “Robust audio watermarks in frequency domain”, *J. of Telecommun. and Inform. Technol.*, no. 2, pp. 12–21, 2014.

[11] R. Tachibana, S. Shimizu, T. Nakamura, and S. Kobayashi, “Audio watermarking method robust against time- and frequency-fluctuation”, in *Proc. of SPIE Int. Conf. on Secur. and Watermark. of Multimed. Cont. III*, San Jose, USA, 2001, vol. 4314, pp. 104–115 (doi: 10.1117/12.435390).

[12] M. Sliskovic, “Sampling frequency offset estimation and correction in OFDM systems”, in *Proc. 8th IEEE Int. Conf. on Electron., Circ. and Syst. ICECS 2001*, Malta, 2001, pp. 437-440 (doi: 10.1109/ICECS.2001.957773).

[13] Z. Piotrowski, “Drift correction modulation scheme for digital audio watermarking”, in *Proc. Int. Conf. on Multimed. Inform. Netw. and Secur. MINES 2010*, Nanjing, Jiangsu, China, 2010, pp. 392–397 (doi: 10.1109/MINES.2010.88).

[14] P. Dymarski and R. Markiewicz, “Time and sampling frequency offset correction in audio watermarking”, in *Proc. of 18th Int. Conf. on Syst., Sig. and Image Process. IWSSIP 2011*, Sarajevo, Bosnia-Herzegovina, 2011.

[15] D. C. Rife and R. R. Boorstyn, “Single-tone parameter estimation from discrete-time observations”, *IEEE Trans. on Inform. Theory*, vol. 20, no. 5, pp. 591–598, 1974 (doi: 10.1109/TIT.1974.1055282).

[16] ITU-T Recommendation P.862: Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs, Feb. 2001.

[17] Ł. Monecki “Transmission of digital messages accompanying analog speech communication”, B.Sc. thesis, Warsaw University of Technology, Sept. 2018 (supervised by P. Dymarski).

Przemysław Dymarski received his M.Sc. and Ph.D. degrees, both in Electrical Engineering, from the Wrocław University of Technology, Poland, in 1974 and 1983, respectively. In 2004 he received a D.Sc. degree in Telecommunications from the Faculty of Electronics and Information Technology, Warsaw University of Technology. Now he is with the Institute of Telecommunications, Warsaw University of Technology. His research interests include various aspects of digital signal processing, particularly speech and audio compression for telecommunications and multimedia, audio watermarking and applications of Hidden Markov Models.

https://orcid.org/0000-0001-6884-8065
E-mail: dymarski@tele.pw.edu.pl
Institute of Telecommunications
Faculty of Electronics and Information Technology
Warsaw University of Technology
Nowowiejska 15/19
00-665 Warsaw, Poland

Paper

Method for Determining Broadcaster Advised Emergency Wake-up Signal for ISDB-T Digital Television Receivers

Satoshi Takahashi

Graduate School of Information Sciences, Hiroshima City University, Hiroshima, Japan

https://doi.org/10.26636/jtit.2019.131619

Abstract—There is a way to automatically wake up television receivers when a broadcaster sends out an emergency alert. In the Integrated Services Digital Broadcasting-Terrestrial (ISDB-T) digital television standard, the emergency wake-up procedure is called an Emergency Warning System (EWS). In ISDB-T, the special signal is embedded in a control message known as transmission and modulation configuration control (TMCC). However, improper identification of the wake-up signal, often encountered in mobile reception, leads to unnecessary wake ups. In this paper, a method of reliably determining a wake-up signal is proposed by assuming that broadcasters will not change the TMCC message except for the wake-up signal when the broadcaster sends out an emergency alert. A change in the wake-up bit leads to variation parity, and the proposed method also relies on such variations. Mutual information to be obtained by the wake-up receiver is evaluated using the memoryless binary asymmetric channel model. Results showed that the proposed method provided mutual information even at a E_b/N_0 being lower than 10 dB. Mutual information of the proposed method with intermittent reception is also analyzed as a function of the duty ratio of the intermittent receiver.

Keywords—binary asymmetric channel, emergency warning system (EWS), intermittent reception, transmission and modulation coding configuration (TMCC).

1. Introduction

When tremors are felt on the ground, we may turn on television sets or radios to listen to alert messages about earthquakes, tsunami alerts, etc. Prompt audio-and-visual alerts during emergencies have been important, as the requirement to place fire sensors inside houses in the United States significantly reduces the number of people who have died while sleeping.

One could easily come up with ideas to provide automatic wake-up television receivers with a broadcaster assigned emergency signal. This idea has been implemented by broadcasters who send special signals to advise of emergencies. Some components of a television receiver, such as the tuner and the power supply controller, are still active and the receiver continues to receive signals from the broadcaster when it is idle, until special messages are sent. Emergency alerts are sent out based on requests from local governments or meteorological agencies.

Emergency alerts provided by means of broadcasting signals are now available. An Emergency Alert System (EAS) is in operation in the United States. The wake-up signal in EAS is encoded into the main audio channel using frequency shift keying (FSK) modulation, and therefore the EAS wake-up signal is mainly available on analog television sets and radios [1]. Digital television sets in the United States employed the Advanced Television Standard Committee (ATSC) standard after analog television broadcasting ended. ATSC-class television receivers had to be woken up by EAS, by decoding the transport stream (TS), the audio channel and the alert message during the idle state. One possible way of designing ATSC wake-up receivers is to use a guard band between broadcasting channels [2]. ATSC 3.0, a successor of ATSC, plans to implement a new EAS wake-up method [3]. The wake-up signal in the Digital Video Broadcasting-Terrestrial (DVB-T) television standard is also encoded into the main audio channel, just as it was the case with FSK modulation in TS, and is delivered by an optional Announcement Service [4]. This procedure is also called EWS. A method of displaying an emergency pop-up message on the screen has been proposed [5].

The emergency wake-up procedure of ISDB-T is defined as EWS. The wake-up signal bit is embedded in a control signal named TMCC, while detailed emergency content is included in the Program Map Table (PMT) in Transport Stream (TS) [6]. Therefore, accurate determination of the specific bit in a TMCC message may serve as a substitute to improve the effectiveness of the wake-up procedure. An emergency wake-up will also be available in the Terrestrial Digital Multimedia Broadcasting (T-DMB) standard [1]. In T-DMB, both the wake-up signal and the detailed information are embedded in the Fast Information Channel (FIC) that also carries multiplexing, service, and conditional access-related information.

It is important for mobile receivers to decrease their power consumption. Separation of transmissions of the wake-up

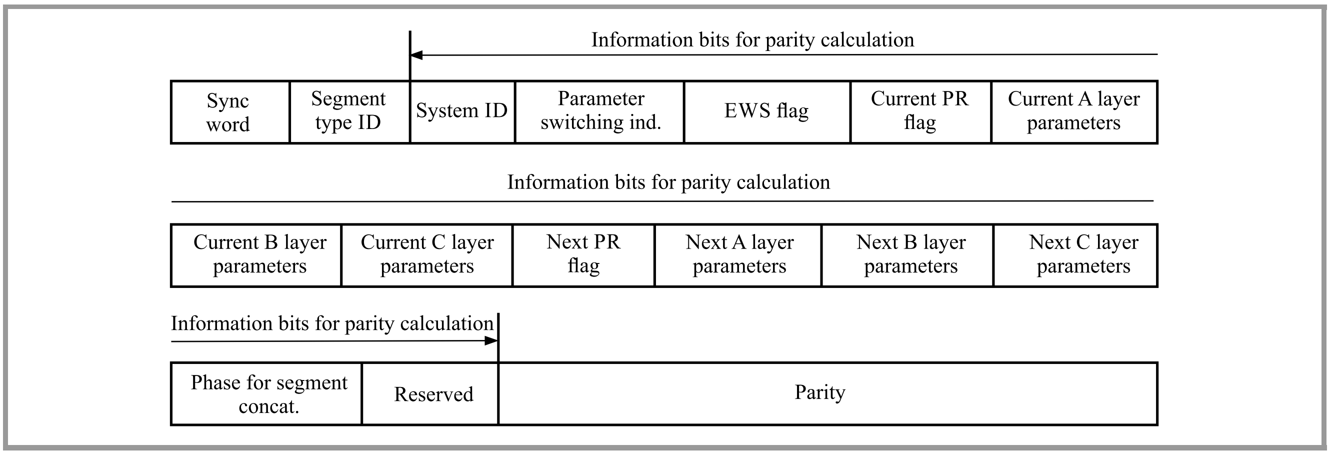

Information bits for parity calculation
Sync word
Segment type ID
System ID
Parameter switching ind.
EWS flag
Current PR flag
Current A layer parameters
Information bits for parity calculation
Current B layer parameters
Current C layer parameters
Next PR flag
Next A layer parameters
Next B layer parameters
Next C layer parameters
Information bits for parity calculation
Phase for segment concat.
Reserved
Parity

Fig. 1. The TMCC message structure.

signal and the signal containing detailed information decreases, in ISDB-T, power consumption of television receivers remaining at an idle state, because all the idle receiver has to do is only to receive the control signal. A way of potentially decreasing power consumption further is to employ intermittent reception. It enables the receiver to sleep periodically, for a certain time, in order to reduce mean power consumption. Because the frequency of disasters is extremely low, the use of intermittently-active receivers could be possible, but on the other hand, intermittent reception may also lead to misdetection of emergency alerts.

In this paper, a method of determining the wake-up signal in ISDB-T is proposed. Performance evaluation is carried out using the probability of misdetection, the probability of a false alarm and mutual information. Mutual information is also analyzed in intermittent reception to determine the trade-off between performance and power consumption.

2. Emergency Wake-up Procedure for ISDB-T Television Receivers

2.1. ISDB-T Signals

ISDB-T employs band segment transmission-orthogonal frequency division multiplexing (BST-OFDM) with 5616 subcarriers and a phase reference subcarrier. It segments a 5.7 MHz channel band into 13 bands. Each segment consists of two control signals (TMCC and AC[1]) and the payload.

ISDB-T digital television is capable of layering the payload into A, B and C layers. Each layer employs mutually different modulations and coding rates. For example, a broadcaster would choose one-segment QPSK modulation in the A layer for mobile receivers, and 12-segment 64 QAM modulation in the B layer for fixed receivers. Some important messages, such as the network information table (NIT) and conditional access table (CAT) are sent in the A layer. The arrangement, as well as the wake-up signal, are described in the TMCC message.

The TMCC signal is periodically sent in 0.2 s cycles by differential bi-phase shift keying (DBPSK) modulation. ISDB-T broadcasters in Japan use a subcarrier bandwidth called mode 3, and there are four TMCC subcarriers in a segment. Each TMCC signal contains a 204-bit message and the message is periodically sent at 992 bps.

The TMCC message structure is shown in Fig. 1. It consists of a 1-bit length phase reference to demodulate the DBPSK signal (not shown in Fig. 1), a 16-bit length synchronization word (alternation of fixed patterns 0x39EE and 0xCA11 in a hexagonal expression), a 3-bit length segment type identifier, a 2-bit length system identifier, a 4-bit length indication of parameter change and a 1-bit length emergency wake-up signal (denoted by the EWS flag). There is also a 1-bit length current partial reception flag, a 39-bit length current modulation, coding rate, and segment number, a 1-bit length next partial reception flag, a 39-bit length next modulation, coding rate, and segment number, a 3-bit length phase compensation for segment concatenation, a 12-bit length reserved space and an 82-bit length parity.

2.2. Bit Error Rate in Receiving TMCC Messages

The radio signal in stationary reception arriving at the receiver is represented by additive white Gaussian noise (AWGN). The bit error rate, P_e, for DBPSK modulation is

$$P_e = \frac{1}{2}\mathrm{e}^{-\gamma} ,$$

where γ is the signal energy per bit above the noise power density, which is usually denoted by E_b/N_0. γ is also proportional to signal strength. An increase in γ monotonically decreases P_e.

[1] Auxiliary channel (AC) is used for broadcasters and for the early earthquake warning in ISDB-T.

Signal strength observed by a moving receiver varies from time to time due to the multipath phenomenon. Signal fluctuation due to Rayleigh fading is caused by both scattered waves in mobile reception. The P_e is [7]:

$$P_e = \frac{1}{2}\,\frac{1+\gamma(1-\rho_C)}{1+\gamma}\ , \tag{1}$$

where ρ_C is the correlation coefficient of signals between the symbol duration of T_s. The ρ_C of the uniformly spread scatterers model becomes:

$$\rho_C = J_0(2\pi f_D T_s)\ , \tag{2}$$

where $J_0(\cdot)$ is the first-kind Bessel function of the zeroth order, and f_D is the maximum Doppler frequency. f_D is obtained from $f_D = v/\lambda$, and v is the velocity and λ is the wavelength. Because the error is the source of the multipath and the Doppler effect, the P_e in mobile reception depends on v and T_s. Various ρ_C are found in reference [7].

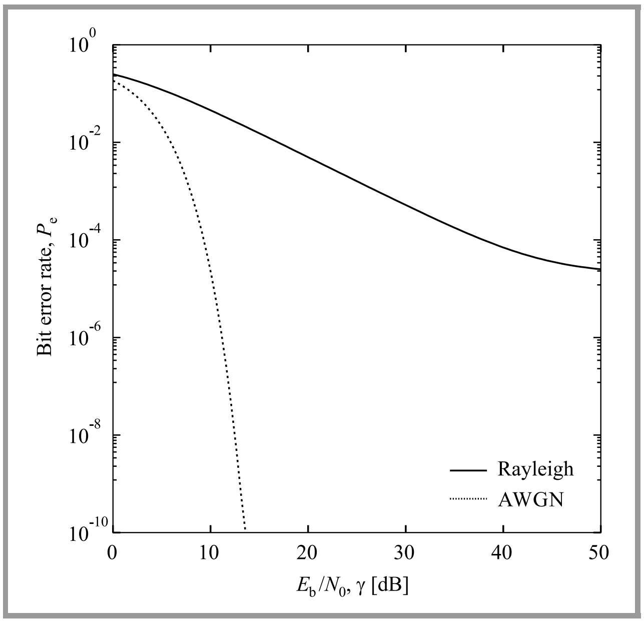


Fig. 2. P_e vs. γ in stationary reception (AWGN) and mobile reception (Rayleigh fading) environments.

P_e in AWGN and Rayleigh fading environments are compared in Fig. 2. In the Rayleigh fading environment, an increase in γ leads to the exhibition of P_e floor, and it is often referred to as an irreducible error. The error floor in Rayleigh fading is calculated by taking the limit of $\gamma \to \infty$ in Eq. (1),

$$P_e = \frac{1}{2}(1-\rho_C)\ . \tag{3}$$

2.3. Error Correction Code Employed in TMCC Message

Parity in a TMCC message is generated by the polynomial of $x^{82}+x^{77}+x^{76}+x^{71}+x^{67}+x^{66}+x^{56}+x^{52}+x^{48}+x^{40}+x^{36}+x^{34}+x^{24}+x^{22}+x^{18}+x^{10}+x^{4}+1$. The $(273, 191)$ difference set cyclic code is a type of the BCH code that is capable of correcting 8 error bits[2]. In general, decoding a BCH code requires cumbersome polynomial factorization. But the code may be decoded by majority determination of the summarized syndromes [8]. The code originally proposed by Weldon is shortened to fit the 102-bit TMCC information. A method of efficiently correcting errors with the $(184, 102)$ shortened code is also proposed, relying on majority determination of the syndrome summary [9].

At the receiver, the TMCC message is divided by the generating polynomial to obtain the 82 syndrome bits $S_0 - S_{81}$. They are further summarized into following 18 work bits:

$$\begin{aligned}
A_1 &= S_{71}+S_{76}\\
A_2 &= S_{17}\\
A_3 &= S_5+S_{23}\\
A_4 &= S_{21}+S_{27}+S_{45}\\
A_5 &= S_3+S_{25}+S_{31}+S_{49}\\
A_6 &= S_{16}+S_{40}+S_{42}+S_{66}\\
A_7 &= S_{35}+S_{52}+S_{56}+S_{78}\\
A_8 &= S_8+S_{44}+S_{61}+S_{65}\\
A_9 &= S_2+S_{11}+S_{47}+S_{64}+S_{68}\\
A_{10} &= S_{10}+S_{13}+S_{22}+S_{58}+S_{75}+S_{79}\\
A_{11} &= S_1+S_{12}+S_{15}+S_{24}+S_{60}+S_{77}+S_{81}\\
A_{12} &= S_{30}+S_{32}+S_{43}+S_{46}+S_{55}\\
A_{13} &= S_6+S_{37}+S_{39}+S_{50}+S_{53}+S_{62}\\
A_{14} &= S_0+S_7+S_{38}+S_{40}+S_{51}+S_{54}+S_{63}\\
A_{15} &= S_{18}+S_{19}+S_{26}+S_{57}+S_{59}+S_{70}+S_{73}\\
A_{16} &= S_9+S_{28}+S_{29}+S_{36}+S_{67}+S_{69}+S_{80}\\
A_{17} &= S_4+S_{14}+S_{33}+S_{34}+S_{41}+S_{72}+S_{74}
\end{aligned}$$

If more than 8 bits out of $A_1 - A_{17}$ are active, the first bit is determined to be wrong and should be inverted. The procedure of the cyclic shift of the syndrome $S_0 - S_{81}$, the calculation of $A_1 - A_{17}$, and the majority determination are repeated to correct errors in the TMCC message. We could correct an error in the wake-up signal by repeating the procedure by 25 times.

3. Proposed Method of Determining Emergency Wake-Up Signal

3.1. Observation of Broadcaster Advised TMCC Message under Normal Conditions

Broadcaster-sent TMCC messages were observed in the Hiroshima area in Japan to determine actual TMCC messages. An ISDB-T front-end decoder by EIDEN 6500A-001 was

[2] Because the Hamming weight is 18, this code can correct up to 8 bits through an erroneous channel.

Table 1
Results of observations of broadcaster-sent TMCC messages in the area of Hiroshima city in Japan

System identification	Terrestrial digital television
Parameter switching	Normal
Wake-up signal (EWS flag)	Inactive
Current information	Partial reception on
A layer	QPSK, coding rate of 2/3 interleave length of 4, one segment
B layer	64QAM, coding rate of 3/4 interleave length of 2, 12 segments
C layer	Unused
Next information	Same as current ones
Concatenate transmission	Unused
Reserved	Unused

used in this observation. Except for the original channel of a cable television (CATV) broadcaster, all six broadcasters sent the same TMCC message listed in Table 1.

Table 2
Parity bit changes according to a change in the wake-up signal

Bit	Change	Coefficient	Bit	Change	Coefficient
122	0 → 1	x^{81}	160	1 → 0	x^{43}
124	1 → 0	x^{79}	163	1 → 0	x^{40}
127	0 → 1	x^{76}	164	1 → 0	x^{39}
128	1 → 0	x^{75}	165	1 → 0	x^{38}
130	1 → 0	x^{73}	169	0 → 1	x^{34}
132	1 → 0	x^{71}	170	1 → 0	x^{33}
133	0 → 1	x^{70}	171	1 → 0	x^{32}
135	0 → 1	x^{68}	173	1 → 0	x^{30}
138	0 → 1	x^{65}	177	1 → 0	x^{26}
139	0 → 1	x^{64}	178	1 → 0	x^{25}
140	0 → 1	x^{63}	182	1 → 0	x^{21}
142	1 → 0	x^{61}	183	0 → 1	x^{20}
144	0 → 1	x^{59}	187	1 → 0	x^{16}
148	0 → 1	x^{55}	194	1 → 0	x^{9}
149	0 → 1	x^{54}	200	0 → 1	x^{3}
150	1 → 0	x^{53}	201	1 → 0	x^{2}
156	1 → 0	x^{47}	202	1 → 0	x^{1}
158	1 → 0	x^{45}			

The TMCC message shown in Table 1 is 3D 25 8B 4B 3F FF 25 8B 4B 3F FF FF FC in a hexagonal representation. The 82-bit length parity summarizes the 102-bit TMCC message. Dividing the polynomial representing the TMCC message by the generating polynomial, we obtain the parity of 2B E8 19 CF AE 72 DB A8 F8 A5 80. If the wake-up signal is sent, the parity becomes 8D 5C F3 F7 84 03 0A 24 B8 26 00.

3.2. Determination Method Proposed

We assume that the broadcasters would not change the TMCC message when sending out an emergency signal, except for the wake-up signal. A change in the wake-up signal varies the parity bits listed in Table 2 in a scenario in which the broadcaster sends the TMCC message listed in Table 2 where the broadcaster sends the TMCC message listed in Table 1.

Therefore, the proposed method of determining the wake-up signal is described by the majority of the corresponding bits matched as the 26-th bit (the wake-up signal) being one, the 122-nd bit being one, the 124-th bit being zero and so on, while the synchronization word exactly matches the definition [10]. It is unlucky that the total number of them is an even number of 36. Here, the receiver determines the wake-up signal by agreeing to more than 18 matching bits.

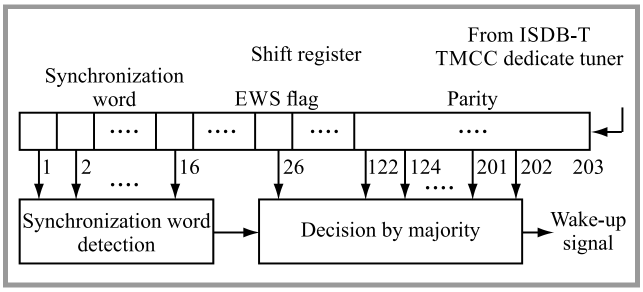


Fig. 3. Schematic block diagram of the proposed method.

It should be noted that parity bit positions to be alternated by change in the wake-up signal state are fixed, regardless of what type of TMCC message is provided, though their values are changed according to the TMCC message kind, as such is the characteristics of linear codes. The schematic block diagram is shown in Fig. 3.

3.3. Misdetection and False Alarm Probabilities

Performance is evaluated in terms of misdetection probability $P_{\rm md}$, and the false alarm probability, $P_{\rm fa}$. Misdetection means the receiver has missed the wake-up signal, and a false alarm means the receiver has wrongly been activated when the wake-up signal was not present.

First of all, we obtain $P_{\rm md}$ and $P_{\rm fa}$ when the receiver determines the wake-up signal only. Such determination is defined, hereinafter, as single-bit determination. The receiver decodes a TMCC message after the frame synchronization that matches the reception bit sequence with the synchronization word. $P_{\rm md}$ for the single-bit determination is the complement probability that all 16-bit synchronization words agrees and that the wake-up signal is correctly detected:

$$P_{\rm md}^{\rm single} = 1-(1-P_e)^{17}\ . \tag{4}$$

On the other hand, P_{fa} is the probability that the synchronization word matches and the wake-up signal is wrongly detected:

$$P_{\mathrm{fa}}^{\mathrm{single}} = (1-P_e)^{16} P_e \ . \tag{5}$$

P_{md} for the proposed method is the complement probability that the synchronization word will agree and that more than 18 bits of the 36 corresponding bits will agree, and is [3]:

$$P_{\mathrm{md}}^{\mathrm{prop}} = 1-(1-P_e)^{16} \cdot \left\{ \sum_{k=0}^{18} {}_{36}C_k \, (1-P_e)^{36-k} P_e^k \right\} \ . \tag{6}$$

The proposed P_{fa} is also obtained where the 16 synchronization bits are correctly received and more than 18 bits of the 36 bits are wrong:

$$P_{\mathrm{fa}}^{\mathrm{prop}} = (1-P_e)^{16} \cdot \left\{ \sum_{k=18}^{36} {}_{36}C_k \, (1-P_e)^{36-k} P_e^k \right\} \ . \tag{7}$$

P_{fa} is a decreasing function for a sufficiently small P_e, but it is also an increasing function where P_e is near to 0.5, which is the highest value. Therefore, the P_{fa} is a convex function of P_e.

The receiver's error correction of the wake-up signal is expressed as "ec" It seems possible to obtain P_{md} for an ec of the receiver as:

$$\tilde{P}_{\mathrm{md}}^{\mathrm{ec}} = 1-(1-P_e)^{16} \cdot \left\{ \sum_{k=0}^{8} {}_{184}C_k \, (1-P_e)^{184-k} P_e^k \right\} \ , \tag{8}$$

but it is virtually impossible to solve the equation, since the number of combinations becomes huge, while the exponent of P_e rapidly approaches zero. Therefore, transmission performance has been evaluated by the Monte Carlo method that uses computer-generated random numbers [11]. But an analytical evaluation using the summarized syndrome in Subsection 3.2 is proposed [12]. P_{md} is the complement probability that the synchronization word is correctly received and that 8 or fewer bits out of the 17 work bits are active:

$$P_{\mathrm{md}}^{\mathrm{ec}} = 1-(1-P_e)^{16} \cdot \left\{ \sum_{k=0}^{8} {}_{17}C_k \, (1-P_e)^{17-k} \cdot P_e^k \right\} \ . \tag{9}$$

Because a single syndrome is obtained with exclusive-or operations of the received sequence, the active ratio of the single syndrome is P_e and the work bit active ratio is also P_e. Therefore, P_{fa} is also obtained when the synchronization word is correctly received and 9 or more work bits are active,

$$P_{\mathrm{fa}}^{\mathrm{ec}} = (1-P_e)^{16} \cdot \left\{ \sum_{k=9}^{17} {}_{17}C_k \, (1-P_e)^{17-k} P_e^k \right\} \ . \tag{10}$$

The frequency of 600 MHz, the moving velocity of 10 m/s, and Rayleigh fading were also assumed. Substituting Eqs. (1) and (2) into Eqs. (4), (6) and (9), we obtain P_{md}. They are compared in Fig. 4a. All P_{md} were monotonically decreased as γ increased, and were almost the same, because the probability that the synchronization word would agree was dominant over the probability of agreeing corresponding bits. P_{md} also indicated a floor value, and the high P_{md}, even for a higher γ, was the remaining problem for the all methods.

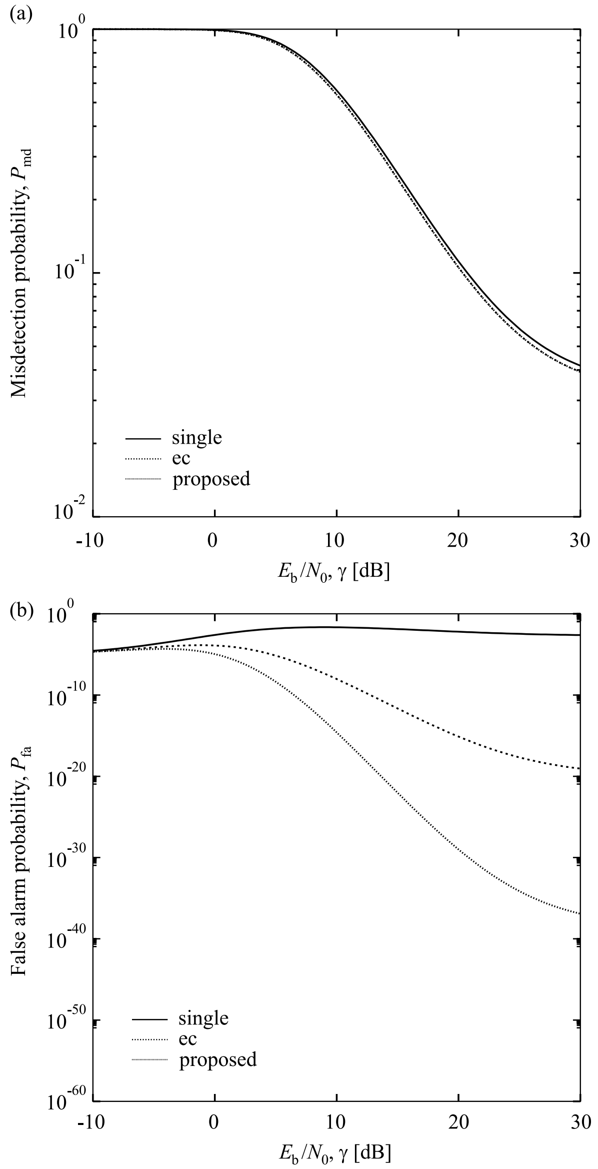


Fig. 4. P_{md} and P_{fa} comparisons in a Rayleigh fading environment: (a) P_{md} and (b) P_{fa}.

P_{fa} are also derived by substituting Eqs. (1) and (2) into Eqs. (5), (7) and (10). They are shown in Fig. 4b. For a higher γ, P_{fa} decreased as γ increased. But for a lower γ, P_{fa} decreased as γ decreased, because synchronization tended to be lost and the receiver did not received any

[3] The combinations number of k out of n, ${}_nC_k$ for a large n can be calculated using the gamma function as $n! = \Gamma(n+1)$ and ${}_nC_k = \dfrac{n!}{k!\,(n-k)!}$.

alerts. Therefore, P_{fa} were a convex in shape. The proposed method provided the lowest P_{fa}. The proposed method uses the parity capability only for correcting the wake-up signal, while the ec receiver uses that capability for correcting all information bits. Focusing on the said capability in the proposed method reduced P_{fa}.

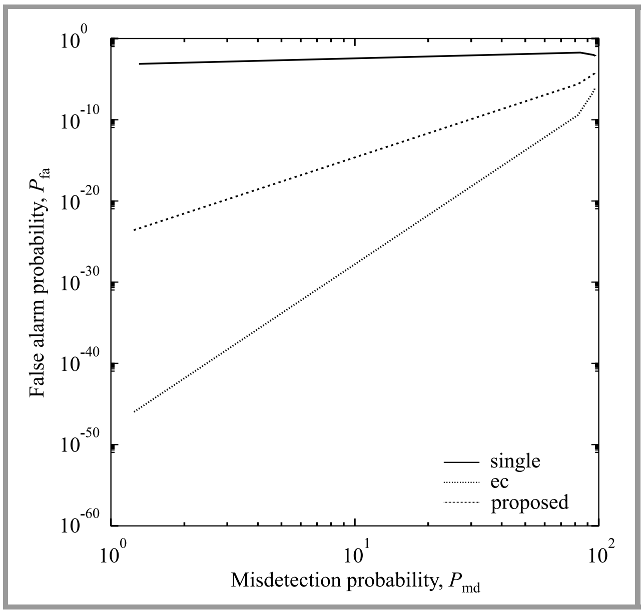


Fig. 5. P_{fa} vs. P_{md} at $v = 10$ m/s.

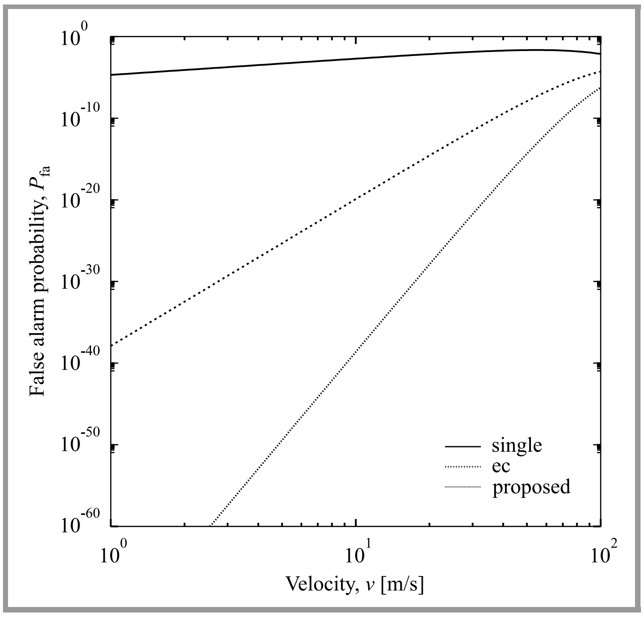


Fig. 6. P_{fa} as a function of v for various methods at $\gamma \rightarrow \infty$.

Both P_{md} and P_{fa} depended on P_e. A strict determination would decrease P_{fa} but would increase P_{md}, while a looser determination may decrease P_{md} but increases P_{fa}. For comparing the trade-off, P_{fa} as a function of P_{md} is plotted in Fig. 5. According to the figure, the proposed method indicated the lowest P_{fa} among all other methods. If we assume that power consumption in the television receiver portion was more dominant than in in portion related to wake-up signal determination, power consumption during the idle state could also be decreased by the same rate as in false alarms.

P_{fa} for a sufficiently high γ is obtained to evaluate the P_e floor effect on P_{fa}. The P_{fa} is obtained using Eqs. (2), (3), (5), (7), and (10) and the results are shown in Fig. 6 as a function of v. A decrease in v also decreased P_{fa}, and the proposed method decreased P_{fa} significantly more than other methods did.

4. Mutual Information Obtained from Wake-up Receiver

Performance can also be compared using the mutual information, instead of using P_{md} and P_{fa}. For evaluating mutual information, we use the line diagram shown in Fig. 7, in which the broadcaster sends an wake-up signal X. We use 0 for expressing the situation in which the broadcaster does not send the wake-up signal and 1 is used for sending out the wake-up signal. Then, P_{md} and P_{fa} can be expressed as p_{10} and p_{01}, respectively. Because the curve shapes of P_{fa}

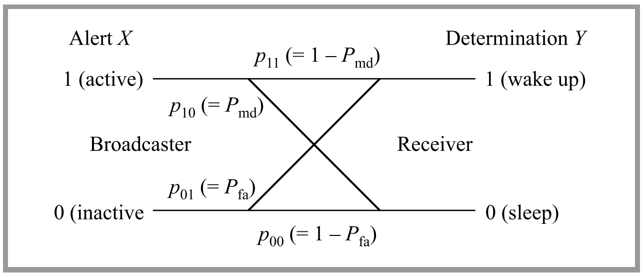

Alert X
Determination Y
p_{11} (= 1 – P_{md})
1 (active)
1 (wake up)
p_{10} (= P_{md})
Broadcaster
Receiver
p_{01} (= P_{fa})
0 (inactive
0 (sleep)
p_{00} (= 1 – P_{fa})

Fig. 7. Line diagram between the broadcaster's alert X and receiver's determination Y.

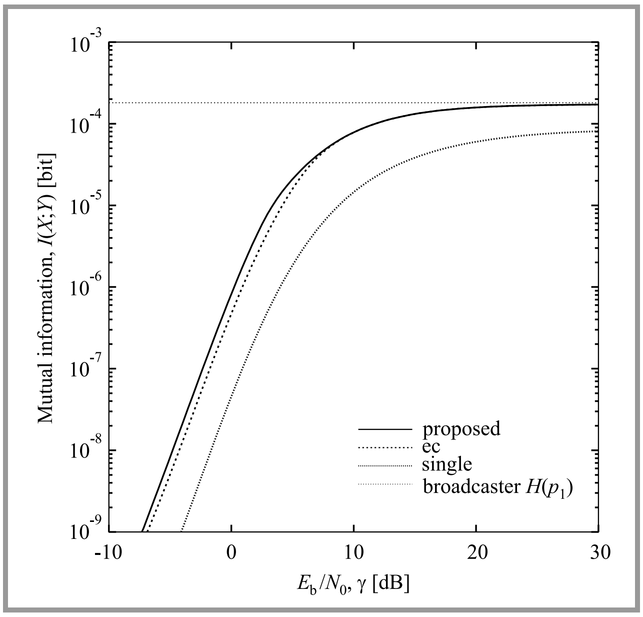


Fig. 8. $I(X;Y)$ as a function of γ at $p_1 = 10^{-5}$.

and P_{md} as a function of γ were different, the line diagram is asymmetric. Because we can also assume that the current determination does not affect future determination, the channel is memoryless. Mutual information $I(X;Y)$ is:

$$\begin{aligned} I(X;Y) &= (1-p_1)\left\{(1-P_{\mathrm{fa}})\log_2\frac{1-P_{\mathrm{fa}}}{q_0}+P_{\mathrm{fa}}\log_2\frac{P_{\mathrm{fa}}}{q_1}\right\} \\ &\quad + p_1\left\{P_{\mathrm{md}}\log_2\frac{P_{\mathrm{md}}}{q_0}+(1-P_{\mathrm{md}})\log_2\frac{1-P_{\mathrm{md}}}{q_1}\right\}\ , \\ q_0 &= (1-p_1)(1-P_{\mathrm{fa}})+p_1P_{\mathrm{md}}\ , \\ q_1 &= (1-p_1)P_{\mathrm{fa}}+p_1(1-P_{\mathrm{md}})\ , \end{aligned} \tag{11}$$

where p_1 is the probability of an emergency alert.

$I(X;Y)$ are plotted in Fig. 8 where we assume $p_1 = 10^{-5}$. p_1 corresponds to the probability of an emergency alert being issued for 26 min over a one year period. The entropy of the broadcaster-issued emergency alerts:

$$H(p_1) = -p_1\log_2(p_1)-(1-p_1)\log_2(1-p_1)\ , \tag{12}$$

was also shown in the figure. For a lower region of γ, $I(X;Y)$ increased along with the increase in γ. On the other hand, $I(X;Y)$ saturated to $H(p_1)$ for the higher γ. Saturation means that we cannot extract more information from the receiver. The γ that $I(X;Y)$ saturated are about 5 dB for the proposed method and the error correction method, and γ is about 10 dB for the single-bit determination.

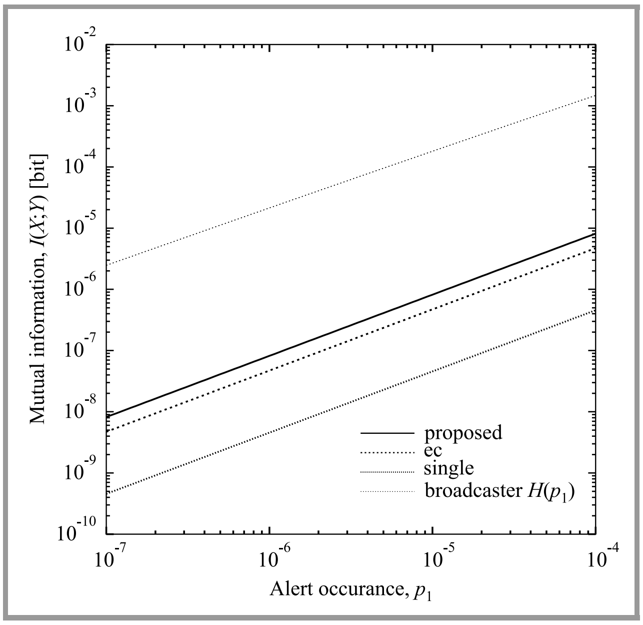


Fig. 9. $I(X;Y)$ as a function of p_1 at $\gamma = 0$ dB.

$I(X;Y)$ as a function of p_1 is calculated and shown in Fig. 9 at $\gamma = 0$ dB. $I(X;Y)$ increased linearly along with the increase in p_1. $I(X;Y)$ for the proposed method and the error correction method were almost the same, and $I(X;Y)$ for the single-bit determination was lower than the above mentioned values.

The dependence of $I(X;Y)$ on v is shown in Fig. 10. In the figure, $I(X;Y)$ was obtained assuming a sufficiently high γ.

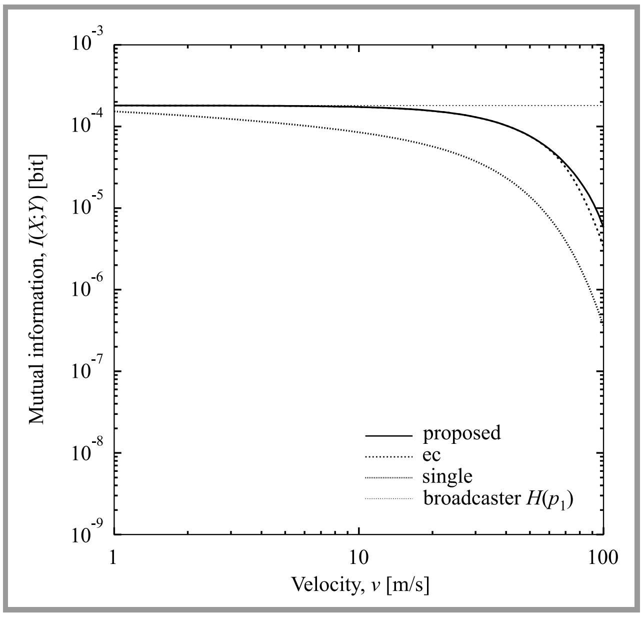


Fig. 10. $I(X;Y)$ as a function of v at $p_1 = 10^{-5}$ and $\gamma \to \infty$.

Therefore, $I(X;Y)$ at v of 30 m/s or less are the same in both the methods, but $I(X;Y)$ for the single-bit determination decreased at the higher v because of the higher P_e.

5. Change in False Alarm and Misdetection Probabilities due to Intermittent Reception

Here, we introduce the intermittent reception of the wake-up signal for the proposed method. The following analysis of intermittent reception will indicate that both false alarm

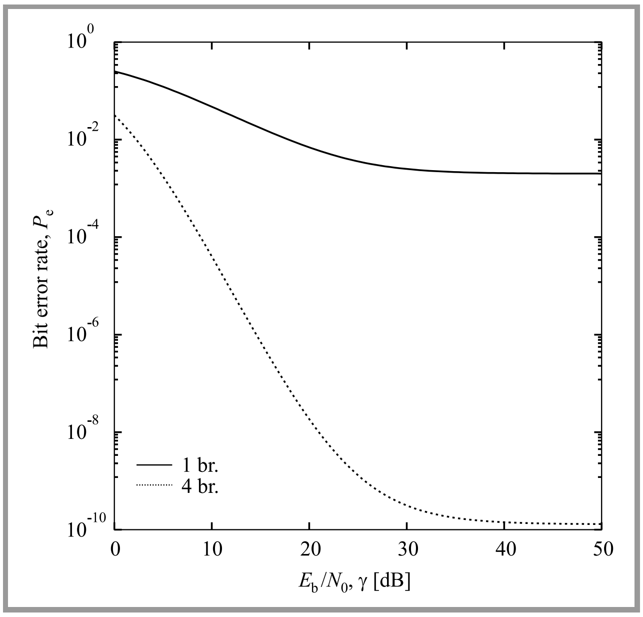


Fig. 11. P_{e} for TMCC with 1-subcarrier branch (1 br.) and 4-subcarrier branch (4 br.) diversity in a Rayleigh fading environment.

and misdetection rates are increased significantly. Therefore, we employ the subcarrier frequency diversity of 4 TMCC signals. P_e with a 4-branch maximum ratio combining the TMCC signals is also derived [7]:

$$P_e = \frac{1}{2}\left\{\frac{1+\gamma(1-\rho_C)}{1+\gamma}\right\}^4. \tag{13}$$

The comparison of P_e for a 1-subcarrier branch and a 4-subcarrier branch diversity in a Rayleigh fading environment is plotted in Fig. 11.

P_{md} and P_{fa} in TMCC with a 1-subcarrier branch and a 4-subcarrier branch diversity are compared in Fig. 12. Subcarrier diversity has significantly reduced both P_{md} and P_{fa}.

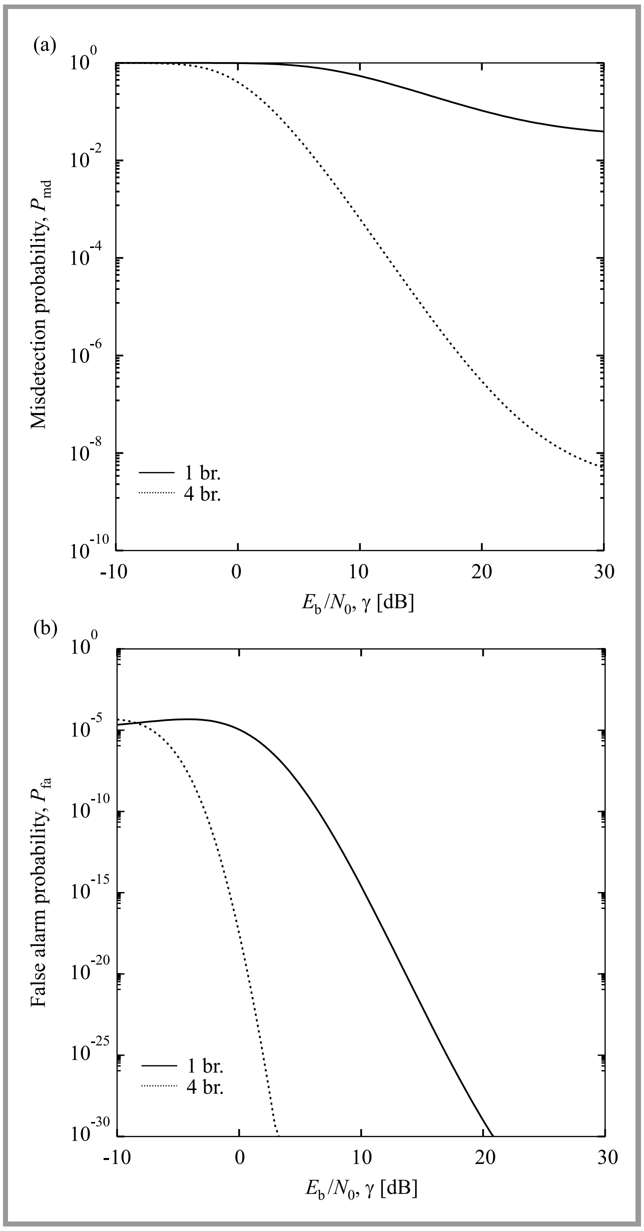


Fig. 12. P_{md} and P_{fa} of the proposed method with and without subcarrier diversity: (a) P_{md} and (b) P_{fa}.

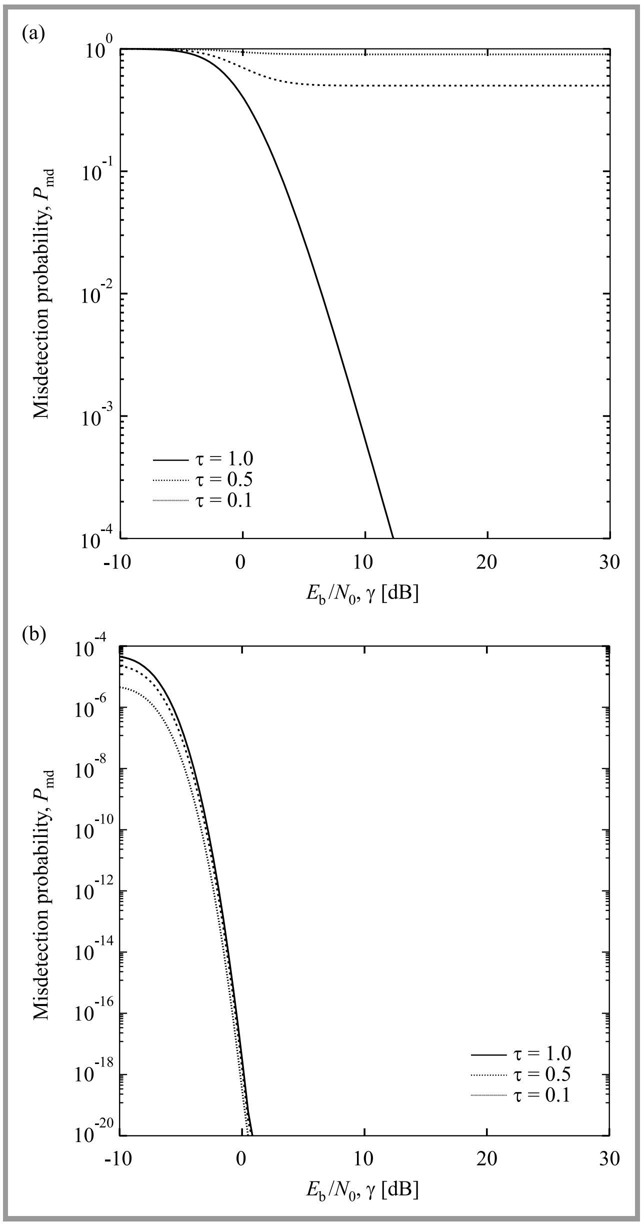


Fig. 13. P_{md} and P_{fa} for various τ: (a) P_{md} and (b) P_{fa}.

While the receiver is in the sleep mode, it does not misdetect wake-up signals, nor does it produce false alarms. For the actual duty ratio τ, $(0 < \tau \leq 1)$, the false alarm probability $\tilde{P_{\mathrm{fa}}}$ is:

$$\tilde{P_{\mathrm{fa}}} = \tau P_{\mathrm{fa}}. \tag{14}$$

On the other hand, the complementary event of misdetection probability (i.e. detection probability) is increased by τ times over the complementary event of P_{md}. The misdetection probability, $\tilde{P_{\mathrm{md}}}$, becomes:

$$\tilde{P_{\mathrm{md}}} = 1 - \tau(1 - P_{\mathrm{md}}). \tag{15}$$

Substituting Eq. (13) into Eqs. (6), (7), (15), and (14), we obtain $\tilde{P_{\mathrm{md}}}$ and $\tilde{P_{\mathrm{fa}}}$ as in Fig. 13. This figure indicated that an increase in both γ and τ decreased P_{md}, and that τ significantly impacted P_{md}. On the other hand, an increase in τ increased P_{fa}.

$I(X;Y)$ is also calculated and plotted in Fig. 14, where $\tau = 0.1$, 0.5, and 1.0. $I(X;Y)$ decreased as along with the decrease in τ, and $I(X;Y)$ is almost proportional to τ. A receiver with a smaller τ produces a smaller $I(X;Y)$. The saturation value of $I(X;Y)$ also decreased with a decrease in τ. The intermittent reception reduced the mutual information that could not be compensated by increasing γ.

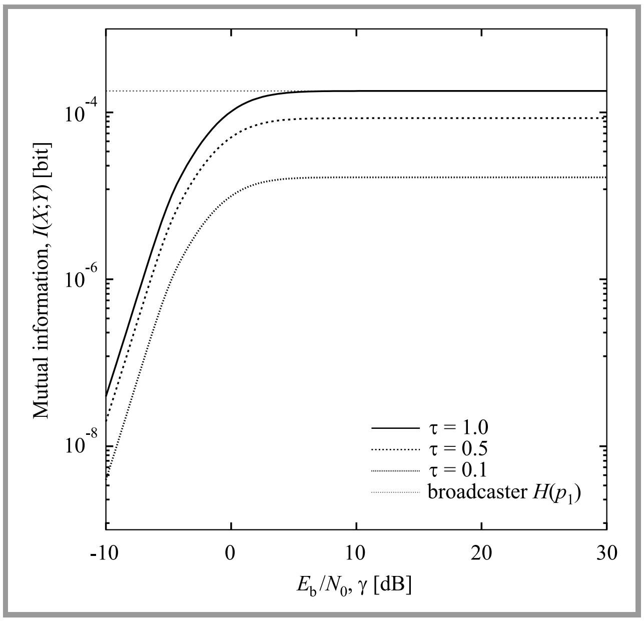


Fig. 14. $I(X;Y)$ as a function of γ for various τ.

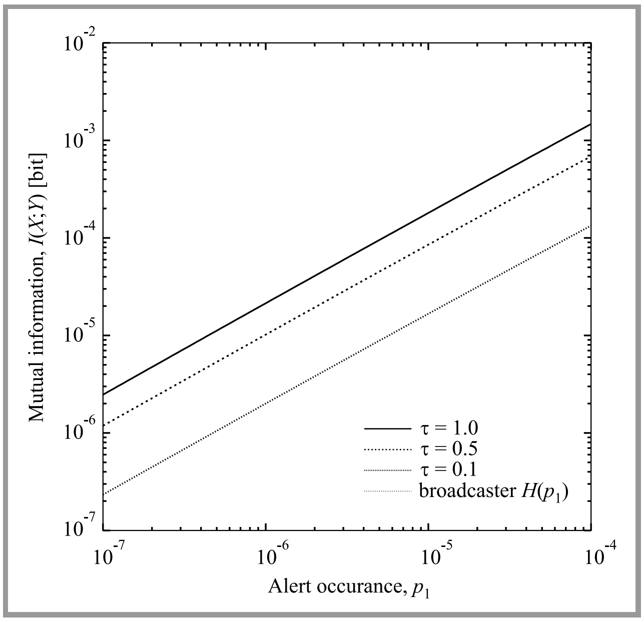


Fig. 15. $I(X;Y)$ as a function of p_1 for various τ.

For obtaining the highest $I(X;Y)$ for various p_1 and τ, the P_e floor value is calculated. $I(X;Y)$ is plotted in Fig. 15 by taking the limit of $\gamma \to \infty$. $I(X;Y)$ increased along with the increase in p_1, and $I(X;Y)$ decreased along with a decrease in τ. $I(X;Y)$ approached $H(p_1)$ at $\tau = 1$.

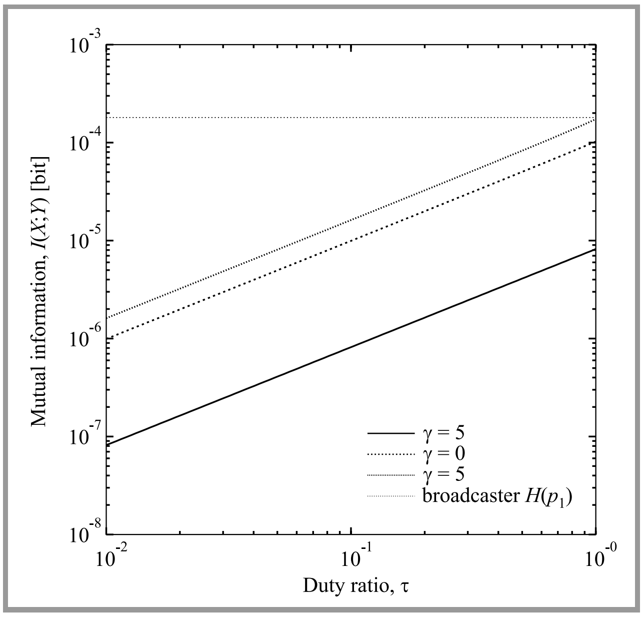


Fig. 16. $I(X;Y)$ as a function of τ for $\gamma = -5, 0,$ and 5 dB.

For obtaining $I(X;Y)$ in a weak signal reception environment, $I(X;Y)$ is calculated and plotted in Fig. 16 as a function of τ at $\gamma = 0$ dB. According to the figure, $I(X;Y)$ increased along with the increase in τ at a constant rate, and no significant change was observed.

6. Conclusion

A method of identifying wake-up signals was proposed to reduce the number of false alarms in ISDB-T digital television receivers during idle state. It has been assumed, in this research, that broadcasters did not change the TMCC message except for a situation in which a wake-up signal is sent out. This paper proposed the majority decision method concerning the wake-up signal and corresponding parity bits. The proposed method decreased the number of false alarms, especially for low-mobility users. Mutual information on intermittent reception was also analyzed using the memoryless binary asymmetrical channel model. Intermittent reception always decreased the mutual information that could not be compensated with a higher E_b/N_0. The mutual information exhibited full saturation in a high E_b/N_0 region.

Acknowledgments

This work was supported under the JSPS Kakenhi Grant Number 17K06436 and under the Hiroshima City University Grant for Special Academic Research (General Studies).

References

[1] S. J. Choi, "Analysis of emergency alert services and systems", in P*roc. IEEE Int. Conf. on Converg. Inform. Technol. ICCIT 2007*, Gyeongju, South Korea, 2007, pp. 657–662
(doi: 10.1109/ICCIT.2007.362).

[2] K. Ryu, I. Park, and H. M. Kim, "Wake-up system design criteria for emergency alerting using DTV guard band", in *Proc. IEEE Int. Symp. on Broadband Multimedia Systems and Broadcasting BMSB 2013*, London, UK, 2013, pp. 1–3 (doi: 10.1109/BMSB.2013.6621794).

[3] B. Kovacs, "EAS and AWARN: planning the future in emergency alerts", *IEEE Broadcast Technol.*, vol. Fourth Quarter, pp. 4–8, 2017 [Online]. Available: https://bts.ieee.org/images/files/newsletters/Preview_Quarter_4.pdf

[4] "DVB Emergency Warning System (EWS)" [Online]. Available: http://www.dvb.org/resources/public/factsheets/DVB-EWS-Fact-sheet.pdf (accessed on 4th Nov., 2018)

[5] R. Azmi, H. Budiarto, and R. Widyanto, "A proposed disaster emergency warning system standard through DVB-T in Indonesia", in *Proc. of the Int. Conf. on Electrical Engineering and Informatics*, Bandung, Indonesia, 2011, pp. 1–4, 2011 (doi: 10.1109/ICEEI.2011.6021746).

[6] Association of Radio Industries and Business (ARIB) ed., Transmission System for Digital Terrestrial Television Broadcasting, ARIB STD-B31, 2.2 ed., Tokyo, 2014 [Online]. Available: http://www.arib.or.jp/english/html/overview/doc/6-STD-B31v2_E1.pdf

[7] A. Goldsmith, *Wireless Communications*. New York: Cambridge University Press, 2005 (ISBN: 0521837162).

[8] E. J. Weldon, "Difference-set cyclic codes", *Bell Syst. Tech. J.*, vol. 45, no. 7, pp. 1045–1055, 1966 (doi: 10.1002/j.1538-7305.1966.tb01686.x).

[9] O. Yamada, "Development of an error-correction method for data packet multiplexed with TV signals", *IEEE Trans. on Commun.*, vol. 35, no. 1, pp. 21–31, 1987 (doi: 10.1109/TCOM.1987.1096669).

[10] S. Takahashi, "A novel method of determining EWS wake-up trigger for ISDB-T digital television receivers", in *Proc. 10th Int. Conf. on Wirel. and Mob. Comput., Network. and Commun. WiMob 2014*, Larnaca, Cyprus, 2014, pp. 407–412 (doi: 10.1109/WiMOB.2014.6962193).

[11] O. Yamada, "Introduction to coding theory; application of coding theory to broadcasting technology", *J. of the Inst. of Television Engin. of Japan*, vol. 45, no. 8, pp. 970–980, 1991 (doi: 10.3169/itej1978.45.970) [in Japanese].

[12] S. Takahashi, "Comparison of two majority determination methods of detecting emergency wake-up trigger for ISDB-T terrestrial digital television receivers", in *Proc. of the Int. Wirel. Commun. & Mob. Comput. Conf. IWCMC 2015*, Dubrovnik, Croatia, 2015, pp. 194–198 (doi: 10.1109/IWCMC.2015.7289081).

Satoshi Takahashi received his B.E., M.E. and Ph.D. degrees from Tokyo Denki University, Japan, in 1990, 1992, and 2001, respectively. He joined Hitachi, Ltd. in 1992, where he was involved in conducting research on radio propagation for indoor wireless systems. He was a research engineer at YRP Key Tech Labs Co. Ltd. in 1996, where he was engaged in research on radio propagation and systems for future generation mobile radio communication methods. He joined the Communications Research Laboratory (CRL) in 2002, now operating under the name of the National Institute of Information and Communications Technology (NICT). He was engaged there in researching intelligent transport systems (ITS) and future radio communication systems. Since 2005, he has been an associate professor at the Hiroshima City University. Dr. Takahashi is a senior member of IEICE, as well as a member of IEEE, ITE and IPSJ.

E-mail: s.takahashi@m.ieice.org
Graduate School of Information Sciences
Hiroshima City University
3-4-1 Ozuka-Higashi, Asa-Minami
Hiroshima 731-3194, Japan

Paper

WannaCry Ransomware: Analysis of Infection, Persistence, Recovery Prevention and Propagation Mechanisms

Maxat Akbanov[1], Vassilios G. Vassilakis[2], and Michael D. Logothetis[3]

[1] *Department of Computer Science, University of York, York, United Kingdom*
[2] *University of York, York, United Kingdom*
[3] *WCL, Dept. of Electrical and Computer Engineering, University of Patras, Patras, Greece*

https://doi.org/10.26636/jtit.2019.130218

Abstract—In recent years, we have been experiencing fast proliferation of different types of ransomware targeting home users, companies and even critical telecommunications infrastructure elements. Modern day ransomware relies on sophisticated infection, persistence and recovery prevention mechanisms. Some recent examples that received significant attention include WannaCry, Petya and BadRabbit. To design and develop appropriate defense mechanisms, it is important to understand the characteristics and the behavior of different types of ransomware. Dynamic analysis techniques are typically used to achieve that purpose, where the malicious binaries are executed in a controlled environment and are then observed. In this work, the dynamic analysis results focusing on the infamous WannaCry ransomware are presented. In particular, WannaCry is examined, during its execution in a purpose-built virtual lab environment, in order to analyze its infection, persistence, recovery prevention and propagation mechanisms. The results obtained may be used for developing appropriate detection and defense solutions for WannaCry and other ransomware families that exhibit similar behaviors.

Keywords—dynamic malware analysis, ransomware, WannaCry.

1. Introduction

Ransomware threat is currently considered to be the main moneymaking scheme for cyber criminals and the key threat to Internet users [1], [2]. In recent years, the appearance of new types of ransomware has been observed, combining the use of worm-like spreading mechanisms and advanced recovery prevention schemes. Recent examples include WannaCry [3], [4] and Petya [5], [6], which exploit the weaknesses of Microsoft Windows, as well as BadRabbit [7], which spreads via insecure compromised websites.

From the defense perspective, the design of new countermeasures is considered, in addition to traditional security approaches, an important and trending task in this field. Such a design, however, requires a comprehensive analysis of ransomware functionality and behavior. This typically involves a wide range of malware analysis tools and techniques. Such techniques may be broadly classified as *static* and *dynamic*. Static analysis is performed without executing the malicious binary, while dynamic analysis involves executing the binary in an isolated environment.

In one of our previous works [8], we performed an initial static and dynamic analysis of WannaCry to identify its resources and functions, as well as its use of dynamic-link libraries (DLLs) and communication protocols. In this work, we have performed a comprehensive dynamic analysis, focusing on WannaCry's infection, persistence, recovery prevention and propagation mechanisms. The techniques presented are also applicable in the cases of other ransomware families whose characteristics are similar to that of WannaCry, such as worm-spreading mechanisms and public-key based encryption. In particular, the research presented examines WannaCry's behavior during its execution in a safe, purpose-built virtual lab environment at the University of York. The results obtained may form a basis for designing and developing effective ransomware defense solutions.

The rest of the paper is organized as follows. In Section 2, we present the relevant background information on ransomware in general and on WannaCry in particular. In Section 3, the main findings from the dynamic analysis of WannaCry we have performed, including its encryption process, recovery prevention and propagation mechanisms, are presented. Finally, Section 4 draws conclusions and discusses potential future directions.

2. Background

2.1. Ransomware

Ransomware is a type of malicious software (malware) that prevents users from accessing or limits their access to the system or files, either by locking the screen or by encrypting files, until a ransom is paid [9]. In most cases, ransomware leaves the user with very few options, such as only allowing the victim to communicate with the attacker and pay the ransom.

The most common types of ransomware use some form of encryption, including both symmetric and public-key based encryption schemes. Ransomware that relies on public-key encryption is particularly difficult to mitigate, since the encryption keys are stored in a remote command and control (C&C) server. There is usually a time limit for ransom to be paid, the users are provided with a special website to purchase cryptocurrency (e.g. Bitcoins) and step-by-step instructions on how to pay the ransom.

The lifecycle of modern day ransomware typically consists of the following steps [10]: distribution, infection, C&C communications, file search, file encryption and ransom demand.

2.2. WannaCry

WannaCry ransomware (also known as Wana Decrypt0r, WCry, WannaCry, WannaCrypt, and WanaCrypt0r) was observed during a massive attack across multiple countries on 12 May 2017 [11]. According to multiple reports from security vendors, the total of 300,000 systems in over 150 countries had been severely damaged. The attack affected a wide range of sectors, including healthcare, government, telecommunications and gas/oil production.

The difficulty in protecting against WannaCry stems from its ability to spread to other systems by using a worm component. This feature makes the attacks more effective and requires defense mechanisms that can react quickly and in real time. Furthermore, WannaCry has an encryption component that is based on public-key cryptography.

During the infection phase, WannaCry uses the *EternalBlue* and *DoublePulsar* exploits that were allegedly leaked in April 2017 by a group called The Shadow Brokers. EternalBlue exploits the server message block (SMB) vulnerability that was patched by Microsoft on March 14, 2017 and has been described in the security bulletin MS17-010 [12]. This vulnerability allows the adversaries to execute a remote code on the infected machines by sending specially crafted messages to an SMB v1 server, connecting to TCP ports 139 and 445 of unpatched Windows systems. In particular, this vulnerability affects all unpatched Windows versions starting from Windows XP to Windows 8.1, except for Windows 10.

DoublePulsar is a persistent backdoor that may be used to access and execute code on previously compromised systems, thus allowing the attackers to install additional malware on the system. During the distribution process, WannaCry's worm component uses EternalBlue for initial infection through the SMB vulnerability, by actively probing appropriate TCP ports and, if successful, tries to implant the DoublePulsar backdoor on the infected systems.

3. WannaCry Analysis

In this section, we present our findings based on the dynamic analysis of WannaCry we have performed. Samples of WannaCry were obtained from VirusShare [13]. Two executable files were analyzed: the worm component and the encryption component (Table 1).

Table 1
WannaCry components

	Worm component
MD5	db349b97c37d22f5ea1d1841e3c89eb4
SHA1	e889544aff85ffaf8b0d0da705105dee7c97fe26
SHA256	24d004a104d4d54034dbcffc2a4b19a11f39008a575aa614ea04703480b1022c
File type	PE32 executable (GUI) Intel 80386, for MS Windows
	Encryption component
MD5	84c82835a5d21bbcf75a61706d8ab549
SHA1	5ff465afaabcbf0150d1a3ab2c2e74f3a4426467
SHA256	ed01ebfbc9eb5bbea545af4d01bf5f1071661840480439c6e5babe8e080e41aa
File type	PE32 executable (GUI) Intel 80386, for MS Windows

3.1. Testbed

In order to analyze WannaCry, a virtual testbed shown in Fig. 1 was built. The characteristics of the host machine are as follows: Intel Core i7-4700MQ 2.40 GHz and 16 GB RAM. The host machine acts as a virtual switch and is running REMnux [14], which is a free Linux toolkit for reverse engineering and malware analysis. Two virtual machines (VMs), running Windows 7 SP1, were used. The first VM was infected with WannaCry, whereas the other VM was clean. A custom network VMnet 5 – 192.168.180.0/24 was created with the Virtual Network Editor option in VMWare hypervisor. This testbed allows observing domain name system (DNS) queries made by WannaCry during the infection and replication process across internal and external

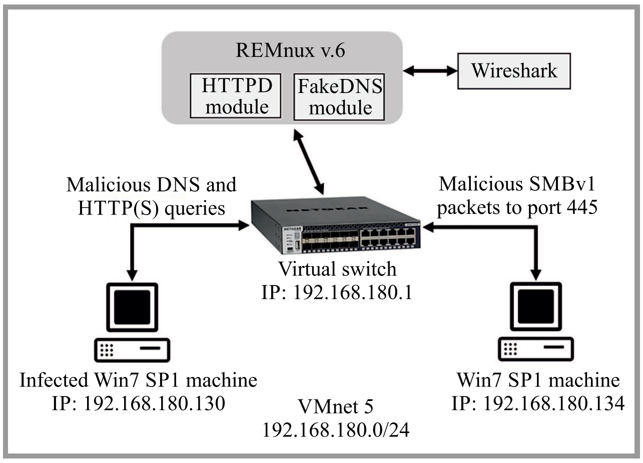


Fig. 1. Testbed for dynamic WannaCry analysis.

networks via port 445 of the SMB v1 protocol. The REMnux machine acts as a DNS and HTTP server, and is able to intercept all network communications using Wireshark. DNS and HTTP services in REMnux were enabled using FakeDNS and HTTP Daemon utilities, respectively.

The system level actions performed by WannaCry were observed on the infected Windows 7 SP1 machine with the 192.168.180.130 IP address. In order to observe and report the actions that WannaCry took while running on the system, the SysAnalyzer tool [15] was used. The main benefit of SysAnalyzer is that it is capable of taking system snapshots before and after malware execution, thus making it possible to inspect system attributes, such as running processes, open ports, DLLs loaded, registry key changes, run time file modifications, scheduled tasks, mutual exclusion objects (mutexes) and network connections. SysAnalyzer is also capable of taking memory dumps and scanning them for specific regular expressions. Before executing the WannaCry sample on the infected machine, the SysAnalyzer's configuration wizard was set to apply a 120 s delay between system snapshots, thus allowing to inspect all system attribute changes.

3.2. Libraries and Functions

Analysis performed with the Pestudio tool [16] revealed that the worm and the encryption components of WannaCry contain DLLs shown in Tables 2 and 3, respectively. During its execution, the worm component invokes *iphlpapi.dll* to retrieve network configuration settings for the infected host. *Kernel32.dll* and *msvcrt.dll* are the two libraries most frequently invoked by the encryption component. This may indicate that the main encryption functionality was implemented by these two malicious libraries. To confirm this, the imported functions of the libraries needed to be examined.

Table 2
DLLs of the worm component

Library	Imports	Description
ws2_32.dll	13	Windows Socket 2.0 32-bit DLL
iphlpapi.dll	2	IP Helper API
wininet.dll	3	Internet Extensions for Win32
kernel32.dll	32	Windows NT Base API Client DLL
advapi32.dll	11	Advanced Windows 32 Base API
msvcp60.dll	2	Windows NT C++ Runtime Library DLL
msvcrt.dll	28	Windows NT CRT DLL

Table 3
DLLs of the encryption component

Library	Imports	Description
kernel32.dll	54	Windows NT Base API Client DLL
advapi32.dll	10	Advanced Windows 32 Base API
user32.dll	1	Multi-User Windows User API Client DLL
msvcrt.dll	49	Windows NT CRT DLL

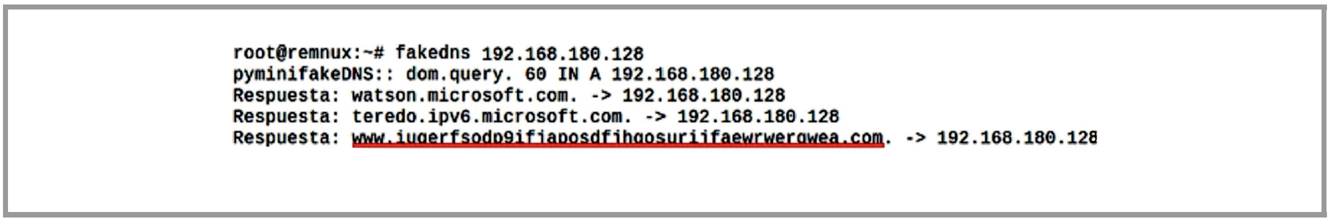
Table 4
Functions of the encryption component

Function	Location
GetCurrentThread	0xa53a
GetStartupInfoA	0xa97a
StartServiceCtrDispatcherA	0xa6f6
RegisterServiceCtrDispatcherA	0xa6d8
CreateServiceA	0xa688
StartServiceA	0xa662
CryptGenRandom	0xa650
CryptAcquireContextA	0xa638
OpenServiceA	0xa714
GetAdaptersInfo	0xa792
InternetOpenUrlA	0xa7c8

Table 5
Functions of the encryption component

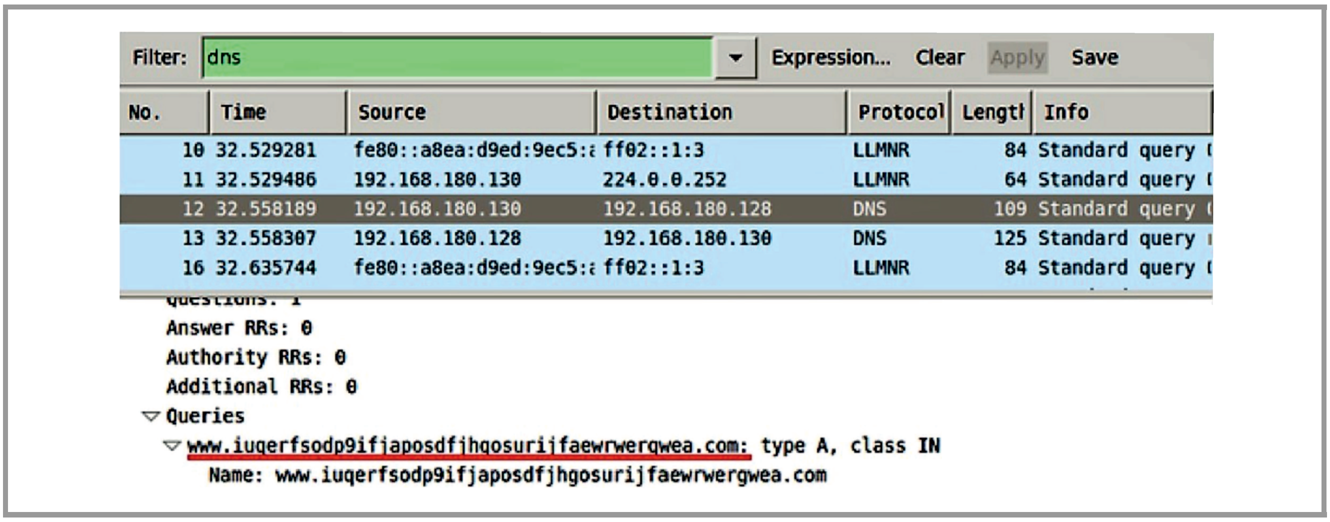
Function	Location
OpenMutexA	0xda84
GetComputerNameW	0xd8b2
CreateServiceA	0xdc2a
OpenServiceA	0xdc62
StartServiceA	0xdc52
CryptReleaseContext	0xdc14
RegCreateKeyW	0xdc04
fopen	0xdcd4
fread	0xdccc
fwrite	0xdcc2
fclose	0xdcb8
CreateFileA	0xd922
ReadFile	0xd964

The imported functions of the samples were observed by Pestudio. The most suspicious functions identified among them are shown in Tables 4 and 5. One may observe that in general, WannaCry uses Microsoft's crypto, file management and C runtime file APIs. The crypto API library is used to generate and manage random symmetric and asymmetric cryptographic keys.



```
root@remnux:~# fakedns 192.168.180.128
pyminifakeDNS:: dom.query. 60 IN A 192.168.180.128
Respuesta: watson.microsoft.com. -> 192.168.180.128
Respuesta: teredo.ipv6.microsoft.com. -> 192.168.180.128
Respuesta: www.iuqerfsodp9ifjaposdfjhgosurijfaewrwergwea.com. -> 192.168.180.128
```

Fig. 2. FakeDNS capture of the malicious DNS request.



No.	Time	Source	Destination	Protocol	Lengtl	Info
10	32.529281	fe80::a8ea:d9ed:9ec5:ε	ff02::1:3	LLMNR	84	Standard query
11	32.529486	192.168.180.130	224.0.0.252	LLMNR	64	Standard query
12	32.558189	192.168.180.130	192.168.180.128	DNS	109	Standard query
13	32.558307	192.168.180.128	192.168.180.130	DNS	125	Standard query
16	32.635744	fe80::a8ea:d9ed:9ec5:ε	ff02::1:3	LLMNR	84	Standard query

```
    Questions: 1
    Answer RRs: 0
    Authority RRs: 0
    Additional RRs: 0
  ▽ Queries
    ▽ www.iuqerfsodp9ifjaposdfjhgosurijfaewrwergwea.com: type A, class IN
        Name: www.iuqerfsodp9ifjaposdfjhgosurijfaewrwergwea.com
```

Fig. 3. Wireshark capture of the malicious DNS request.

3.3. Initial Interactions

The dynamic analysis conducted has revealed that, upon startup, the worm component tries to connect to the following domain, using the *InternetOpenUrl* function:

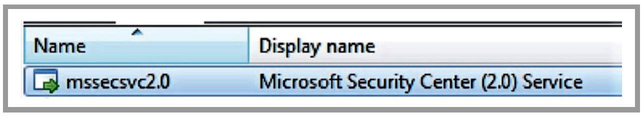
www.iuqerfsodp9ifjaposdfjhgosurijfaewrwergwea.com

The aforementioned domain is a kill-switch domain. This means that if the domain is active, the worm component stops running. On the other hand, if the worm component cannot establish a connection with this domain (e.g. if the domain is not active or if there is no connectivity), it continues to run and registers itself as a "Microsoft Security Center (2.0) Service" *mssecsvs2.0* process on the infected machine. Hence, this kill-switch domain may be used as part of a detection technique when developing a defense system.

The FakeDNS utility at REMnux captures the malicious DNS request on port 80 (Fig. 2), while Wireshark shows (Fig. 3) the DNS packet query field from the infected machine (IP 192.168.180.130) to the DNS server on REMnux (IP 192.168.180.128).

3.4. Persistence Mechanisms

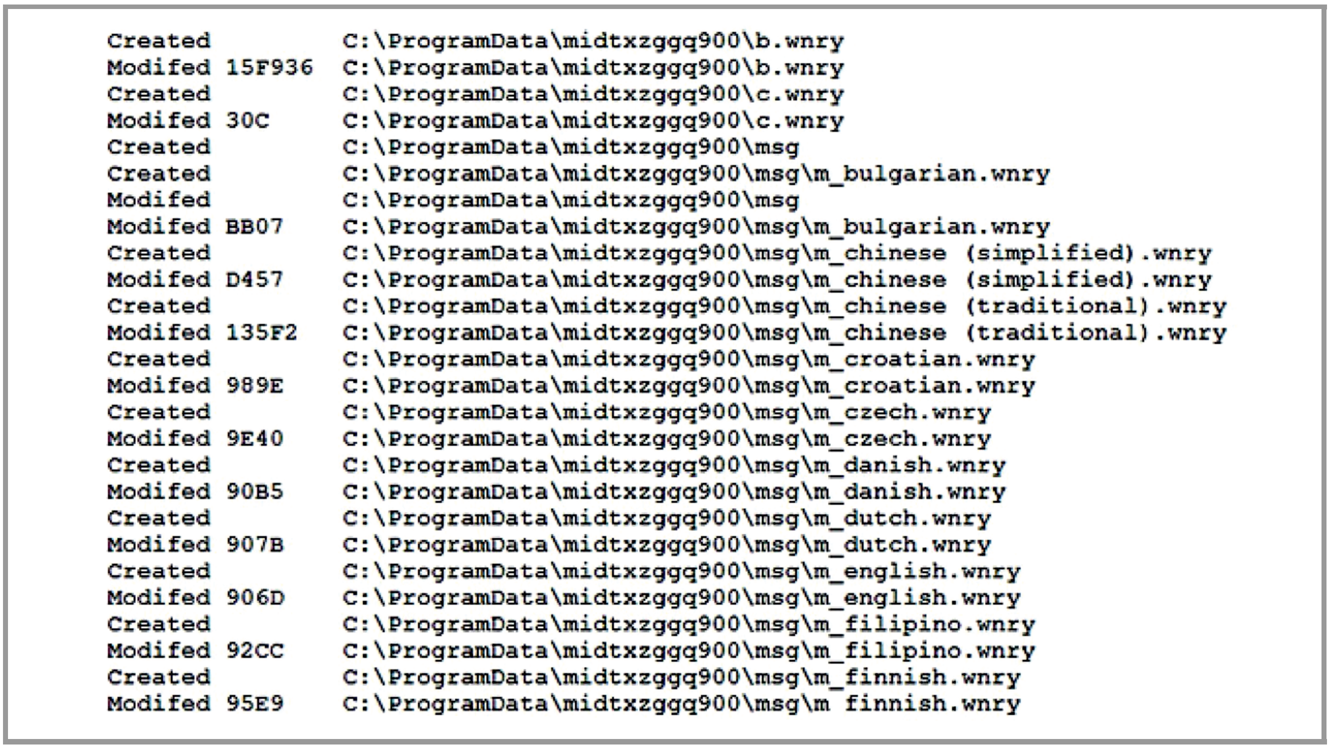
After connection failure with the kill-switch domain, the worm component attempts to create a *mssecsvs2.0* process with the DisplayName of "Microsoft Security Center (2.0) Service". This can be observed in the Process Hacker tool with 4016 PID, indicating that the service has been launched (Fig. 4). In addition to this, the worm component of WannaCry extracts the hardcoded *R resource* binary and then copies it to "C:\Windows\taskche.exe" directory path. The R resource represents the binary of the WannaCry encryption component. After that, the worm runs the executable with the following parameters in the command line: "C:\Windows\taskche.exe/i". Next, the worm tries to move the "C:\Windows\taskche.exe" file to "C:\Windows\qeriuwjhrf", to replace the original file if it exists. This is done to ensure multiple infections and avoid any issues with creating the tasksche.exe process.



Name	Display name
mssecsvc2.0	Microsoft Security Center (2.0) Service

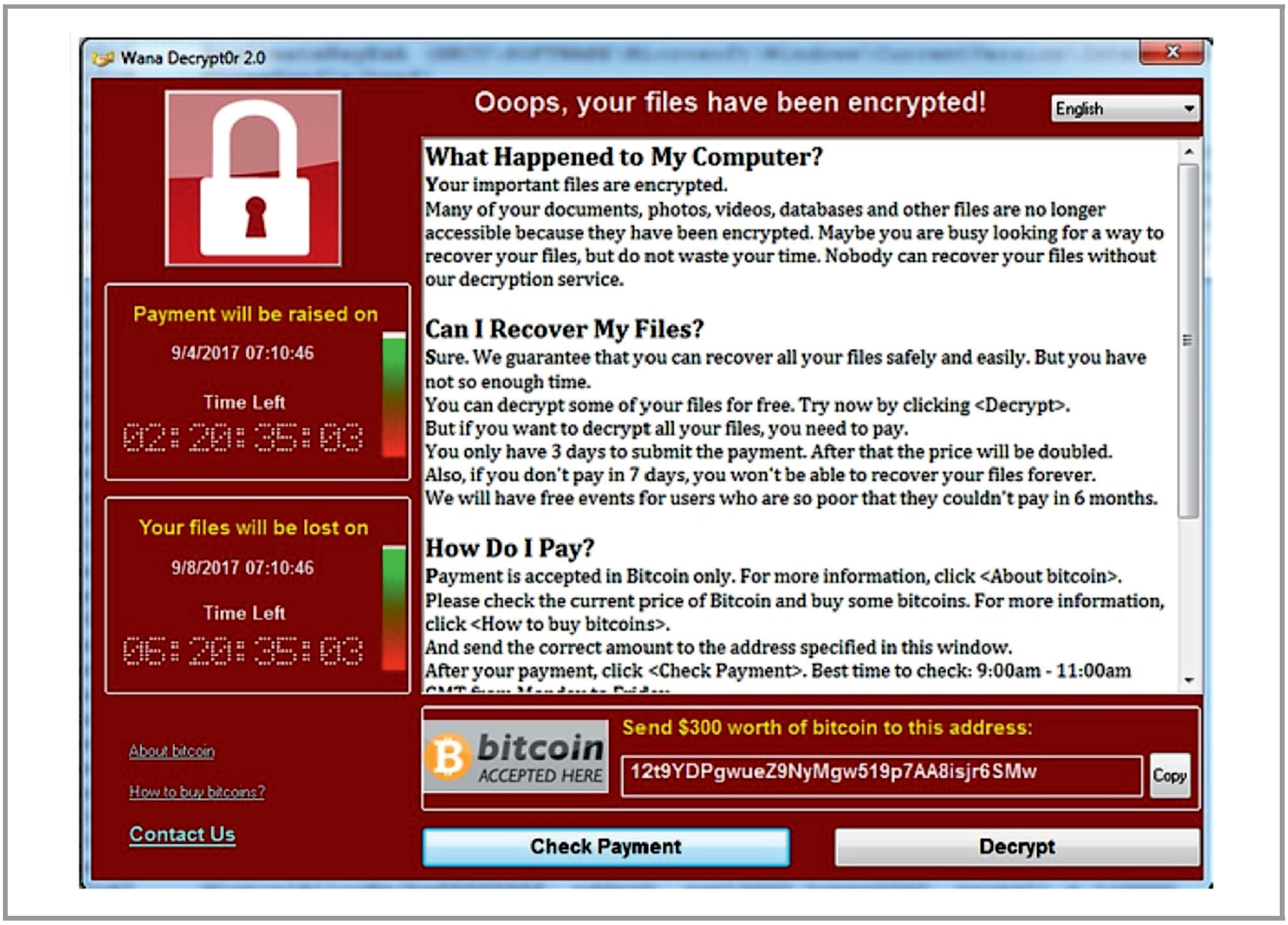
Fig. 4. Microsoft Security Center (2.0) Service.

Finally, WannaCry creates an entry in the Windows registry in order to ensure that it runs every time the computer is restarted. The new entry contains a string (e.g. "midtxzggq900"), which is a unique identifier randomly generated by using the computer name. Once the tasksche.exe component runs, it copies itself to a folder with a randomly generated name in the Common Appdata directory of the infected machine. Then, it attempts to establish memory persistence by adding itself to the AutoRun feature.



```
Created         C:\ProgramData\midtxzgqq900\b.wnry
Modifed 15F936  C:\ProgramData\midtxzgqq900\b.wnry
Created         C:\ProgramData\midtxzgqq900\c.wnry
Modifed 30C     C:\ProgramData\midtxzgqq900\c.wnry
Created         C:\ProgramData\midtxzgqq900\msg
Created         C:\ProgramData\midtxzgqq900\msg\m_bulgarian.wnry
Modifed         C:\ProgramData\midtxzgqq900\msg
Modifed BB07    C:\ProgramData\midtxzgqq900\msg\m_bulgarian.wnry
Created         C:\ProgramData\midtxzgqq900\msg\m_chinese (simplified).wnry
Modifed D457    C:\ProgramData\midtxzgqq900\msg\m_chinese (simplified).wnry
Created         C:\ProgramData\midtxzgqq900\msg\m_chinese (traditional).wnry
Modifed 135F2   C:\ProgramData\midtxzgqq900\msg\m_chinese (traditional).wnry
Created         C:\ProgramData\midtxzgqq900\msg\m_croatian.wnry
Modifed 989E    C:\ProgramData\midtxzgqq900\msg\m_croatian.wnry
Created         C:\ProgramData\midtxzgqq900\msg\m_czech.wnry
Modifed 9E40    C:\ProgramData\midtxzgqq900\msg\m_czech.wnry
Created         C:\ProgramData\midtxzgqq900\msg\m_danish.wnry
Modifed 90B5    C:\ProgramData\midtxzgqq900\msg\m_danish.wnry
Created         C:\ProgramData\midtxzgqq900\msg\m_dutch.wnry
Modifed 907B    C:\ProgramData\midtxzgqq900\msg\m_dutch.wnry
Created         C:\ProgramData\midtxzgqq900\msg\m_english.wnry
Modifed 906D    C:\ProgramData\midtxzgqq900\msg\m_english.wnry
Created         C:\ProgramData\midtxzgqq900\msg\m_filipino.wnry
Modifed 92CC    C:\ProgramData\midtxzgqq900\msg\m_filipino.wnry
Created         C:\ProgramData\midtxzgqq900\msg\m_finnish.wnry
Modifed 95E9    C:\ProgramData\midtxzgqq900\msg\m finnish.wnry
```

Fig. 5. WannaCry dropped files to the working directory.



What Happened to My Computer?

Your important files are encrypted.

Many of your documents, photos, videos, databases and other files are no longer accessible because they have been encrypted. Maybe you are busy looking for a way to recover your files, but do not waste your time. Nobody can recover your files without our decryption service.

Can I Recover My Files?

Sure. We guarantee that you can recover all your files safely and easily. But you have not so enough time.

You can decrypt some of your files for free. Try now by clicking <Decrypt>.

But if you want to decrypt all your files, you need to pay.

You only have 3 days to submit the payment. After that the price will be doubled.

Also, if you don't pay in 7 days, you won't be able to recover your files forever.

We will have free events for users who are so poor that they couldn't pay in 6 months.

How Do I Pay?

Payment is accepted in Bitcoin only. For more information, click <About bitcoin>.

Please check the current price of Bitcoin and buy some bitcoins. For more information, click <How to buy bitcoins>.

And send the correct amount to the address specified in this window.

After your payment, click <Check Payment>. Best time to check: 9:00am - 11:00am GMT from Monday to Friday.

Send $300 worth of bitcoin to this address:

12t9YDPgwueZ9NyMgw519p7AA8isjr6SMw

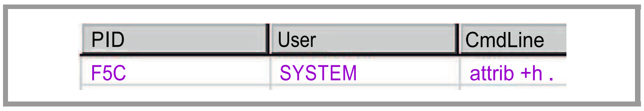
Fig. 6. WannaCry extortion message.

In summary, the dynamic analysis has revealed that, to achieve persistence on the infected machine, WannaCry performs the following actions:

- creates an entry in the Windows registry to ensure that it executes every time the machine is restarted,
- attempts to achieve memory persistence by adding itself to the AutoRun feature of Windows,
- uses Windows *icacls* command to grant itself a full access to all files on the machine,
- deletes all backup (shadow) copies and tries to prevent being booted in *safe mode* by executing several commands in the Windows command line,
- deletes all backup folders,
- by using the Windows command line, creates a VBScript program which generates a single shortcut of the *@WanaDecryptor@.exe* decrypter file,
- tries to kill SQL and MS Exchange database processes by executing several commands in the Windows command line.

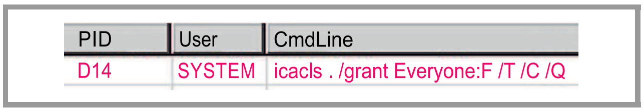
3.5. Configuration Data Load

After the persistence phase, WannaCry loads the *XIA resource*, which corresponds to a password protected ZIP file. It decompresses the files and drops them to the working directory of the running process (Fig. 5), as observed in the DirWatch module of SysAnalyzer.
As one can see, WannaCry loads configuration data from the c.wnry file into memory. WannaCry randomly chooses one of the three available Bitcoin addresses and then writes this address back to the configuration data. This is done in order to display the payment address in the extortion message (Fig. 6). After that, WannaCry sets the hidden attribute (Fig. 7) for the working directory with the help of the CreateProcess function. Next, with the help of the Windows icacls command, WannaCry grants full access to all files on the target system (Fig. 8).



PID	User	CmdLine
F5C	SYSTEM	attrib +h .

Fig. 7. WannaCry sets the hidden attribute for the working directory.



PID	User	CmdLine
D14	SYSTEM	icacls . /grant Everyone:F /T /C /Q

Fig. 8. WannaCry grants full access on the target system.

The next step is to import one of the hardcoded public RSA keys as was identified at offset 0xec00 of the tasksche.exe process (Fig. 9). WannaCry then loads and executes, in memory, the contents of the t.wnry file (Fig. 10) which contains the default encrypted AES key required for decrypting the DLL responsible for the file encryption routine. The first 8 bytes of the file are checked to match the WANACRY! string. Then, the imported public RSA key hardcoded within binary is used to decrypt the AES key stored at the beginning of the t.wnry file. The AES key obtained is then used to decrypt and load the encryption DLL, which can be observed with the help of OllyDbg debugging tool [17] during WannaCry execution, as shown in Fig. 11. This DLL is responsible for file encryption on the infected machine and is summarized in Table 6.

Table 6
Encryption DLL

MD5	f351e1fcca0c4ea05fc44d15a17f8b36
SHA1	7d36a6aa8cb6b504ee9213c200c831e b8d4ef26b
Size	65536 bytes
File type	Dynamic-Link-Library
Internal name	kbdlv.dll
File description	Latvia keyboard layout
Timestamp	Mon, Jul 13 18:12:55 2009

3.6. Encryption Process

The encryption component of WannaCry is invoked with the TaskStart system thread. During its execution, the encryption component checks if one of the following mutexes exists:

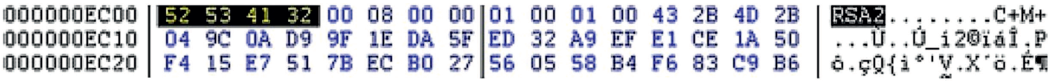
```
GlobalnMsWinZonesCacheCounterMutexA,
GlobalnMsWinZonesCacheCounterMutexW,
MsWinZonesCacheCounterMutexA.
```

If the mutex “MsWinZonesCacheCounterMutexA” is present, then the encryption component automatically stops without taking any further action. If the mutex is not present on the system, the encryption process starts. In particular, TaskStart creates a new mutex named “MsWinZonesCacheCounterMutexA” and reads the contents of the c.wnry file from the current directory. After that, WannaCry creates three configuration files shown in Table 7.

Table 7
WannaCry configuration files

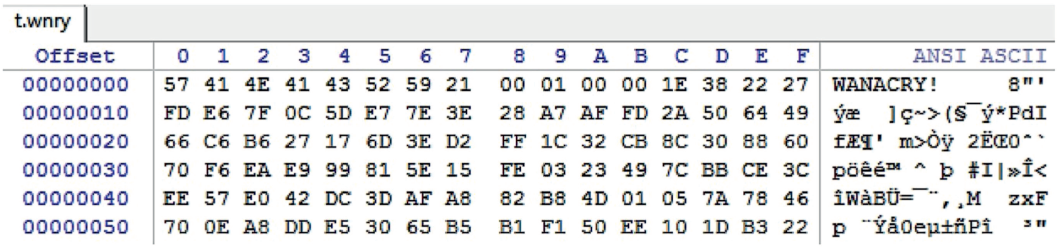
Filename	Description
00000000.res	TOR/C2 info
00000000.pky	Public RSA key
00000000.eky	Encrypted private RSA key

After the configuration files have been created, the encryption component is ready to start encrypting files on the system. To accomplish this, it spawns several threads. First,



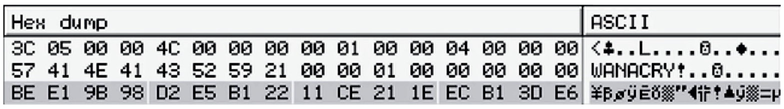
```
000000EC00 | 52 53 41 32 00 08 00 00|01 00 01 00 43 2B 4D 2B | RSA2........C+M+
000000EC10 | 04 9C 0A D9 9F 1E DA 5F|ED 32 A9 EF E1 CE 1A 50 | ...Ù..Ú_í2©ïáÎ.P
000000EC20 | F4 15 E7 51 7B EC B0 27|56 05 58 B4 F6 83 C9 B6 | ô.çQ{ì°'V.X´ö.É¶
```

Fig. 9. Imported RSA private key.



```
t.wnry
Offset     0  1  2  3  4  5  6  7   8  9  A  B  C  D  E  F         ANSI ASCII
00000000  57 41 4E 41 43 52 59 21  00 01 00 00 1E 38 22 27  WANACRY!      8"'
00000010  FD E6 7F 0C 5D E7 7E 3E  28 A7 AF FD 2A 50 64 49  ýæ  ]ç~>(§¯ý*PdI
00000020  66 C6 B6 27 17 6D 3E D2  FF 1C 32 CB 8C 30 88 60  fÆ¶' m>Òÿ 2ËŒ0ˆ`
00000030  70 F6 EA E9 99 81 5E 15  FE 03 23 49 7C BB CE 3C  pöêé™ ^ þ #I|»Î<
00000040  EE 57 E0 42 DC 3D AF A8  82 B8 4D 01 05 7A 78 46  îWàBÜ=¯¨‚¸M  zxF
00000050  70 0E A8 DD E5 30 65 B5  B1 F1 50 EE 10 1D B3 22  p ¨Ýå0eµ±ñPî  ³"
```

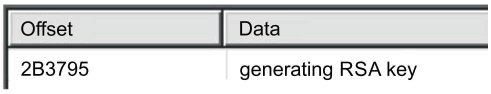
Fig. 10. Loaded and executed t.wnry file.



```
Hex dump                                        | ASCII
3C 05 00 00 4C 00 00 00 00 01 00 00 04 00 00 00 | <♣..L....0..♦...
57 41 4E 41 43 52 59 21 00 00 01 00 00 00 00 00 | WANACRY!..0.....
BE E1 9B 98 D2 E5 B1 22 11 CE 21 1E EC B1 3D E6 | ¥ßøÿÊő▒"◄╬!▲ÿ▒=µ
```

Fig. 11. Decrypted AES key in a memory dump.

WannaCry attempts to load and check the existence of two keys in the 00000000.pky and 00000000.dky files. The 00000000.dky file presents a decryption RSA key which is received upon the payment has been verified. When the victim clicks the "Check Payment" button, WannaCry starts checking for the presence of the 00000000.dky file on the system. If the two aforementioned files do not exist, WannaCry generates a new unique RSA 2048-bit asymmetric key pair, which can be seen in the memory dump made with with SysAnalyzer tool at 0x2B3795 offset (Fig. 12).



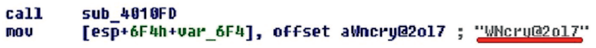
Offset	Data
2B3795	generating RSA key

Fig. 12. Generation of an RSA key pair.

Once the key pair has been generated, WannaCry exports the victim's public RSA key to a 00000000.pky file using Microsoft's *CryptExportKey* function. Next, WannaCry exports the victim's private RSA key and encrypts it with another hard-coded RSA public key. The encrypted private key is stored as a 00000000.eky file. After the key has been safely stored, WannaCry calls upon the *CryptDestroyKey* function to destroy the private key in memory, to limit any key recovery options.

Next, WannaCry starts enumerating, every 3 seconds, information about all logical drives attached to the system. If a new attached drive is not a CD ROM drive, then it begins the encryption process on the new drive. At this stage, WannaCry also starts iterating through all existing directories and searching for predefined file extensions of interest.

To encrypt each file, it generates a 16-byte symmetric AES key using the *CryptGenRandom* function. Then, it encrypts every generated AES key with the public RSA key and stores it inside the file header starting with the WANACRY! string value. Encrypted files are renamed and appended with the .*WNCRY* file extension.



```
call    sub_4010FD
mov     [esp+6F4h+var_6F4], offset aWncry@2ol7 ; "WNcry@2ol7"
```

Fig. 13. Password for a ZIP archive in the encryption component.

The encryption component contains a password-protected ZIP archive. We managed to obtain the password, "WNcry@2ol7", by disassembling the encrypter with the IDA Pro tool [18] (see Fig. 13). The contents of the ZIP archive are summarized in Table 8 and described below:

- *msg* is a folder that contains a list of rich text format (RTF) files with the *wnry* extension. These files are the readme instructions used to show the extortion message to the victim in different languages, based on the information obtained from the system by malicious WannaCry functions;
- *b.wnry* is an image file used for displaying instructions for the decryption of user files. It starts with 42 4D strings, which indicates that this file is a bitmap image;
- *c.wnry* contains a list of Tor addresses with the *.onion* extension and a link to a zipped installation file of the Tor browser from Tor Project [19];

Table 8
Files in the password protected ZIP archive

Name	Size [bytes]	Modified
msg	1,329,657	2017-05-11
b.wnry	1,440,054	2017-05-11
c.wnry	780	2017-05-11
r.wnry	864	2017-05-10
s.wnry	3,038,286	2017-05-09
t.wnry	65,816	2017-05-11
taskdl.exe	20,480	2017-05-11
taskse.exe	20,480	2017-05-11
u.wnry	245,760	2017-05-11

- *r.wnry* is a text file in English with additional decryption instructions to be used by the decryption component (the *u.wnry* file mentioned below);
- *s.wnry* file is a ZIP archive (HEX signature 50 4B 03 04) which contains the Tor software executable. This executable has been obtained with the assistance of the WinHex tool [20] by saving raw binary data with the .zip extension;
- *t.wnry* is an encrypted file with the WANACRY! encryption format. The file header starts with the WANACRY! string;
- *taskdl.exe* is a supporting tool for the deletion of files with the .WNCRY extension. By observing the properties of the file, the following masquerade description can be found: “SQL Client Configuration Utility”;
- *taskse.exe* is a supporting tool for malware execution on remote desktop protocol (RDP) sessions. The following file description was identified: “waitfor – wait/send a signal over a network”;
- *u.wnry* is an executable file (HEX signature 4D 5A) with the name of “@WanaDecryptor@.exe”, which represents the decryption component of WannaCry.

At the same time, another thread calls the taskse.exe process every 30 s, which tries to enumerate active RDP sessions on connected remote machines and to run the @WanaDecryptor@.exe binary file. This file is extracted from the u.wnry file and represents the decryption component of WannaCry. The persistence of RDP session injections is ensured by adding the value in the AutoRun registry key.

3.7. Recovery Prevention

After finishing the encryption process, WannaCry tries to prevent various common data recovery methods by executing several commands on the system. To prevent data recovery, WannaCry executes the following commands:

- vssadmin delete shadows/all/quiet. Deletes all the shadow volumes on the system without alerting the user. By default, these volumes contain backup data in the event of a system fault;
- wmic shadowcopy delete. Ensures deletion of any copies relevant to shadow volumes;
- bcdedit/set default bootstatuspolicy ignoreallfailures. Ensures that the machine is booted, even if errors are found;
- bcdedit/set default recoveryenabled no. Disables the Windows recovery feature, thus preventing the victims from the possibility to reverting their system to a previous build;
- wbadmin delete catalog $-q$. Ensures that victim can no longer use any backup files created by Windows Server.

3.8. Propagation

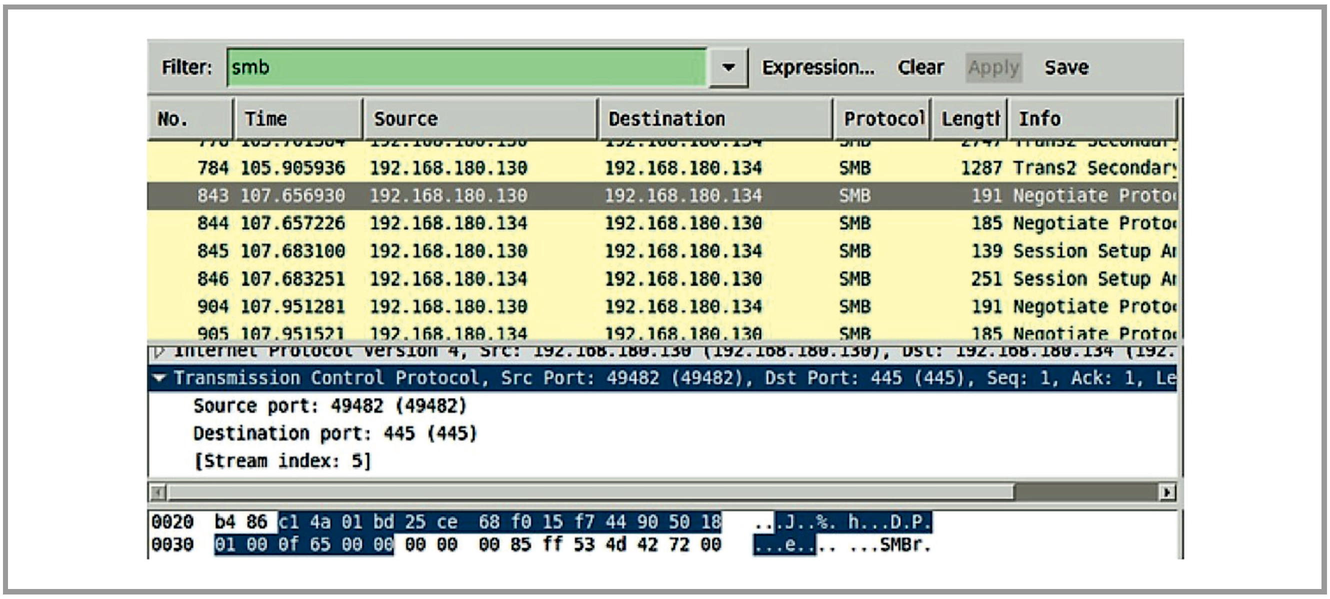
The worm component of WannaCry carries the main propagation and exploit functionality, which utilizes the EternalBlue exploit and the DoublePulsar backdoor to leverage the MS17-010 SMB vulnerability [12]. After performing the initial interactions and checking connectivity with the kill-switch domain, the worm functionality is established by initiating the *mssecsvs2.0* service, which WannaCry installs after being executed. This service tries to spread WannaCry payload through the SMB vulnerability on any vulnerable systems on both internal and external networks.

In order to perform this, WannaCry creates and spawns two separate threads that simultaneously replicate worm payload in all detected networks. In the internal network, before starting the propagation process, the component obtains the IP addresses of local network interfaces by invoking the *GetAdaptersInfo* function, and determines the subnets existing in the network.

After that, the worm component tries to connect to all possible IP addresses in any available local network on port 445, which is the default port for SMB over IP service. If successful, the worm attempts to exploit the service for the MS17-010 vulnerability. In our testbed, connection attempts were observed with Wireshark on a REMnux machine, when the infected machine (IP 192.168.180.130) sent SMB probe packets to the clean machine (IP 192.168.180.134), as shown in Fig. 14.

During the SMB probing, one of the unique features of the generated traffic is that it contains two hardcoded IP addresses: 192.168.56.20 and 172.16.99.5. They can be observed by extracting strings from the binary. In particular, WannaCry sends three NetBIOS session setup packets, where two of them contain the aforementioned hardcoded IP addresses.

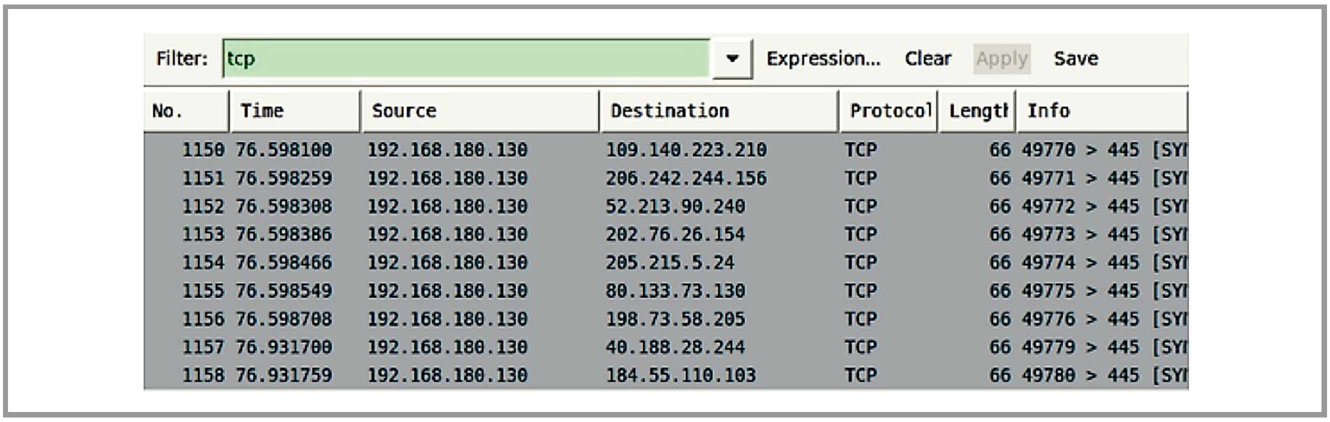
At the same time, the worm component attempts to spread across the external networks by generating various IP addresses and by trying to connect to TCP port 445. This can be observed with Wireshark on REMnux, as shown



```
No.  Time        Source           Destination      Protocol Length Info
 784 105.905936  192.168.180.130  192.168.180.134  SMB      1287 Trans2 Secondary
 843 107.656930  192.168.180.130  192.168.180.134  SMB       191 Negotiate Proto
 844 107.657226  192.168.180.134  192.168.180.130  SMB       185 Negotiate Proto
 845 107.683100  192.168.180.130  192.168.180.134  SMB       139 Session Setup Ar
 846 107.683251  192.168.180.134  192.168.180.130  SMB       251 Session Setup Ar
 904 107.951281  192.168.180.130  192.168.180.134  SMB       191 Negotiate Proto
 905 107.951521  192.168.180.134  192.168.180.130  SMB       185 Negotiate Proto
▷ Internet Protocol Version 4, Src: 192.168.180.130 (192.168.180.130), Dst: 192.168.180.134 (192.
▼ Transmission Control Protocol, Src Port: 49482 (49482), Dst Port: 445 (445), Seq: 1, Ack: 1, Le
    Source port: 49482 (49482)
    Destination port: 445 (445)
    [Stream index: 5]

0020  b4 86 c1 4a 01 bd 25 ce  68 f0 15 f7 44 90 50 18   ...J..%. h...D.P.
0030  01 00 0f 65 00 00 00 00  00 85 ff 53 4d 42 72 00   ...e.... ...SMBr.
```

Fig. 14. WannaCry internal network traffic attempting the SMB exploit.



```
No.   Time       Source           Destination      Protocol Length Info
1150  76.598100  192.168.180.130  109.140.223.210  TCP      66 49770 > 445 [SYN
1151  76.598259  192.168.180.130  206.242.244.156  TCP      66 49771 > 445 [SYN
1152  76.598308  192.168.180.130  52.213.90.240    TCP      66 49772 > 445 [SYN
1153  76.598386  192.168.180.130  202.76.26.154    TCP      66 49773 > 445 [SYN
1154  76.598466  192.168.180.130  205.215.5.24     TCP      66 49774 > 445 [SYN
1155  76.598549  192.168.180.130  80.133.73.130    TCP      66 49775 > 445 [SYN
1156  76.598708  192.168.180.130  198.73.58.205    TCP      66 49776 > 445 [SYN
1157  76.931700  192.168.180.130  40.188.28.244    TCP      66 49779 > 445 [SYN
1158  76.931759  192.168.180.130  184.55.110.103   TCP      66 49780 > 445 [SYN
```

Fig. 15. WannaCry external network traffic attempting the SMB exploit.

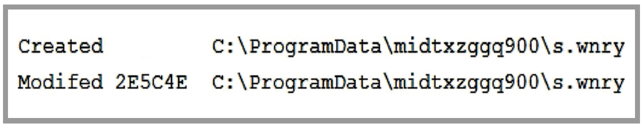
in Fig. 15. As it can be seen, the worm attempts to probe external Internet IP addresses for the MS17-010 vulnerability. This explains the reason for the widespread infection seen during the massive outbreak on 12 May 2017. The full list of WannaCry generated IP addresses obtained during the analysis is presented in Table 9.

Table 9
External IP addresses generated
by WannaCry

IP address : port
109.140.223.210 : 445
206.242.244.156 : 445
52.213.90.240 : 445
202.76.26.154 : 445
205.215.5.24 : 445
80.133.73.130 : 445
198.73.58.205 : 445
40.188.28.244 : 445
184.55.110.103 : 445

3.9. C&C Communication

During its execution, the software also tries to contact the C&C servers. To this end, WannaCry unpacked and dropped files from the s.wnry file, containing the Tor executable, into the installation directory as shown in Fig. 16. Before unpacking, it starts listening on the localhost address 127.0.0.1:9050. This address, with the specified 9050 port, is typically used for configuring the



```
Created          C:\ProgramData\midtxzggq900\s.wnry
Modifed 2E5C4E   C:\ProgramData\midtxzggq900\s.wnry
```

Fig. 16. Tor executable dropped into the installation directory.

Tor browser application. If the contents of the s.wnry file are corrupted, then WannaCry tries to download the Tor executable from a hardcoded URL. After the successful extraction of the Tor executable, it copies "TaskData\Tor\tor.exe" to "TaskData\Tor\taskhsvc.exe" and executes it. Next, WannaCry parses the contents of the c.wnry file, which specifies the configuration data, including the following .onion addresses to connect and the zipped Tor browser installation file:

```
gx7ekbenv2riucmf.onion
57g7spgrzlojinas.onion
xxlvbrloxvriy2c5.onion
76jdd2ir2embyv47.onion
cwwnhwhlz52maqm7.onion
https://dist.torporject.org/torbrowser/6.5.1/tor
        -win32-0.2.9.10.zip
```

After that, WannaCry sends the first eight bytes of the 00000000.res file content to the C&C server. These bytes specify the host and user name of the infected machine. The 00000000.res file, which is dropped during encryption process, accumulates in total 88 bytes of configuration data, including internal flags, counters, and timestamps.

During its communication with Tor addresses, WannaCry establishes a secure HTTPS channel to port 443, and uses common Tor ports, 9001 and 9050, for network traffic and directory information.

4. Conclusions and Future Work

We have performed a comprehensive dynamic analysis of WannaCry ransomware in a purpose-built virtual testbed. We analyzed the WannaCry version which was observed during the massive attacks on 12 May 2017. The analysis has revealed that the given ransomware is composed of two distinctive components, which enable the worm-like self-propagating mechanism and combined encryption process. Both worm and encryption components of WannaCry have been examined.

The focus of this study was on WannaCry's initial interactions and the infection process, its persistence mechanism, encryption process, recovery prevention as well as its propagation mechanisms and communication with C&C servers. The analysis has revealed important characteristics and behaviors of WannaCry during its execution. In particular, we identified Tor addresses used for C&C, observed TCP and DNS connections, SMB probes, as well as actions related to WannaCry persistence and obfuscation.

The worm component of WannaCry weaponized by the functionality enabling it to exploit and propagate via Microsoft's MS17-010 on unpatched systems by sending SMB probing packets on port 445. In addition to the modular nature of WannaCry, it was also observed that it has embedded RSA keys used to decrypt the required malicious DLL representing the encryption component. It was identified that the worm component scans both internal and external networks for MS17-010 vulnerability, by generating a list of local and global IP addresses. The worm tries to probe the hosts from the generated list by sending packets to port 445. Before its execution, WannaCry also performs an initial check with the kill-switch domain.

At the same time, the analysis has identified two hardcoded IP addresses (192.168.56.20 and 172.16.99.5), which are sent during the SMB probing. Depending on the condition of the s.wnry file dropped during execution, WannaCry can also communicate with embedded .onion addresses via a secure channel on port 443 and via common Tor ports 900 and 9050 to download the Tor browser installation software from a specified URL.

The findings of this work could be used for designing effective mitigation mechanisms for WannaCry and other ransomware families that exhibit similar behavior. This is left as future work. In particular, we plan to investigate the use of software-defining networking (SDN) [21], [22] for ransomware detection and mitigation. SDN is an emerging paradigm of programmable networks that decouples the control and data planes. SDN controllers maintain a view of the entire network and implement policy decisions. On the other hand, each device at the data plane maintains one or more *flow tables*, where the packet handling rules are stored. This changes the way that networks are designed and managed, and enables new SDN-based security solutions [23]–[25], such as firewalls and intrusion detection systems for various types of malware, including ransomware mitigation [26], [27].

References

[1] D. O'Brien, "Ransomware 2017", Internet Security Threat Report, Symantec, July 2017 [Online]. Available: https://www.symantec.com/content/dam/symantec/docs/security-center/white-papers/istr-ransomware-2017-en.pdf

[2] K. Savage, P. Coogan, and H. Lau, "The evolution of ransomware", Security Response, Symantec, June 2015 [Online]. Available: http://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/the-evolution-of-ransomware.pdf

[3] A. Zeichnick, "Self-propagating ransomware: What the WannaCry ransomworm means for you", May 2017 [Online]. Available: https://www.networkworld.com/article/3196993/security/self-propagating-ransomware-what-the-wannacry-ransomworm-means-for-you.html

[4] "Ransom.Wannacry", Symantec, May 2017 [Online]. Available: https://www.symantec.com/security-center/writeup/2017-051310-3522-99/

[5] "Petya – taking ransomware to the low level", Malwarebytes Labs, Jun. 2017 [Online]. Available: https://blog.malwarebytes.com/threat-analysis/2016/04/petya-ransomware/

[6] "Petya ransomware eats your hard drives", Kaspersky Labs, Jun. 2017 [Online]. Available: https://www.kaspersky.com/blog/petya-ransomware/11715

[7] "Bad Rabbit: A new ransomware epidemic is on the rise", Kaspersky Labs, Oct. 2017 [Online]. Available: https://www.kaspersky.com/blog/bad-rabbit-ransomware/19887/

[8] M. Akbanov, V. G. Vassilakis, I. D. Moscholios, and M. D. Logothetis, "Static and dynamic analysis of WannaCry ransmware", in *Proc. IEICE Inform. and Commun. Technol. Forum ICTF 2018*, Graz, Austria, 2018.

[9] C. Everett, "Ransomware: To pay or not to pay?", *Comp. Fraud & Secur.*, vol. 2016, no. 4, pp. 8–12, 2016 (doi: 10.1016/S1361-3723(16)30036-7).

[10] "Understanding ransomware and strategies to defeat it", McAfee Labs, White Paper, 2016 [Online]. Available: https://www.mcafee.com/enterprise/en-us/assets/white-papers/wp-understanding-ransomware-strategies-defeat.pdf

[11] "What you need to know about the WannaCry ransomware", Symantec, Threat Intelligence, Oct. 2017, [Online]. Available: https://www.symantec.com/blogs/threat-intelligence/wannacry-ransomware-attack

[12] Microsoft Security Bulletin MS17-010 – Critical, March 14, 2017 [Online]. Available: https://docs.microsoft.com/en-us/security-updates/securitybulletins/2017/ms17-010

[13] ViRus Share malware repository [Online]. Available: https://virusshare.com (accessed Nov. 30, 2018).

[14] "REMnux: A Linux toolkit for reverse-engineering and analyzing malware" [Online]. Available: https://remnux.org (accessed Nov. 30, 2018).

[15] SysAnalyzer – Automated malcode analysis system [Online]. Available: https://github.com/dzzie/SysAnalyzer (accessed Nov. 30, 2018).

[16] Pestudio, Malware Assessment Tool [Online]. Available: https://www.winitor.com (accessed Nov. 30, 2018).

[17] OllyDbg – A 32-bit assembler level debugger for Microsoft Windows [Online]. Available: http://www.ollydbg.de/ (accessed Nov. 30, 2018).

[18] IDA: Pro [Online]. Available: https://www.hex-rays.com/products/ida (accessed Nov. 30, 2018).

[19] Tor Project [Online]. Available: https://www.torproject.org (accessed Nov. 30, 2018).

[20] "WinHex: Computer forensics and data recovery software" [Online]. Available: https://www.x-ways.net/winhex (accessed Nov. 30, 2018).

[21] B. Nunes, M. Mendonca, X. N. Nguyen, K. Obraczka, and T. Turletti, "A survey of software-defined networking: Past, present, future of programmable networks", *IEEE Commun. Surveys & Tutor.*, vol. 16, no. 3, pp. 1617-1634, 2014 (doi: 10.1109/SURV.2014.012214.00180).

[22] V. G. Vassilakis, I. D. Moscholios, B. A. Alzahrani, and M. D. Logothetis, "A software-defined architecture for next-generation cellular networks", in *Proc. IEEE Int. Conf. on Commun. ICC 2016*, Kuala Lumpur, Malaysia, 2016 (doi: 10.1109/ICC.2016.7511018).

[23] C. Yoon, T. Park, S. Lee, H. Kang, S. Shin, and Z. Zhang, "Enabling security functions with SDN: A feasibility study", *Comp. Netw.*, vol. 85, pp. 19–35, 2015 (doi: 10.1016/j.comnet.2015.05.005).

[24] J. M. Ceron, C. B. Margi, and L. Z. Granville, "MARS: An SDN-based malware analysis solution", *Proc. IEEE Symp. on Comp. and Commun. ISCC 2016*, Messina, Italy, 2016 (doi: 10.1109/ISCC.2016.7543792).

[25] V. G. Vassilakis, I. D. Moscholios, B. A. Alzahrani, and M. D. Logothetis, "On the security of software-defined next-generation cellular networks", in *Proc. IEICE Inform. and Commun. Technol. Forum ICTF 2016*, Patras, Greece, 2016.

[26] K. Cabaj and W. Mazurczyk, "Using software-defined networking for ransomware mitigation: The case of CryptoWall", *IEEE Network*, vol. 30, no. 6, pp. 14–20, 2016 (doi: 10.1109/MNET.2016.1600110NM).

[27] K. Cabaj, M. Gregorczyk, and W. Mazurczyk, "Software-defined networking-based crypto ransomware detection using HTTP traffic characteristics", *Comp. & Elec. Engin.*, vol. 66, pp. 353–386, 2018 (doi: 10.1016/j.compeleceng.2017.10.012).

Maxat Akbanov received the B.Sc. degree in Information and Communications System Security from the National Technical University of Ukraine "Kyiv Polytechnic University", Kyiv, Ukraine, in 2011, and the M.Sc. degree in Cyber Security from the University of York, York, UK, in 2018. In 2008 and 2016, he received the prestigious Kazakhstan governmental "Bolashak" scholarship to fund his studies abroad. He holds merit and distinction awards for B.Sc. and M.Sc. degrees, respectively. He is currently working for the private sector in Kazakhstan and is involved in developing several startup projects for the government-sponsored "Digital Kazakhstan" and "Cyber Shield" strategies. His main research interests include network and malware forensics, software-defined networking, covert channels, cryptography, Internet of Things, machine learning and artificial intelligence.

E-mail: maxat.akbanov@gmail.com
Department of Computer Science
University of York
Deramore Lane
Heslington
York YO10 5GH, United Kingdom

Vassilios G. Vassilakis received his Ph.D. degree in Electrical and Computer Engineering from the University of Patras, Greece in 2011. He is currently a lecturer in Cyber Security at the University of York, UK. He's been involved in EU, UK, and industry funded R&D projects related to the design and analysis of future mobile networks and Internet technologies. His main research interests are in the areas of network security, Internet of Things, next-generation wireless and mobile networks, and software-defined networks. He has published over 90 papers in international journals/conferences. He has served as a Guest Editor in IEICE Transactions on Communications, IET Networks, and Elsevier Optical Switching & Networking, and in the TPC of IEEE ICC and IEEE Globecom.

E-mail: vv274@cl.cam.ac.uk
University of York
York YO10 5DD, United Kingdom

Michael D. Logothetis received his B.Eng. degree and Ph.D. in Electrical Engineering, both from the University of Patras, Patras, Greece, in 1981 and 1990, respectively. From 1991 to 1992 he was a Research Associate at NTT's Telecommunication Networks Laboratories, Tokyo, Japan. In 2009 he was elected (Full) Professor at the ECE Department of the University of Patras. His research interests include teletraffic theory, simulation and performance optimization of telecommunications networks. He has published over 200 conference/journal papers. He has become a Guest Editor in: Mediterranean Journal of Electronics and Communications, Mediterranean Journal of Computers and Networks, IET Circuits, Devices and Systems, IET Networks and Ubiquitous Computing and Communication Journal. He is a member of the IARIA (Fellow), IEEE (Senior), IEICE (Senior), FITCE and the Technical Chamber of Greece (TEE).

E-mail: mlogo@upatras.gr
Wire Communications Laboratory
Department of Electrical and Computer Engineering
University of Patras
265 04 Patras, Greece

Paper

Theoretical and Experimental Analysis of Cryptographic Hash Functions

Jacek Tchórzewski[1,2] and Agnieszka Jakóbik[2]

[1] *AGH University of Science and Technology, Cracow, Poland*
[2] *Cracow University of Technology, Cracow, Poland*

https://doi.org/10.26636/jtit.2019.128018

Abstract—The paper presents a theoretical introduction to the cryptographic hash function theory and a statistical experimental analysis of selected hash functions. The definition of hash functions, differences between them, their strengths and weaknesses are explained as well. Different hash function types, classes and parameters are described. The features of hash functions are analyzed by performing statistical analysis. Experimental analysis is performed for three certified hash functions: SHA1-160, SHA2-512 and SHA3-512. Such an analysis helps understand the behavior of cryptographic hash functions and may be very helpful for comparing the security level of the hashing method selected. The tests may serve as a basis for examination of each newly proposed hash function. Additionally, the analysis may be harness as a method for comparing future proposals with the existing functions.

Keywords—cryptographic hash function, hashing metod, security.

1. Introduction

As they play an important role in ensuring the security and confidentiality of information, identification and authentication methods are approached with an ever greater attention, both in civilian (personal information, passwords, PIN codes) and military domains. Hashing is one of the techniques enabling to meet some of the demands described above. Practical applications of cryptographic hash functions include message integrity checking, digital signatures, authentication procedures and other information security-related applications.

The paper is organized as follows. In Section 2 we describe the properties of one way functions, as well as the properties and classes of hash functions. In Section 3, methods of creating hashing functions are presented. In Section 4, the strengths and weaknesses of hashing functions certificated by NIST are presented. Section 5 is devoted to statistical tests involving SHA1-160, SHA-512 and SHA3-512, with their results described. Section 6 summarizes the work and offers conclusions.

2. Hash Functions: Properties, Classes and Types

Let us start with the definition of a one way function, which is given below [1]:

$$\forall x \in X, f : x \to y \land \neg(\exists g : y \to x) \ . \qquad (1)$$

It means that for all function arguments x there exist a value y, but it is impossible to identify a function which will assume this value y as an argument and return x. Hash functions belong to family of one way functions, but are bound by an additional restriction. Formally, they are defined as follows [2]:

$$h : \{0,1\}^* \to \{0,1\}^n, n \geq 1 \ , \qquad (2)$$

where $\{0,1\}^*$ is an input set (formally its elements may be of any length), and will be further denoted by M. Elements from M will be denoted by m ($m \in M$). $\{0,1\}^n$ is an set of output hashes, each with a fixed length and a finite number of combinations, and will be further denoted by H (note, that n is greater than or equal to 1) [2]. Hashes from H will be denoted by $h(m)$ ($h(m) \in H$).

In article [3] Carter and Wegman presented three basic hash function classes.

2.1. Universal Hash Functions Classes

Class H1 is designated for computers which are capable of fast multiplication of the input bit string. Hashes from this class may become inconvenient when the input bit string is too long to multiply it in a single machine instruction. The basic formula of hashes from this group is: for 2 elements, let us call them m and n, the hash is calculated as follows [3]:

$$h_{m,n}(x) = (mx + n) \mod p \ . \qquad (3)$$

In class H3, only simple linear transformations are used instead of multiplication. Formally, the class is defined as follows: if the hash function transforms elements from set A (each element is a binary number with length i) to set B (each element is a binary number with length j), M is an array of size i and contains elements from B, and m are elements of M ($m \in M$) where $m(k)$ is the k-th bit of element m, then for any $x \in A$ (with the same bit indexing as m), the hash function is represented by [3]:

$$h_m(x) = x_1 m(1) \oplus x_2 m(2) \oplus \cdots \oplus x_i m(i). \tag{4}$$

Then H3 is a set defined in the following manner [3]:

$$\{f_m : m \in M\} \ . \tag{5}$$

Class H2 is very similar to class H3. The difference is that hashing functions from this class require more space for hash computation, but need less time. The key point is to find a function g which maps an input bit string into a longer input bit stream containing fewer less '1s'. Then, H2 can be defined [3]:

$$\{f * g : f \in H3\} \ . \tag{6}$$

To define when a set of hashing functions becomes universal, we have to introduce a certain notation. Let us consider hash function h which maps set A into set B. It is always assumed that $|A| > |B|$. Then, it is possible to define function δ_h in the following manner [3]:

$$\delta_h(x,y) = \begin{cases} 1 & \text{if } x \neq y \text{ and } h(x) = h(y) \\ 0 & \text{otherwise} \end{cases} \ , \tag{7}$$

where $x, y \in A$. We can say that collection of hashing functions C is universal when for all x and y in A $\delta_C(x,y) \leq \frac{|C|}{|B|}$ [3]. In practice, this means that no pair of distinct inputs from A collides under more than $(\frac{1}{|B|})$–th of the functions [3]. All three classes (H1, H2, and H3) are universal, but H2 and H3 classes are the most popular ones [3].

2.2. Hash Function Types

Cryptographic hash functions may be divided into two groups [2]:

- keyed hash functions – require a secret key and are known as message authentication code (MAC) [2],
- un-keyed hash functions – do not require any secret key and may be referred to as manipulation detection code (MDC).

  Generally, the term hash functions refer to un-keyed hash functions [2].

In this paper, we will focus on un-keyed hash functions which can be divided into three subgroups, based on their additional properties:

1. One way hash functions (OWHF) – defined by Merkle [4] and fulfilling the following requirements:
   - hash function does not give any constraint on input data size,
   - output hash has constant length,
   - output hash should be easy to compute,
   - "given h and $h(x)$, it is computationally infeasible to determine x" – a preimage resistance feature,
   - "given h and x, it is computationally infeasible to find an $x' \neq x$ such feature that $h(x) = h(x')$" – the second preimage resistance.
2. Collision resistant hash functions (CRHF) – belonging to the OWHF group and fulfilling an additional requirement: it is impossible to find a pair (x, x') where $x \neq x'$, which have the same hash value ($h(x) = h(x')$). This condition is known as collision resistance. The difference between the second preimage resistance depends on the selection of arguments. In the second preimage resistance condition, the attacker has a given value x and has to find x'. In the collision resistance condition, the selection of both: x and x' is a free choice of the attacker.
3. Universal one way hash functions – a family in which the probability of finding a second preimage for a randomly chosen hash function is negligible [2], [5]. These functions are faster than CRHF and allow to omit trapdoors during digital signature creation. They are used when it is impossible to make a decision in which the hash function should be chosen before computation starts.

3. MDC Construction Method

3.1. Hash Function Based on Block Ciphers

To describe the general concept of creation of hash functions based on block ciphers, the following set have to be defined [6], [7]:

$$S \in \{M_i, M_i \text{ XOR } X_i, X_i, C\} \ , \tag{8}$$

where M_i is one block of a message, X_i is a chaining value from the previous step and C is a chosen constant value [7]. Note that all these values are given for i-th round of hash computation, and that secure block cipher B was already chosen. Then the construction of i-th round is:

1. Choose a private key P for B from set S.
2. Choose an input I from set S.
3. Pass I and P to the algorithm B and calculate cipher value CV.
4. Choose value T from set S.
5. Calculate $X_{i+1} = T \oplus CV$.
6. Update set S with values X_{i+1} and M_{i+1} (next block of message) according to the formula (8). If it is impossible, computation ends.
7. Go to step 1.

The selection of variables depends on the algorithm design, but at least one variable should be M_i. The output hash should be as big as block size of B, or twice as big [6]. This is caused by the small size (mainly 64 bits) of the block. Hash should be bigger to avoid collisions. The speed of hash functions based on block ciphers is equal to the number of encryptions to process r plaintext bits, where r is defined as block size [6].
Most hash functions constructed in this way suffer from numerous security vulnerabilities [7] and cannot be used in practice. However, an opposite situation may occur when the block cipher construction is based on the hash function, for example in SHACAL and SHACAL-2, with both being based on the SHA-1 cryptographic hash function [2]. A good example of a hash function with its length equal to the size of the block is described by Meyer and Oseas in [8]. More examples may be found in [6].

3.2. Hash Functions Based on Cellular Automata

Cellular automata (CA) can be used for ciphers generations and for hash functions design [7], since Wolfram [9] developed a pseudorandom generator based on CA rule 30. Cellhas, as described in [10], is a good example of a hashing function based on CA.

3.3. Hash Functions Based on Math

There are three ways of creating of hashing functions based on mathematical constructions:

1. Hashing function based on mathematical primitives is based on modular arithmetic, discrete logarithm problem and factorization problem [11].

    - Factorization problem is based on the difficulty of finding two factors, for any positive integer, which, when multiplied, will give these integers. This problem can be also described by the following formula: for a given integer I it is hard to find a and b such that $ab = I$.
    - Discrete logarithm problem, such as that for a given abelian group O, generator of this group o and an element e which belongs to O, finding (if it is possible) x such that $o^x = e$. The difficulty of the discrete logarithm problem depends on group O.

2. Hashing function based on Knapsack NP-complete problem. From a cryptographic point of view its formula may be formulated as [6]:

$$\sum_{i=1}^{n} a_i \cdot x_i = S \ , \tag{9}$$

where each a_i is a m bit integers: $\{a_1, a_2, \ldots, a_n\}$, S is a p bit integer and $p \approx m + \log_2 n$, and X is a vector of elements $x_i \in \{0,1\}$, [6], [7] and [11].

3. Hashing function based on algebraic matrices, developed by Harari [12]. Here, the key K is a $n \times n$ random matrix and M is the $1 \times n$ message matrix. Then the digest D is: $M^T KM$ or equivalently $K^T MK$ [7], [12]:

$$D = M^T KM. \tag{10}$$

or

$$D = K^T MK. \tag{11}$$

Unfortunately, collisions appeared in the Harari hash function proposition [7].

3.4. Dedicated Hash Functions

Dedicated hash functions created only for hashing operation. Their security can be proved mainly in an empirical way, because they very often do not base on any hard problem, like factorization or discrete logarithm problem. Examples are MD4, MD5, SHA1, SHA2 or SHA3. Based on them algorithms were designed to be as fast as possible in software implementations rather than hardware [11].

3.5. Standardization of Hashing Functions

After the first collision for MD5 was discovered, the National Institute of Standard and Technology, USA, created a hashing standard – Secure Hash Algorithm (SHA). The first version of SHA, known as SHA-0, was published in 1993. In 1995 SHA-0 was replaced by a new version – SHA-1. In 2005 vulnerabilities were identified in SHA-1 and NIST introduced SHA-2, which is used currently. In 2007 an open competition for the next generation SHA-3 was announced. The evaluation criteria are as follows [13]:

- applications of the hash functions – the wider variety of cryptographic usage, the better. The new standard should be useful for the creation of hashed message authentication code (HMAC), as well as for the creation of digital signatures or random bit generators [14].
- specific requirements when hash functions are used to support HMAC, pseudo-random functions (PRFs), or randomized hashing – each algorithm had to have at least one scheme to support HMAC as PRFs [14]. These PRFs have to be secure against known attacks which require less than $2^{\frac{n}{2}}$ queries or which require less computation then the preimage attack [14]. If the hashing function is capable of randomized hashing, it has to have n security bits against attacks mentioned in [14].
- additional security requirements of hash functions – for a digest with size n: $\frac{n}{2}$ bits of collision resistance, n bits of preimage resistance and $n-k$ bits second preimage resistance for any message shorter than 2^k bits [14]. All these rules should be fulfilled with m replacing n for any m size subset taken from the digest [14].

- evaluation of attack resistance- hashing functions were attacked with well-known and popular methods discovered during the security evaluation phase. Other validation methods were based on statistical and behavioral tests, such randomness of hash creation [14].
- other consideration factors – for example quality of security proofs, proper analysis, documentation and simplicity of the algorithm, as well as opinions by NIST and the cryptographic community.

The remaining criteria included speed of the algorithm, code size, memory and hardware implementation requirements, flexibility and simplicity [13]. The final report announcing the winner (the Keccak algorithm) was published in 2012 [15].

4. Theoretical Analysis of Security Parameters

In this section, we will present the results of an analysis of the dependence between the length of the digested messages and the security parameters of hashing functions.

The security level of a cryptographic primitive is expressed in bits, where n-bit security means that the attacker would have to perform 2^n operations to break it. The security level of a cryptographic hash function has been defined using the following properties:

- collision resistance bits of security,
- preimage attacks bits of security,
- second preimage resistance bits of security.

To compromise collision resistance using the brute force method, the attacker needs to hash a huge number of variants of the message m, and hash a huge number of variants m', go through the lists and see if there are values that are equal. For example, in a 160-bit hash output, the attacker needs 2^{80} inputs to test in both lists. Therefore, in the case of this hash function, the number of bits of security against this attack is equal to 80, due to the Birthday Paradox.

While breaking preimage and second preimage resistance, the attacker cannot apply the Birthday Paradox. For a mbox160-bit hash output, the attacker needs to examine 2^{160} input messages, which means that 160 bits of security are achieved.

In Table 1, security parameters of selected hashing functions, as accepted by NIST, are presented [16]. In Table 1, function $\mathrm{L}(M)$ is defined as:

$$L(M) = \left\lceil \log_2 \frac{\mathrm{len}(M)}{B} \right\rceil . \tag{12}$$

where M is the input message, B is the block size of the hash function and $\lceil . \rceil$ denotes the least integer not strictly lower than the argument in the brackets.

Table 1
NIST-approved security parameters of hash functions

		Bits of security		
Function	Output size	Collision	Pre-image	Second preimage
SHA-1	160	< 80	160	160-L(M)
SHA-224	224	112	224	min [224, 256-L(M)]
SHA-256	256	128	256	256-L(M)
SHA-384	384	192	384	384
SHA-512	512	256	512	512-L(M)
SHA3-224	224	112	224	224
SHA3-256	256	128	256	256
SHA3-384	384	192	384	384
SHA3-512	512	256	512	512

Using the brute force method, there always exists a generic attack comprising $2^{\frac{n}{2}}$, 2^n and 2^n steps, respectively, where n is the hash length [17]. This is the maximum (ideal) security level which can be achieved for any hash function. As it can be seen in Table 1, SHA-1 has a lower-than-ideal security level in terms of collision attacks and second preimage attacks. SHA-1 offers the maximum potential strength in terms of second preimage attacks, when the message size (in bits) is up to 160. With bigger message sizes, Eq. (12) is growing up to 1 (Fig. 1).

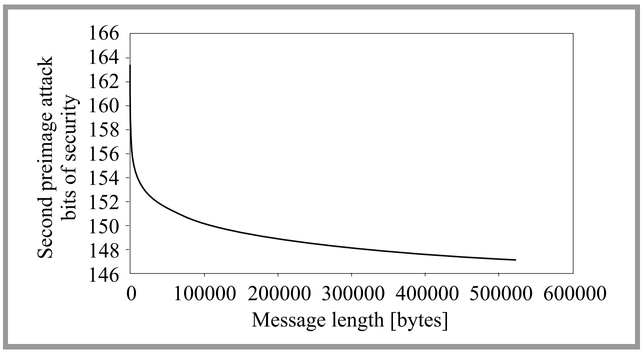


Fig. 1. SHA-160 second preimage attack bits of security.

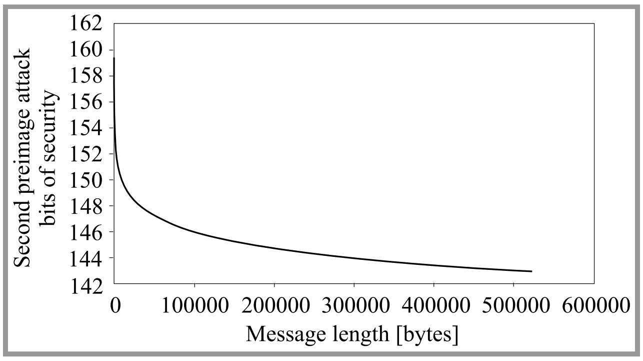


Fig. 2. SHA-256 second preimage attack bits of security.

The SHA-2 family is collision resistant but in every case (except for SHA-384), the security bit parameter of the second preimage attack cannot be ideal when the length of M is greater than the block size B. Dependencies between the number of bits of security and the message length for SHA-256 and SHA-512 are presented in Figs. 2 and 3. For SHA-256, resistance to second preimage attacks is not perfect when the message size is over 256 bits. the maximum message size was measured in the same way as in the case of SHA-1.

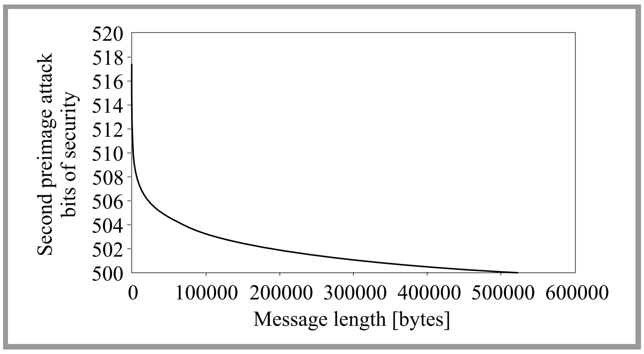


Fig. 3. SHA-512 second preimage attack bits of security.

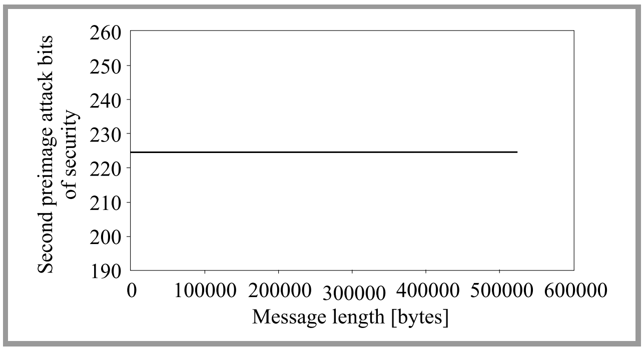


Fig. 4. SHA-224 second preimage attack bits of security.

The case of SHA-512 is similar to that of SHA-256, but this time the maximum secure message size is 512 bits. It is so because Eq. (12) is greater than or equal to 1 only when $\log_2 \frac{\text{len}(M)}{B}$ is greater than 0. $\log_2 \frac{\text{len}(M)}{B}$ is greater than 0 only when the message length is greater than the block size. SHA-224 is different than its predecessors. As can be seen in Fig. 4 the dependence between the number of bits of security and the message size is constant, but according to NIST (Table 1), SHA-224 is not totally resistant against second preimage attacks. This is because the security bits are defined in this case by $\min[224, 256 - \text{L}(M)]$. A vulnerability appears when $\text{L}(M) = 33$ $(56-33 = 223)$. This situation may occur when $\left\lceil \frac{\text{len}(M)}{512} \right\rceil = 2^{33}$. Thus, $\frac{\text{len}(M)}{512}$ should be in the range of $(2^{33}-1, 2^{33})$. $\frac{\text{len}(M)}{512} = 2^{33}-1$ when $\text{len}(M) = 4398046510592$ bits. It means that SHA-224 becomes insecure against second preimage attacks when the size of M is about 512 GB.

The SHA3 function family currently offers perfect security against all three attacks: collision, preimage and second preimage.

5. Experimental Analysis for Selected NIST Hash Functions

For statistical analysis, we have chosen the strongest (the longest) hash function from the SHA1, SHA2 and SHA3 families, i.e. SHA-512, SHA3-512 and SHA1-160, respectively. A data sample consisting of 10,000 messages was considered. All tests were implemented in the JAVA programming language (JDK 1.8) and hashes were generated with the use of the Bouncy Castle library [18].

All 10000 random messages were binary strings. Each input had the same length as the output digest size (160 bit inputs for SHA1 function and 512 bit inputs for SHA-512 and SHA3-512 functions). All input data was generated one by one, separately for SHA1, SHA-512 and SHA3-512, with the use of SecureRandom Java class [19].

By hashing those inputs, we have received the same number pairs: $(input, digest)$ for every hash function.

Three statistical tests were performed: hamming distance test, bits probability test and series test. The details, results, conclusions and comparisons are described in the following subsections.

5.1. Hamming Distance Test

The idea of this test was to measure how small (or even micro) changes in input data influence the output hash. Hash function is passing the test when the statistic |Z| from T-Student test (13) is within the $(0, 1.96)$ interval. The expected value is equal to $\frac{Hashsize}{2}$ (50%). Significance level α was set to 5%. The T-Student formula is:

$$|Z| = \left| \frac{Average\,value - Expected\,value}{Standard\ deviation} \sqrt{Sample\,size} \right|. \quad (13)$$

Firstly, we generated, for each of the pairs $(input, hash)$, another pair $(input', hash')$, where $input'$ was the original input with one random bit changed to the opposite (1 into 0 or 0 into 1), and $hash'$ was a digest calculated from input$'$. We have received two very similar inputs and two hashes. The aim of the experiment was to measure the Hamming distance [20] between these hashes and to repeat this procedure for all inputs generated and for all hashing functions chosen. The hamming distance may be defined as follows. If $S1$ is the first bit string, $S2$ is the second bit string and $\text{len}(S1) = \text{len}(S2)$, the Hamming distance between $S1$ and $S2$ is the number of 1s in the string $S3 = S1 \oplus S2$. It is the number of positions in which $S1$ have different values than $S2$, which can be interpreted as the distance between $S1$ and $S2$.

Results of the experiment for the SHA1 hashing function are presented in Fig. 5 and in Table 2. The horizontal black line is set to 80 because it is the expected value (distance). 4673 out of 10, 000 values were over the black line, but the score is close to 50%. The critical values presented in Table 2 indicate that the average is almost 50%. The $|Z|$ statistic was equal to 1.17, so the T-Student test had been passed. The conclusion is that micro changes in input data

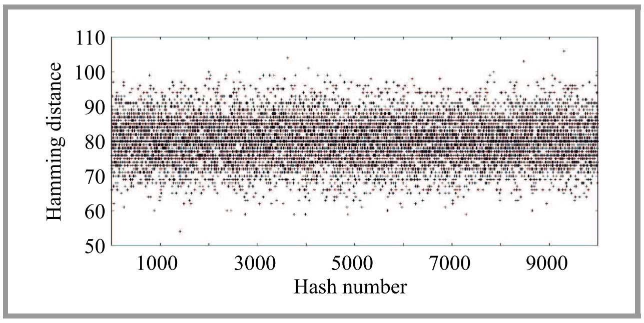


Fig. 5. SHA1 Hamming distance test.

make SHA1 hashes at least 50% different, so the Hamming distance test has been passed.

Results of the experiment for the SHA-512 hashing function are presented in Fig. 6 and Table 2.

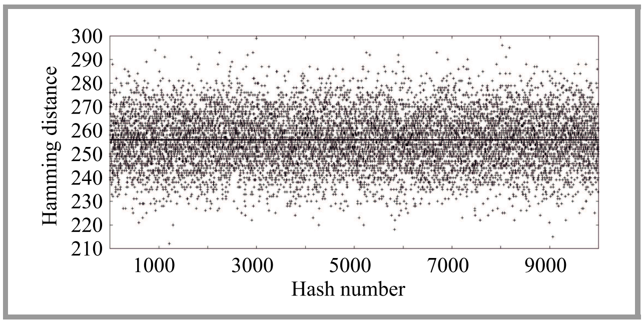


Fig. 6. SHA-512 Hamming distance test.

Table 2
Comparison of Hamming distance critical values

	Hamming distance [% values]			
Function	Max	Min	AVG	SD
SHA-1	66.25	33.75	50.04	±4.02
SHA-512	58.40	41.41	50.00	±2.21
SHA3-512	59.38	41.99	50.00	±2.22

The black line is equal to 256, because it is the expected value (distance). In 4836 out of $10,000$ cases, the difference between hashes was lower than 50%, but in some case it was also close to 50%. The average distance between hash and hash$'$ is equal to 50% and the $|Z|$ value was 0.148. T-Student test was also passed, but the score achieved was much better than in the SHA1 case. SHA-512 has also passed the Hamming distance test.

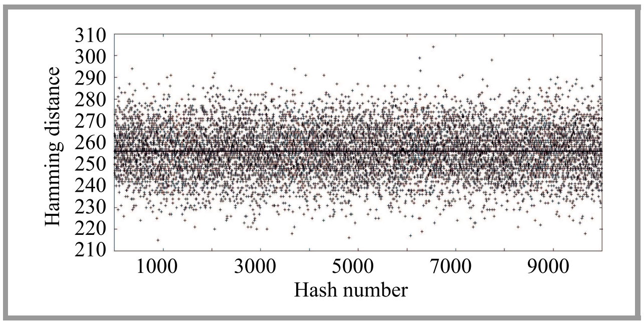


Fig. 7. SHA3-512 Hamming distance test.

The research results for SHA3-512 hashing function are presented in Fig. 7 and in Table 2.

The black line is equal to 256, because it is the expected value. In 4799 out of $10,000$ cases, the difference between hashes was lower than 50%. The critical values are very similar to SHA-512 (average, standard deviation). $|Z|$ statistic was equal to 0.44, thus SHA3-512 has also passed the Hamming distance test.

All three hashing functions passed the Hamming distance test, however, statistically, the SHA-512 is the best, SHA3-512 ranks second and SHA1 ranks third.

5.2. Bits Probability Test

This time, the aim was to check whether bits in the digest may be predicted or not. To measure it, we had to estimate the probability of 1s in every bit position. The ideal situation is when every bit has a 50% probability of being a 1, and a 50% probability of being a 0:

$$P_1(i) = 50\%, \quad i = 1, \ldots, l \ , \tag{14}$$

where i denotes the bit position and l is the hash length. For each hashing function, we used 10,000 generated digests to estimate the probability of ‘1’:

$$P_1(i) = \frac{\sum\limits_{j=1}^{10000} hashes[j][i]}{10000}, \quad i = 1, \ldots, l \ , \tag{15}$$

where *hashes* is a table of generated digests, j denotes j-th hash from *hashes*. We used Eq. (13) to calculate the $|Z|$ statistic. The test is passed when $|Z| < 1.96$ (significance level of $\alpha = 5\%$). The expected value is 50%.

Results of the experiment for the SHA1 hashing function are presented in Fig. (8) and in Table 3.

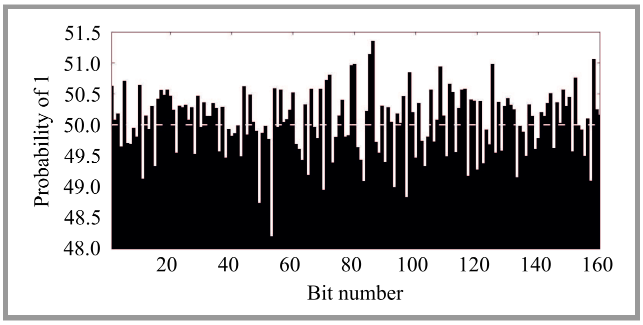


Fig. 8. SHA1 bits prediction.

As one can see, the average value is very close to 50% and the standard deviation is low. Despite the fact that none of the bits have the probability that is equal to 50%, the fluctuations are very small. The $|Z|$ statistic is equal to 1.04, so the conclusion is that none of 160 SHA1 bits can be predicted.

Table 3
Comparison of bits prediction test values

	Probability of 1			
Function	Max	Min	AVG	SD
SHA1	51.35	48.20	50.04	±0.52
SHA-512	51.46	48.49	49.99	±0.53
SHA3-512	51.74	48.76	49.99	±0.51

Results of the test for the SHA-512 hashing function are presented in Table 3 and in Fig. 9.

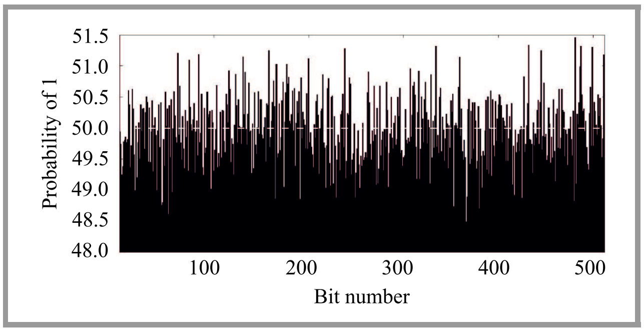


Fig. 9. SHA-512 bits prediction.

Performance of SHA-512 is a similar to that of SHA1 in the context of bits prediction. 510 of the bits have a $P_1(i)$ value that is different than 50%. However, despite small fluctuations, the average value and $|Z|$ equal to 0.863 clearly prove that none of the 512 bits of SHA-512 can be predicted.

Results of the experiment results for the SHA3-512 hashing function are presented in Table 3 and in Fig. 10.

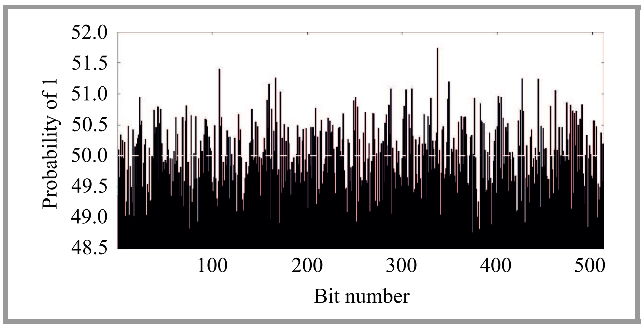


Fig. 10. SHA3-512 bits prediction.

The results for SHA3-512 are similar to those for SHA-512. 506 out of 512 bits fail to satisfy $P_1(i) = 50\%$, but the differences are small. $|Z|$ is equal to 0.317, so the conclusion is that none of 512 bits of the SHA3-512 hashing function may be predicted.

All tested functions pass the bits probability test. This test shows that in every bit position its value is random (ones and zeroes are equally probable). The best score was achieved by SHA3-512, SHA-512 ranked second and SHA1 third.

5.3. Series Test

This test allows to measure whether all hashes were generated in a random manner. Previously, in the bits prediction test, we considered each bit of digest separately, but in the context of all generated hashes. This time the existence of internal dependencies of each hash of each chosen hashing function was measured. To do this, we performed the Wald-Wolfowitz series test [21].

This measure is the subsequence taken from a sequence consisting of the same values only (0 or 1). The number of all series found in one hash will be further denoted by R. n_1 is the number of subsequences consisting only of 1s, and n_0 is the number of subsequences consisting only of 0s. For example, in the 00101101 sequence, the parameters are: $R = 6$, $n_1 = 3$ and $n_0 = 3$. The null hypothesis H_0 claims that the investigated sequence (in this case digest) is random. The alternative hypothesis H_a claims that the investigated sequence was not produced in a random manner. To decide whether H_0 is true or not, a proper test statistic value has to be calculated. Because every generated hash has $n_0 > 20$ and $n_1 > 20$, test statistics tend to have normal distribution $N(0,1)$ (when H_0 is true) and will be denoted by Z. Test statistic Z for each hash was calculated from [21]:

$$Z = \frac{R - \overline{R}}{SD} , \tag{16}$$

where $\overline{R}$ is the expected number of all series, such as:

$$\overline{R} = \frac{2n_0n_1}{n_0 + n_1} + 1 , \tag{17}$$

and SD is the standard deviation:

$$SD = \sqrt{\frac{2n_0n_1(2n_0n_1 - n_0 - n_1)}{(n_0 + n_1)^2(n_0 + n_1 - 1)}} . \tag{18}$$

We have chosen significance level of $\alpha = 5\%$. Thus when $|Z| > Z_{0.975}$ the H_0 is true and the hash investigated was created randomly. Parameter $Z_{0.975}$ is equal to 1.96.

Results of the experiment for the SHA1 hashing function are presented in Table 4 and Fig. 11.

Table 4
Comparison of series test critical values

	Z statistic values			
Function	Max	Min	AVG	SD
SHA1	4.23	$\tilde{0}$	0.79	±0.60
SHA-512	4.04	$\tilde{0}$	0.80	±0.61
SHA3-512	4.39	$\tilde{0}$	0.79	±0.59

The black horizontal line is indicating $Z = 1.96$ ($\alpha = 5\%$). The average value and the standard deviation show that, generally, SHA1 passes the series test, but one may notice

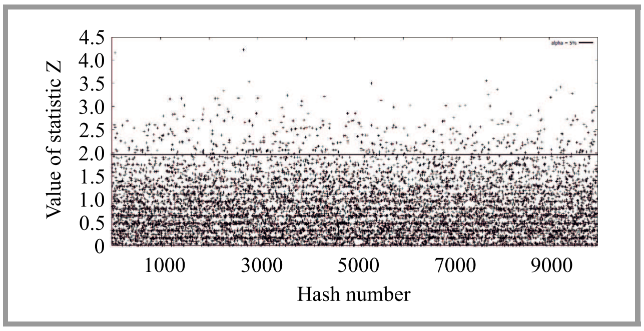


Fig. 11. SHA1 series test.

in the chart that definitely not all digests do so. 480 out of 10,000 samples (4.8%) are considered to have failed.

Results of the test for the SHA-512 hashing function are presented in Table 4 and Fig. 12.

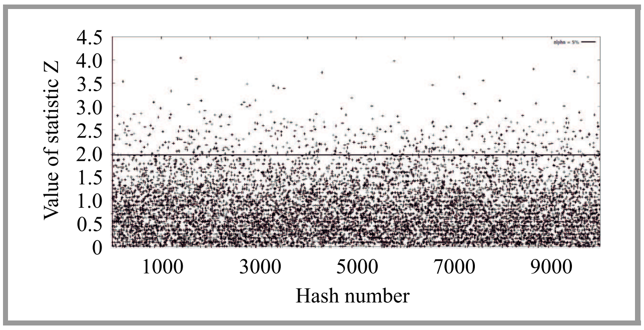


Fig. 12. SHA-512 series test.

In the case of SHA-512, the test was failed in the case of 498 out of 10,000 samples (4.98%). The value is higher than in the case of SHA1 and SHA3-512. The average Z value is the highest, however it is still far from the critical region. We can say that SHA-512 passes the series test with the worst results achieved.

Research results for the SHA3-512 hashing function are shown in Table 4 and Fig. 13.

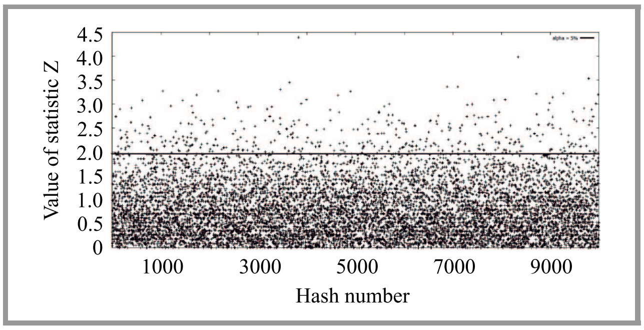


Fig. 13. SHA3-512 series test.

In contrast to SHA-512, SHA3-512 achieved the best test results. Only 417 out of 10,000 samples failed (4.17%), which is the lowest value among all hashing functions tested. The average Z value is also closest to 0. SHA3-512 definitely passes the test.

All three hashing functions have passed our last test. The best score was achieved by SHA3-512 and the worst by SHA-512. In all cases, the test was not passed by less than 5% of samples, so it may be stated that, statistically (with a significance level set to 5%), all hashes were generated randomly.

6. Summary

The aim of this paper was to describe the types, classes and main characteristics of cryptographic hash functions. The formal definition of a hash function was presented and universal hash function classes were described. Then, several methods for the construction of hash functions were disclosed. Standardization procedures for hash functions, as drawn up by NIST, USA, finalize the theoretical part of this paper.

In the research-related sections, we provided an analysis on the influence of the hashed message length on the theoretical security of hash functions, described as the number of bits of security.

The paper describes numerous experiments evaluating the basic features of SHA1, SHA-512 and SHA3-512. The randomness of such functions in terms of input spreading, single bit prediction ability and randomness inside each single bit output, were illustrated. Three tests were performed. The first was based on the Hamming distance measurement, the second examined the frequency of zeros and ones in a large sample, and the third was a series test. Numerous experiments proved that the features of certified hash functions differ, but they all offer very good characteristics in terms of collision resistance, preimage resistance and second preimage resistance attacks.

The analysis provided may be very useful for testing new or proprietary hash functions.

References

[1] I. Gomaa, "Global information assurance certification paper", SANS Institute, May 2011.

[2] R. Sobti and G. Ganesan, "Cryptographic hash functions: A review", *Int. J. of Com. Sci. Issues*, vol. 9, no. 2, pp. 461–479, 2012.

[3] J. L. Carter and Mark N. Wegman, "Universal classes of hash functions", *J. of Comp. and Syst. Sciences*, vol. 18, no. 2, pp. 143–154, 1979 (doi:10.1016/0022-0000(79)90044-8).

[4] R. C. Merkle, "Secrecy, authentication, and public key systems", Ph.D. Thesis, Department of Electrical Engineering, Stanford University, CA, USA, 1979 [Online]. Available: http://www.merkle.com/papers/Thesis1979.pdf

[5] M. Naor and M. Yung, "Universal one-way hash functions and their cryptographic applications", March 1995 [Online]. Available: http://www.wisdom.weizmann.ac.il/~naor/PAPERS/uowhf.pdf

[6] B. Preneel, "Cryptographic hash functions: An overview", in *Proc. of the 6th Int. Comp. Secur. and Virus Conf. ICSVC 1993*, Lueven, Belgium, 1993 [Online]. Available: https://www.esat.kuleuven.be/cosic/publications/article-289.pdf

[7] S. Bakhtiari, R. Safavi-Naini, and J. Pieprzyk, "Cryptographic hash functions: A survey", Tech. Rep. 95-09, vol. 4, Department of Computer Science, University of Wollongong, 1995.

[8] S. Matyas, C. Meyer, and J. Oseas, "Generating strong one-way functions with cryptographic algorithm", *IBM Techn. Disclosure Bull.*, vol. 27, no. 10A, 1985.

[9] S. Wolfram, "Random sequence generation by cellular automata", *Adv. in Appl. Mathem.*, vol. 7, no. 2, pp. 123–169, 1986 (doi: 10.1016/0196-8858(86)90028-X).

[10] J. Daemen, R. Govaerts, and J. Vandewalle, "A framework for the design of one-way hash functions including cryptanalysis of damgård's one-way function based on a cellular automaton", in *Advances in Cryptology ASIACRYPT 91*, H. Imai, R. L. Rivest, and T. Matsumoto, Eds. *LNCS*, vol. 739, pp. 82–96. Berlin Heidelberg: Springer, 1993 (doi: 10.1007/3-540-57332-1_7).

[11] B. Preneel, "The first 30 years of cryptographic hash functions and the nist SHA-3 competition", in *Topics in Cryptology – CT-RSA 2010. The Cryptographers' Track at the RSA Conference 2010, San Francisco, CA, USA, March 1-5, 2010. Proceedings*, J. Pieprzyk, Ed. *LNCS*, vol. 5985, pp. 1–14. Berlin, Heidelberg, 2010 (doi: 10.1007/978-3-642-11925-5_1).

[12] S. Harari, "Non linear non commutative functions for data integrity", in *Advances in Cryptology*, T. Beth, N. Cot, and I. Ingemarsson, Eds. *LNCS*, vol. 209, pp. 25–32. Berlin Heidelberg: Springer, 1985 (doi: 10.1007/3-540-39757-4_4).

[13] M. S. Turan *et al.*, "NISTIR 7764: Status report on the second round of the sha-3 cryptographic hash algorithm competition", Tech. Rep., NIST, 2011 (doi: 10.6028/NIST.IR.7764).

[14] "Announcing request for candidate algorithm nominations for a new cryptographic hash algorithm (SHA-3) family", Tech. Rep., NIST, 2007.

[15] S. J. Chang *et al.*, "NISTIR 7896: Third-round report of the SHA-3 cryptographic hash algorithm competition", Tech. Rep., NIST, 2012 (doi: 10.6028/NIST.IR.7896).

[16] Fips pub 202: M. J. Dworkin, "SHA-3 standard: Permutation-based hash and extendable-output functions", Tech. Rep. no. 202, NIST, 2015 (doi: 10.6028/NIST.FIPS.202).

[17] A. K. Lenstra, "Key lengths. Contribution to the handbook of information security", 2004 [Online]. Available: http://plan9.bell-labs.co/who/akl/key_lengths.pdf

[18] Bouncy Castle library [Online]. Available: https://www.bouncycastle.org (accessed: 08.2018).

[19] Class SecureRandom [Online]. Available: https://docs.oracle.com/javase/7/docs/api/java/security/SecureRandom.html (accessed: 08.2018).

[20] "Hamming Distance and Error Correcting Codes", Hamming Distance [Online]. Available: http://www.oxfordmathcenter.com/drupal7/node/525 (accessed: 08.2018).

[21] Wald-Wolfowitz series test [Online]. Available: https://www.itl.nist.gov/div898/handbook/eda/section3/eda35d.htm (accessed: 08.2018).

Jacek Tchórzewski received his B.Sc. and M.Sc. degrees in Computer Science, with distinctions, from the Cracow University of Technology, Poland, in 2016 and 2017, respectively. Currently, he is a Research and Teaching Assistant at the Cracow University of Technology and a Ph.D. student at AGH Cracow University of Science and Technology.

https://orcid.org/0000-0002-0188-4253
E-mail: jacek.tchorzewski@onet.pl
AGH University of Science and Technology
30 Mickiewicza Av.
30-059 Kraków, Poland

Cracow University of Technology
24 Warszawska St
31-155 Kraków, Poland

Agnieszka Jakóbik (Krok) received her M.Sc. in Stochastic Processes from the Jagiellonian University, Cracow, Poland and the Ph.D. degree in Neural Networks from the Tadeusz Kosciuszko Cracow University of Technology, Poland, in 2003 and 2007, respectively. From 2009 she has been an Assistant Professor at the Faculty of Physics, Mathematics and Computer Science, Tadeusz Kosciuszko Cracow University of Technology. Her main scientific and didactic interests are focused mainly on cloud computing security, artificial neural networks, genetic algorithms, and additionally on cryptography.

https://orcid.org/0000-0003-4568-3944
E-mail: agneskrok@gmail.com
Cracow University of Technology
24 Warszawska St
31-155 Kraków, Poland

Information for Authors

Journal of Telecommunications and Information Technology **(JTIT)** is published quarterly. It comprises original contributions, dealing with a wide range of topics related to telecommunications and information technology. **All papers are subject to peer review**. Topics presented in the JTIT report primary and/or experimental research results, which advance the base of scientific and technological knowledge about telecommunications and information technology.

JTIT is dedicated to publishing research results which advance the level of current research or add to the understanding of problems related to modulation and signal design, wireless communications, optical communications and photonic systems, voice communications devices, image and signal processing, transmission systems, network architecture, coding and communication theory, as well as information technology.

Suitable research-related papers should hold the potential to advance the technological base of telecommunications and information technology. Tutorial and review papers are published only by invitation.

Manuscript. TEX and LATEX are preferable, standard Microsoft Word format (.doc) is acceptable. The authors JTIT LATEX style file is available:
http://www.nit.eu/for-authors

Papers published should contain up to 10 printed pages in LATEX authors style (Word processor one printed page corresponds approximately to 6000 characters).

The manuscript should include an abstract about 150200 words long and the relevant keywords. The abstract should contain statement of the problem, assumptions and methodology, results and conclusion or discussion on the importance of the results. Abstracts must not include mathematical expressions or bibliographic references.

Keywords should not repeat the title of the manuscript. About four keywords or phrases in alphabetical order should be used, separated by commas.

The original files accompanied with pdf file should be submitted by e-mail: redakcja@itl.waw.pl

Figures, tables and photographs. Original figures should be submitted. Drawings in Corel Draw and PostScript formats are preferred. Figure captions should be placed below the figures and can not be included as a part of the figure. Each figure should be submitted as a separated graphic file, in .cdr, .eps, .ps, .png or .tif format. Tables and figures should be numbered consecutively with Arabic numerals.

Each photograph with minimum 300 dpi resolution should be delivered in electronic formats (TIFF, JPG or PNG) as a separated file.

References. All references should be marked in the text by Arabic numerals in square brackets and listed at the end of the paper in order of their appearance in the text, including exclusively publications cited inside. Samples of correct formats for various types of references are presented below:

[1] Y. Namihira, Relationship between nonlinear effective area and mode field diameter for dispersion shifted fibres, *Electron. Lett.*, vol. 30, no. 3, pp. 262264, 1994.

[2] C. Kittel, *Introduction to Solid State Physics*. New York: Wiley, 1986.

[3] S. Demri and E. Orłowska, Informational representability: Abstract models versus concrete models, in *Fuzzy Sets, Logics and Knowledge-Based Reasoning*, D. Dubois and H. Prade, Eds. Dordrecht: Kluwer, 1999, pp. 301314.

Biographies and photographs of authors. A brief professional authors biography of up to 200 words and a photo of each author should be included with the manuscript.

Galley proofs. Authors should return proofs as a list of corrections as soon as possible. In other cases, the article will be proof-read against manuscript by the editor and printed without the author's corrections. Remarks to the errata should be provided within one week after receiving the offprint.

Copyright. Manuscript submitted to JTIT should not be published or simultaneously submitted for publication elsewhere. By submitting a manuscript, the author(s) agree to automatically transfer the copyright for their article to the publisher, if and when the article is accepted for publication. The copyright comprises the exclusive rights to reproduce and distribute the article, including reprints and all translation rights. No part of the present JTIT should not be reproduced in any form nor transmitted or translated into a machine language without prior written consent of the publisher.
For copyright form see: http://www.nit.eu/for-authors

A copy of the JTIT is provided to each author of paper published.

Journal of Telecommunications and Information Technology has entered into an electronic licencing relationship with EBSCO Publishing, the worlds most prolific aggregator of full text journals, magazines and other sources. The text of *Journal of Telecommunications and Information Technology* can be found on EBSCO Publishings databases. For more information on EBSCO Publishing, please visit www.epnet.com.

(*Contents Continued from Front Cover*)

INSTYTUT ŁĄCZNOŚCI
PAŃSTWOWY INSTYTUT BADAWCZY

Editorial Office

National Institute
of Telecommunications
Szachowa st 1
04-894 Warsaw, Poland

tel. +48 22 512 81 83
fax: +48 22 512 84 00
e-mail: redakcja@itl.waw.pl
http://www.nit.eu