Informatics Control Measurement
in Economy and Environment Protection

1/2022

Informatyka Automatyka Pomiary

www.e-IAPGOS.pl

W GOSPODARCE i OCHRONIE ŚRODOWISKA

ISSN 2083-0157

Kwartalnik Naukowo-Techniczny

POLITECHNIKA
LUBELSKA
WYDZIAŁ ELEKTROTECHNIKI
I INFORMATYKI

FUNDACJA ROZWOJU POLITECHNIKI LUBELSKIEJ

netrix SA
RESEARCH & DEVELOPMENT CENTRE

SDwPL

International Interdisciplinary PhD Workshop 2022

15 - 17 March 2022

Lublin, POLAND

International Interdisciplinary PhD Workshop 2022

The International Interdisciplinary PhD Workshop 2022 took place in Lublin between 15-17 March 2022.The workshops are a cyclical undertaking, the individual editions of which were organized in various European countries. The first IIPhDW workshop was held in Szklarska Poręba in 2009. In 2013–2019, the event was organized annually in various countries – Poland, Czech Republic, Slovakia and Germany. The next workshop was scheduled for 2020, but was postponed due to the COVID 19 pandemic. The current edition of IIPhDW 2022 was realized after a two-year break.

The organizers of the conference were:
- Department of Electronics and Information Technology of the Lublin University of Technology,
- Lublin University of Technology Doctoral School,
- Foundation for the Development of Lublin University of Technology,
- Research and Development Center, Netrix S.A.

The companies supporting the event were: Regional Dairy Cooperative „OSM Krasnystaw" and „Pszczółka" Candy Factory Ltd.

Nearly 80 participants took part in the workshop. During the 3 days of the conference there were 2 invited speeches, 8 oral sessions (50 speeches) and 1 poster session (12 posters). In addition, 1 session conducted remotely was organized, during which Russian-speaking doctoral students presented their research achievements.

Participants of the conference had the opportunity to take part in a city game, during which they visited the most important attractions of Lublin. The guests also had the opportunity to hear the history of the Perła Brewery Underground and learn about the ingredients and the course of the beer production process.

During this year's edition of IIPhDW 2022, the main goal of the conference was realized, which was to gather doctoral students and young doctors in order to share knowledge and discuss problems related to their research and scientific interests. The workshops enabled the participants to exchange experiences and integrate the scientific community. This event gave an opportunity to intensify cooperation with national and foreign research centers.

The next edition of the IIPhDW workshop will take place on May 3-5, 2023 in Wismar, Germany. Detailed information about the event can be found on the website https://www.hs-wismar.de/iiphdw-2023/.

foto: Tomasz Maślona

1/2022

styczeń – marzec

Wydanie pod redakcją naukową
prof. dr hab. inż. Waldemara Wójcika

INFORMATYKA AUTOMATYKA POMIARY

W GOSPODARCE i OCHRONIE ŚRODOWISKA

Informatics Control Measurement in Economy and Environment Protection

p-ISSN 2083-0157, e-ISSN 2391-6761, www.e-iapgos.pl

PRINTING HOUSE – DRUKARNIA

Soft Vision Mariusz Rajski
https://www.printone.pl

nakład: 100 egzemplarzy

OTHER INFORMATION – INNE INFORMACJE

Czasopismo jest indeksowane w bazach:

DOAJ	doaj.org
BazTech	baztech.icm.edu.pl
IC Journals Master List	www.journals.indexcopernicus.com
Google Scholar	scholar.google.pl
POL-index	pbn.nauka.gov.pl
Sherpa RoMEO	www.sherpa.ac.uk
OAJI	oaji.net
SCOPUS	www.scopus.com

Czasopismo *Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska* zostało objęte finansowaniem przez Ministerstwo Nauki i Szkolnictwa Wyższego w ramach programu *Wsparcie dla czasopism naukowych* w latach 2019-2020.

Czasopismo znajduje się w wykazie czasopism naukowych opublikowanym w Komunikacie Ministra Edukacji i Nauki z dnia 1 grudnia 2021 r., Unikatowy Identyfikator Czasopisma: 200167 – z przypisaną liczbą punktów przyznawanych za publikację artykułu równą 20.

Zasady publikowania artykułów, przygotowania tekstów, zasady etyczne, procedura recenzowania, wykazy recenzentów oraz pełne teksty artykułów dostępne są na stronie internetowej czasopisma:

www.e-iapgos.pl

W celu zwiększenia oddziaływania czasopisma w środowisku naukowym redakcja zaleca:

- w artykułach publikowanych w IAPGOS cytować artykuły z renomowanych czasopism międzynarodowych (szczególnie indeksowanych w bazach Web of Science oraz Scopus) używając oficjalnych skrótów nazw czasopism,
- w artykułach publikowanych w innych czasopismach (zwłaszcza indeksowanych w bazach Web of Science oraz Scopus) cytować prace publikowane w IAPGOS – zwłaszcza posługując się numerami DOI, np.:
Kluszczyński K. *Modelowanie – umiejętność czy sztuka?* Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska – IAPGOS, 1/2016, 4–15, https://doi.org/10.5604/20830157.1193833.

CONTENTS – SPIS TREŚCI

http://doi.org/10.35784/iapgos.2877 *received: 02.02.2022 | revised: 16.03.2022 | accepted: 28.03.2022 | available online: 31.03.2022*

METHODS FOR ENSURING DATA SECURITY IN MOBILE STANDARDS

Serhii Moroz, Anatolii Tkachuk, Mykola Khvyshchun, Stanislav Prystupa, Mykola Yevsiuk
Lutsk National Technical University, Faculty of Computer and Information Technologies, Department of Electronics and Telecommunications, Lutsk, Ukraine

Abstract. *The analysis of mobile communication standards is carried out, the functional structure and interfaces of interaction between the structural elements of the cellular network are considered. To understand the principle of communication according to the GSM standard, a block diagram of a mobile switching center (MSC), base station equipment (BSS), control and service center (MCC), mobile stations (MS) is presented. The main algorithms for ensuring the confidentiality and security of mobile subscribers' data, in different types of standards, as well as the vulnerabilities of information flows are considered. In particular, the following dangerous types of attacks have been identified, to which mobile network subscribers are sensitive: sniffing; leakage of personal data; leakage of geolocation data; spoofing; remote capture of SIM-card, execution of arbitrary code (RCE); denial of service (DoS). It is established that the necessary function of the mobile network is the identification of subscribers, which is performed by IMSI, which is recorded in the SIM card of the subscriber and the HLR of the operator. To protect against spoofing, the network authenticates the subscriber before starting its service. In the case of subscriber identification, the subscriber and the network operator are protected from the effects of fraudulent access. In addition, the user must be protected from eavesdropping. This is achieved by encrypting the data transmitted over the radio interface. Thus, user authentication in UMTS, as well as in the GSM network, is carried out using encryption with a common key using the "hack-response" protocol (the authenticating party sends a random number to the authenticated party, which encrypts it according to a certain algorithm using a common key and returns the result back).*

Keywords: mobile communication, radio communication equipment, encryption, authentication

METODY ZAPEWNIENIA BEZPIECZEŃSTWA DANYCH W STANDARDACH MOBILNYCH

Streszczenie. *Przeprowadzana jest analiza standardów komunikacji mobilnej, rozważana jest struktura funkcjonalna i interfejsy interakcji między elementami strukturalnymi sieci komórkowej. Aby zrozumieć zasadę komunikacji w standardzie GSM, przedstawiono schemat blokowy centrali ruchomej (MSC), wyposażenia stacji bazowej (BSS), centrum sterowania i obsługi (MCC), stacji ruchomych (MS). Rozważane są główne algorytmy zapewniające poufność i bezpieczeństwo danych abonentów telefonii komórkowej w różnych typach standardów, a także podatności na przepływ informacji. W szczególności zidentyfikowano następujące niebezpieczne rodzaje ataków, na które podatni są abonenci sieci komórkowych: sniffing; wyciek danych osobowych; wyciek danych geolokalizacyjnych; podszywanie się; zdalne przechwytywanie karty SIM, wykonanie dowolnego kodu (RCE); odmowa usługi (DoS). Ustalono, że niezbędną funkcją sieci komórkowej jest identyfikacja abonentów, która jest realizowana przez IMSI, która jest zapisywana na karcie SIM abonenta i HLR operatora. Aby zabezpieczyć się przed podszywaniem się, sieć uwierzytelnia subskrybenta przed uruchomieniem usługi. W przypadku identyfikacji abonenta, abonent i operator sieci są chronieni przed skutkami nieuprawnionego dostępu. Ponadto użytkownik musi być chroniony przed podsłuchem. Osiąga się to poprzez szyfrowanie danych przesyłanych przez interfejs radiowy. Tak więc uwierzytelnianie użytkownika w UMTS, jak również w sieci GSM, odbywa się z wykorzystaniem szyfrowania wspólnym kluczem z wykorzystaniem protokołu „hack-response" (strona uwierzytelniająca wysyła do strony uwierzytelnianej losową liczbę, która ją szyfruje zgodnie z określnym algorytmem używając wspólnego klucza i zwraca wynik).*

Słowa kluczowe: komunikacja mobilna, sprzęt radiokomunikacyjny, szyfrowanie, uwierzytelnianie

Introduction

Mobile communications are currently viewed as a necessity, and mobile communications technologies are the most in-demand and rapidly growing. Mobile communication systems are developing at a very fast pace. Considering the evolution of mobile communication systems, we come to the concept of "generations" [2]. The first to be fully digital were second-generation (2G) mobile communication systems. They have brought significant benefits in terms of offering improved services, capacity, and quality to subscribers. GSM (Global System for Mobile communication) refers to 2G technology and is the fundamental technology on which the technologies of existing mobile communication systems have grown.

The increase in the number of users and the expansion of the volume of services has led to the creation of the concept of third-generation (3G) systems that will allow communication, exchange of information, and the provision of various entertainment services focused on a wireless terminal. The development of such services has already begun for 2G systems, but to support these services, the system must have a high capacity and bandwidth of radio channels, as well as interoperability between systems. An example of a 3G system is the Universal Mobile Telecommunications System (UMTS) [17].

4G is the fourth generation of cellular communications. Unlike previous generations, namely the third, 4G does not support traditional circuit-switched telephony services. 4G uses IP telephony, that is, packet switching. For telephony services, calls and messages can be switched over existing 2G and 3G networks. The 4G standard allows information to be transmitted at a speed five to seven times higher than that of 3G [18]. The main difference between 4G networks and earlier ones is that 4G technology is based entirely on packet data transfer protocols, while 3G combines both packet switching and circuit switching. For transmission in 4G, VoLTE technologies are provided [15]. The fifth generation of 5G mobile communications, operates on the basis of telecommunications standards, more modern than the existing 4G / IMT-Advanced standards. 5G will provide higher bandwidth than 4G technologies, which will enable greater availability of mobile broadband and the Internet of Things [3].

Let's consider how the user is authenticated and the encryption key is generated in GSM networks (Fig. 1).

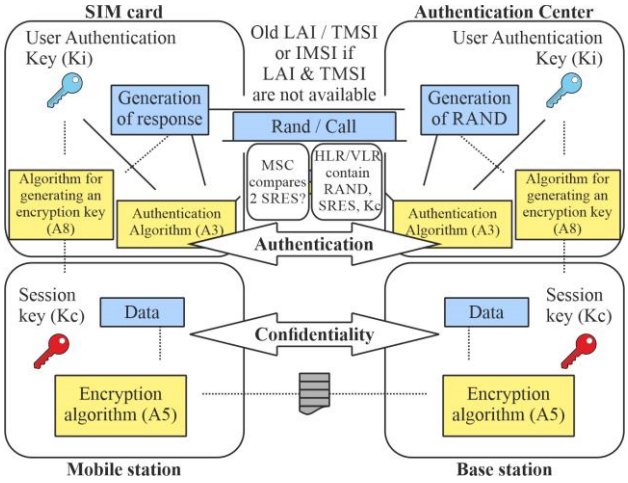


Fig. 1. User authentication and encryption key generation in GSM networks

Figure 1 shows schematically the following steps:

1. The phone connects to the network.
2. To confirm its identity, the phone sends a special identification code TMSI (Temporary Mobile Subscriber Identity).
3. The Authentication Center (CA) generates a 128-bit random number RAND and sends it to the Mobile Station (MS).
4. MS encrypts the received RAND using its private key Ki and the A3 authentication algorithm.

5. MC takes the first 32 bits from the sequence obtained in the previous step (SRES (signed response)) and sends them back to the DA.
6. The CA performs the same operation and receives a 32-bit XRES (expected response) sequence.
7. CA compares SRES and XRES. If both values are equal, the phone is considered authenticated.
8. MS and CA calculate the session encryption key using the private key Ki and the key generation algorithm A8 Kc = A8ki (RAND).

1. Literature review

Consider the process of encryption and authentication in the GSM network. In addition to IMSI, the SIM card stores a randomly generated sequence "Ki", which it returns only in a hashed form. Ki is stored in the HLR of the operator and is not transmitted in the open [10]. The authentication process is based on the principle of "four-way handshake":

1. The subscriber performs Location Update Request, then provides IMSI.
2. The network sends a pseudo-random RAND value.
3. The phone's SIM card hashes Ki and RAND according to the A3 algorithm. A3 (RAND, Ki) = SRAND.
4. The network hashes Ki and RAND by algorithm A3.
5. If the value of SRAND on the part of the subscriber coincided with the calculated on the network side, then the subscriber has been authenticated (Fig. 2).

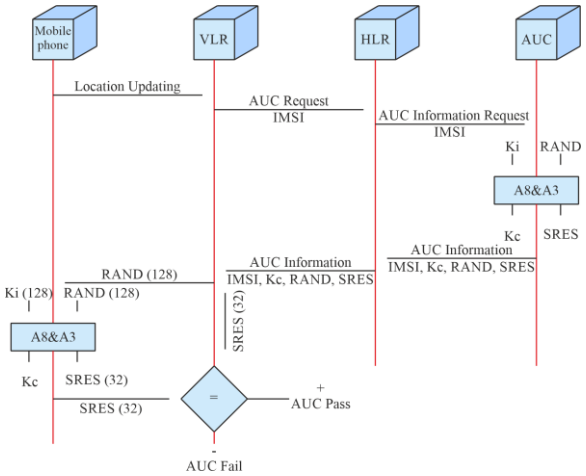


Fig. 2. Authentication procedure scheme

In order to achieve the required level of security, algorithm A3 must be a unidirectional function [7]. This means that the calculation of SRES for known Ki and RAND should be simple, and the reverse – the calculation of Ki for known RAND and SRES – should be as complex as possible. This determines the level of security of the system. The value calculated by the A3 algorithm must be 32 bits long [4]. Cryptographic methods make it possible to achieve a high level of security by simple means. GSM uses common methods to protect all data: user information, signal transmission, message transmission, and system signal transmission [1]. When the authenticity of the subscriber has been verified, thus protecting both the subscriber and the network operator from the influence of fraudulent access, the user must be protected from eavesdropping, this is achieved by encrypting data transmitted over the radio using the key Kc and secret algorithm A5 (Fig. 3) [19].

Kc is generated during authentication using Ki, RAND, and the secret algorithm A8, which is stored on the SIM card. Like the A3 algorithm, the A8 is not unique and can also be generated by an operator [21]. The Kc keys for each user are calculated by the AuC of the home network and transmitted to the VLR as part of a set of triplets, where each triplet and, accordingly, the key Kc, is assigned a key number – CKSN. In some implementations, algorithms A3 and A8 are combined into a single algorithm A38, which uses RAND and Ki to generate Kc and SRES [9]. Unlike A3 and A8, which may be different for each individual operator, A5 is selected from a list of 7 possible options. The encryption is preceded by an information exchange phase, during which it is determined which version of the A5 will be used. If the network and the mobile station do not have common versions of A5, the connection must continue in open mode or the connection will be lost. Algorithm A5 uses a 64-bit key Kc and a 22-bit TDMA frame number to compute the two 114-bit encryption words BLOCK1 and BLOCK2, which are used in transmission and reception, respectively [8]. Encryption words – EXORed with 114 bits of data in each frame. Because the encrypted data is computed using the TDMA frame number, the words change from frame to frame and are not repeated during the hyper frame (~ 3.5 hours). Before starting the encryption, the mobile station (MS) sends to the VLR the number of the CKSN encryption key, which is stored in its memory since the last authentication procedure. CKSN does not contain secret data, but serves to allow the MS to tell the network which key Kc it "remembers" [5]. The VLR then sends a command to the MS to enable encryption and transmits to the base station (BTS) the key Kc from the triplet that corresponds to the CKSN number received from the MS. Thus, the choice of the encryption key is agreed upon between the MS and the VLR without transmitting the key itself over the radio interface [20].

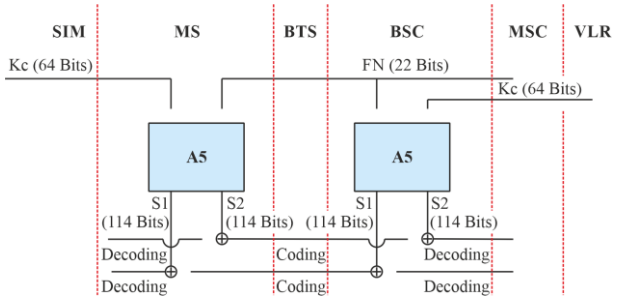


Fig. 3. Encryption and decryption procedure in the GSM standard

The GSM network is characterized by the following restrictions in ensuring information security:

- one-way authentication (provides for user authentication only). This is the reason for the lack of protection, for example, against attack by a false base station BTS;
- there is no mechanism for ensuring message authentication, protecting data integrity;
- fixed security mechanisms without the possibility of flexible shifts in order to enhance protection;
- The data is encrypted only between the mobile station (MS) and the base station (BTS). In the section between the BTS and the base station controller BSC, this data is transmitted in cleartext.

In UMTS, as well as in GSM, the temporary TMSI is used to protect the privacy of the location and to ensure the confidentiality of the IMSI in the area of the CC core network. Similar temporary identifiers are used in the area of the KP core network. To distinguish them from TMSI, the term Packet Temporary Mobile Subscriber Identity (P-TMSI) is used here.

When the connection is established, the temporary TMSI / P-TMSI is periodically changed. Since the temporary identifier is local in nature, then, as in GSM, it must be accompanied by the identifier of the location area. Therefore, the Local Area Identity (LAI) location area identifier is added to the TMSI, and the Routing Area Identity (RAI) routing area identifier is added to the P-TMSI.

Each location area LA (Local Area) is uniquely identified by the identifier LAI, consisting of:

LAI = MCC + MNC + LA code, where MCC is the mobile country code (three digits),

MNC – mobile network code (2–3 digits),

LA is a number that identifies this area.

The KP core network area has its own location registration procedure, which is based on the concept of the RA (Routing Area) routing area. The RA is defined similarly to the LA (as the area within which the UE can move without the routing area update procedure). At the same time, the RA region is a "subset".

2. Researches methodology

Consider mutual authentication of users and networks. UMTS provides mutual user and network authentication. The 3G network structure can be integrated into a system with many different terrestrial radio access technologies, which in turn requires network authentication [11]. The authentication mechanism is based on a master key (master key) K, which is shared between the USIM and the home network database. This 128-bit key never becomes visible between two points on the network. Along with mutual authentication, encryption and integrity keys are generated [12]. At the same time, the principle of cryptography (Kirchhoff's principle) is fulfilled in limiting the duration of using a permanent key to a minimum and using temporary keys [13]. The authentication procedure starts after the user is identified in the serving network by the procedure for ensuring the privacy of the location of the mobile station [16]. User identification is carried out as a result of signaling messages containing TMSI or P-TMSI (in exceptional cases IMSI) to the VLR or SGSN [6, 14]. In Fig. 4 shows a diagram of mutual authentication.

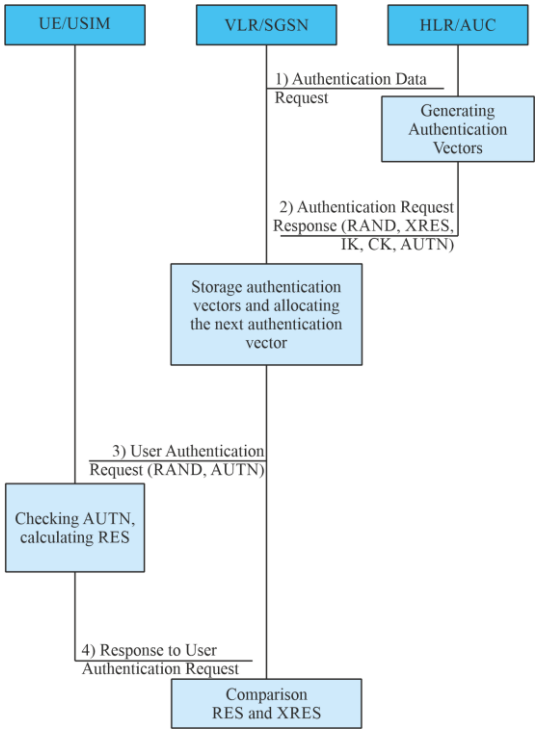


Fig. 4. Mutual authentication framework

In message 1, the VLR or SGSN sends a request for user authentication data to the AUC. This message includes the user ID IMSI. Based on the key K, the AUC authentication center generates and transmits to the HLR authentication vectors for the user with the IMSI.

Each authentication vector contains:

1. Random number RAND (hail);
2. Expected response to the hail, XRES;
3. Encryption key, CK (Cipher Key);
4. Integrity key, IK (Integrity Key);
5. Parameter (label) of network authentication AUTN (An Authentication Token).

In message 2 (response to a request for authentication data), the HLR sends the authentication vector back to the VLR or SGSN. Messaging 1 and 2 are exchanged using the SS7 MAP mobile application protocol. Message 3 (User Authentication Request) contains two parameters from the Authentication Vector, RAND and AUTN.

This message 3 is transmitted to the subscriber identification module USIM, which is located in a protected environment from unauthorized access (to a UMTS smart card – UICC, UMTS Integrated Circuit Card). The USIM user uses the AUTN parameter value to verify the authenticity of the connected network. The USIM user uses RAND to compute the RES response to the user's authentication request.

Message 4 (response to user authentication request) contains the parameter RES. This message is sent back from the UE to the VLR / SGSN where RES is compared to the expected XRES value. If RES and XRES match, the network verifies the identity of the user.

3. Results

In Fig. 5 shows a scheme for generating an authentication vector in HLR / AUC for generation 3G networks. The Authentication Center contains the user's master keys and, based on the international mobile station identifier IMSI, generates authentication vectors for the user. Note that in a GSM network, such a key generates an authentication vector from two parameters – a revocation and an encryption key. Here the number of authentication vector parameters is greater.

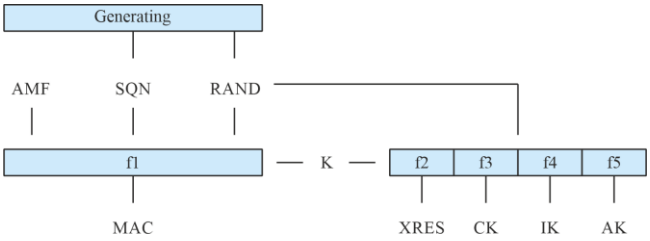


Fig. 5. Generation of the authentication vector: SQN (Sequence Number) – the sequence number of the authentication is used to protect against the "retry" threat; AK (Anonymity Key) – the key used to encrypt SQN; AMF (Authentication Management Field) – administrative control field of authentication, can be used to specify multiple encryptions; MAC (Message Authentication Code) – message authentication code, MAC = f (K, AMF, SQN, RAND)

Five one-way functions are used to compute the authentication vector – f1, f2, f3, f4, f5. The f1 and f5 functions are used to authenticate the network, and the f2 function to authenticate the user. The AV authentication vector, in addition to the AUTN authentication parameter, the XRES response (function f2), also includes the encryption key CK (function f3) and the integrity key IK (function f4): {\ displaystyle AV = RAND \ | XRES \ | CK \ | IK \ | AUTN}. Network Authentication Option {\ displaystyle AUTN = SQN \ oplus AK \ | AMF \ | MAC}.

The protection of a subscriber in the 4G generation of cellular communication is that in the process of servicing it is hidden with temporary identifiers. To close data in LTE networks, streaming encryption is used by superimposing a pseudo-random sequence on open information using the XOR operator. These networks use the principle of tunneling connections to ensure security within the network. S1 and X2 packets can be encrypted using IPsec ESP, and signaling messages on these interfaces are encrypted. At the moment of connection or activation of the user equipment (UE) in the network, the network starts the authentication procedure and AKA (Authentication and Key Agreement). The purpose of this procedure is to mutually authenticate the subscriber and the network and generate an intermediate KASME key. The algorithm for authentication and key generation is shown in Fig. 6.

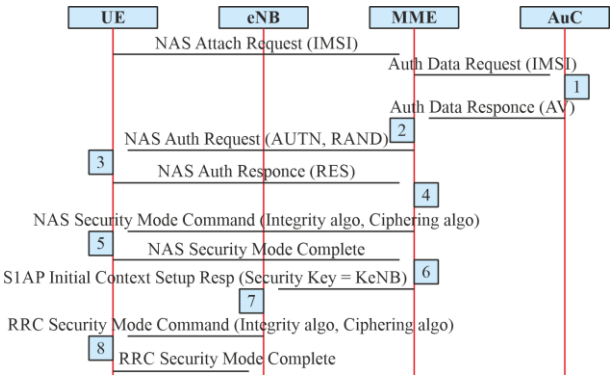


Fig. 6. Authentication and key generation diagram

Step 1. Request for a network connection from a mobile station (UE). The MME requests authentication data related to a specific IMSI by sending an Authentication Data Request. The AuC / HSS selects the PSK associated with a specific IMSI and computes the PSK authentication data. AuC / HSS sends back AV with Authentication Data Response.

Step 2. MME gets IK, CK, XRES, RAND, and AUTH from AV. MME sends AUTH and RAND with Authentication Request to UE.

Step 3. The UE authenticates the NW by checking the received AUTH. Then it calculates IK, CK, RES, XMAC from its security key, AMF, (OP), AUTH and RAND. She sends RES with an Authentication response.

Step 4. After receiving the RES, the MME compares it with XRES and if they match, then the authentication was successful, otherwise the MME sends an Authentication failure to the UE. MME resets DL NAS counter. Calculates KASME, KeNB, Knas-int, Knas-enc. Sends the NAS security mode command (integrity algorithm, encryption algorithm, NAS keyset ID, UE security function) with the integrity protected but not encrypted using Knas-inc.

Step 5. After the NAS receives the security mode command, the UE calculates KASME, KeNB, Knas-int, Knas-enc.

Step 6. After the NAS receives the security mode command from the UE, the MME sends the KeNB to the eNB with the S1AP initial setting of the initial context (security key).

Step 7. After receiving KeNB, eNB calculates Krrc-int, Krrc-enc, Kup-enc. It then sends the RRC security key command with the AS integrity of the algorithm and the AS encryption algorithm.

Step 8. After receiving the RRC command of the security key, the UE calculates Krrc-int, Krrc-enc, Kup-enc. The UE sends the RRC the executed encryption key to the eNB.

After all the described actions, all NAS and AS messages will be reliably protected and encrypted, in contrast to user data, which will only be encrypted. The LTE security architecture defines a security mechanism for both the NAS and AS layers (figure 7).

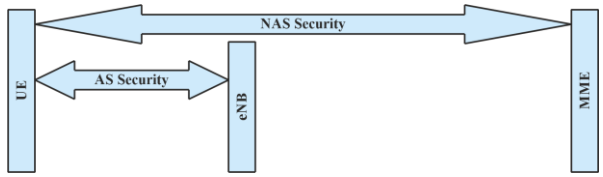


Fig. 7. Security layers

NAS (Non-Access Stratum) security. Made for NAS messages and belongs to the UE and MME. In this case, integrity is required when transmitting NAS messages between the UE and the MME, protected and encrypted with an optional NAS security header.

AS security (Access Stratum). Implemented for RRC and user data plane belonging to the UE and eNB. The PDCP layer on the UE and eNB sides is responsible for encryption and integrity protection. RRC messages are integrity protected and encrypted, however, U-Plane data is only encrypted.

In 5th generation networks, the trust in the elements of the network decreases as the elements move away from the core of the network. This concept influences the solutions implemented in the 5G security architecture. Thus, we can talk about a 5G network trust model that determines the behavior of network security mechanisms. From the user's side, the trust domain is formed by the UICC and USIM. On the network side, the trust domain has a more complex structure. The radio access network is subdivided into two components – DU (Distributed Units) and CU (Central Units) (Fig. 8).

Together they form the gNB – the radio interface of the 5G network base station. DUs do not have direct access to user data, since they can be deployed in segments of unprotected infrastructure. CUs, on the other hand, should be deployed in protected network segments, since they are responsible for terminating the traffic of AS security mechanisms. At the core of the network is the AMF, which terminates the traffic of the NAS's security mechanisms. In 5th generation networks, the authentication procedure has two components: primary and secondary authentication. Primary authentication is required for all user devices connecting to the network. Secondary authentication can be performed upon request from external networks if the subscriber connects to those. After the successful completion of the primary authentication and the generation of a shared key K between the user and the network, KSEAF is extracted from the key K – special anchor (root) key of the serving network. Subsequently, keys are generated from this key to ensure the confidentiality and data integrity of the RRC and NAS signaling traffic.

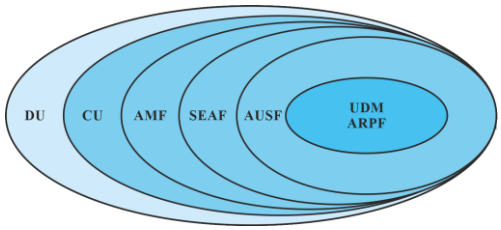


Fig. 8. Components of a 5G radio access network

Initial authentication procedure. On 5G networks, EAP-AKA and 5G AKA are the standard primary authentication mechanisms. Let's split the primary authentication mechanism into two phases: the first is responsible for initiating authentication and choosing an authentication method, the second is for mutual authentication between the user and the network (Fig. 9).

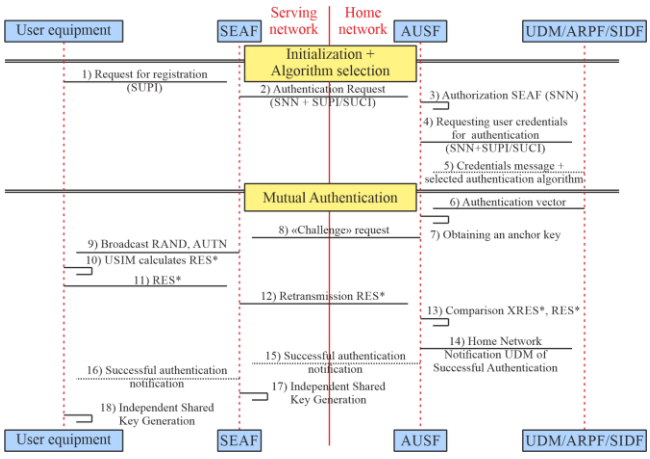


Fig. 9. Initial authentication procedure 5G

Initiation. The user submits a SEAF registration request that contains the user's hidden subscription identifier SUCI.

SEAF sends to AUSF an authentication request message (Nausf_UEAuthentication_Authenticate Request) containing SNN (Serving Network Name) and SUPI or SUCI.

The AUSF checks if the SEAF requestor is allowed to use the given SNN. If the serving network is not authorized to use this SNN, then the AUSF responds with an authorization error "Serving network not authorized" (Nausf_UEAuthentication_Authenticate Response).

Authentication credentials are requested from AUSF in UDM, ARPF, or SIDF over SUPI or SUCI and SNN. Based on SUPI or SUCI and user information, UDM / ARPF chooses an authentication method that will be used further and issues user credentials.

Mutual authentication. When using any authentication method, the UDM / ARPF network functions MUST generate an Authentication Vector (AV). EAP-AKA: UDM / ARPF first generates an authentication vector with AMF splitting bit = 1, then generates CK 'and IK' from CK, IK, and SNN and composes a new AV authentication vector (RAND, AUTN, XRES *, CK ', IK '), which is sent to AUSF with the instruction to use it only for EAP-AKA. 5G AKA: UDM / ARPF obtains the KAUSF key from CK, IK, and SNN, and then generates 5G HE AVangl. 5G Home Environment Authentication Vector). The 5G HE AV authentication vector (RAND, AUTN, XRES, KAUSF) is sent to AUSF with the instruction to use it for 5G AKA only.

The AUSF then obtains the anchor key KSEAF from the KAUSF key and sends a request to SEAF "Challenge" in a "Nausf_UEAuthentication_Authenticate Response" message containing also RAND, AUTN, and RES *. Next, RAND and AUTN are transmitted to the user equipment using a secure NAS signaling message. The user's USIM calculates RES * from the received RAND and AUTN and sends it to SEAF. SEAF will relay this value to AUSF for verification.

AUSF compares the XRES * stored in it and the RES * received from the user. In case of a match, AUSF and UDM in the operator's home network are notified of successful authentication, and the user and SEAF independently generate a KAMF key from KSEAF and SUPI for further communication.

Secondary authentication. The 5G standard supports optional secondary authentication based on EAP-AKA between the user equipment and external data network. In this case, the SMF acts as an EAP authenticator and relies on the external network AAA server to authenticate and authorize the user (Figure 10).

- Mandatory primary user authentication occurs on the home network and develops a common NAS security context with AMF.
- The user sends a session request to the AMF.
- AMF sends a request to establish a session in the SMF with the user's SUPI.
- SMF verifies user credentials in UDM using the provided SUPI.
- SMF sends a response to a request from AMF.
- SMF starts the EAP authentication procedure to obtain permission to establish a session from the AAA server of the external network. To do this, the SMF and the user exchange messages to initiate the procedure.
- The user and the AAA server of the external network then exchange messages to authenticate and authorize the user. The user sends messages to the SMF, which in turn exchanges messages with the external network via the UPF.

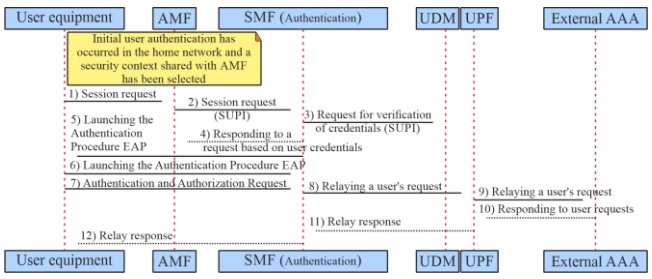


Fig. 10. Secondary authentication procedure 5G

The presented researches allow us to carry out the comparative analysis of different generations of mobile communication from the point of view of safety. In particular, studies of data encryption and authentication procedures for various mobile networks have been conducted. The conducted research will make it possible to identify the main vulnerabilities in the process of establishing a mobile connection and propose measures to eliminate the identified vulnerabilities.

4. Conclusions

As a result of the study, the following types of threats to mobile networks were identified and classified.

A. Unauthorized access to sensitive data (breach of confidentiality).

A1. Eavesdropping: An intruder intercepts messages without being detected.

A2. Masquerading: an intruder tricks an authorized user by posing as a legitimate system in order to obtain confidential information from the user, or an intruder tricks a legitimate system by posing as an authorized user in order to obtain system services or confidential information.

A3. Traffic analysis: the intruder monitors such data as time, the volume of messages, source, and recipient of messages in order to determine the location of the user or find out important information.

A4. Browsing: An intruder browses the stored data in search of sensitive information.

A5. Leakage: An intruder obtains sensitive information through legitimate access to data.

A6. Interference: An intruder observes the system's response by sending requests to it. For example, an intruder can initiate a connection and then monitor the radio interface for data such as the source and recipient of messages, the volume of messages.

B. Unauthorized manipulation of important data (violation of integrity). Manipulation of messages: A message can be deliberately altered, inserted, duplicated, or destroyed by an intruder. Network disruption (resulting in a denial of service or reduced availability).

B1. Intervention: An intruder can prevent an authorized user from serving by deliberately interfering with traffic, signaling, or control data.

B2. Resource exhaustion: An intruder can prevent an authorized user from being served by creating an overload.

B3. Misuse of privileges: A user or service network uses its priorities to obtain unauthorized service or information.

B4. Abuse of services: An offender may abuse certain special services or facilities in order to gain an advantage or cause disruption to the network.

C. Repudiation: A user or network is giving up the actions that took place. Unauthorized access to services.

C1. An intruder can gain access to the service masquerading as a user or network object.

C2. A user or a network object can gain unauthorized access to the service without using their access rights.

Most of the threats to mobile generations have been addressed. However, attackers are looking for new and new vulnerabilities in communication standards. That is why the analysis of the security architecture of various communication technologies is important for counteracting illegal access to mobile network data. The following studies, it is planned to highlight ways to protect the information flows of mobile networks with virtual technical functions (VTF) and their evolution to generations 6G and above.

References

[1] Al-Tawil K., Akrami A.: A new authentication protocol for roaming users in GSM networks. Proceedings IEEE International Symposium on Computers and Communications (Cat. No. PR00250) 1999, 93–99, [http://doi.org/10.1109/ISCC.1999.780775].

[2] Bakhovskyy P. et al.: Stages of the Virtual Technical Functions Concept Networks Development. D. Cagá˘nová et al. (eds.), Advances in Industrial Internet of Things, Engineering and Management. EAI, Springer Innovations in Communication and Computing, 2021, 119–135 [http://doi.org/10.1007/978-3-030-69705-1_7].

[3] Cai Y. et al.: Modulation and Multiple Access for 5G Networks. IEEE Communications Surveys & Tutorials 20(1), 2018, 629–646, [http://doi.org/10.1109/COMST.2017.2766698].

[4] Chen W. et al.: NFC Mobile Transactions and Authentication Based on GSM Network. Second International Workshop on Near Field Communication, 2010, 83–89, [http://doi.org/10.1109/NFC.2010.15].

[5] Deng L. et al.: Mobile network intrusion detection for IoT system based on transfer learning algorithm. Cluster Comput 22, 2019, 9889–9904 [http://doi.org/10.1007/s10586-018-1847-2].

[6] Gupta A., Jha R. K.: A Survey of 5G Network: Architecture and Emerging Technologies. IEEE Access 3, 2015, 1206–1232, [http://doi.org/10.1109/ACCESS.2015.2461602].

[7] Hongfeng Z. et al.: A Novel and Provable Authenticated Key Agreement Protocol with Privacy Protection Based on Chaotic Maps towards Mobile Network. International Journal of Network Security 18(1), 2016, 116–123.

[8] Melnyk V. et al.: Design and implementation of interdomain communication mechanism for high performance data processing. Eastern-European Journal of Enterprise Technologies 1(9), 2016, 10–15, [http://doi.org/10.15587/1729-4061.2016.60629].

[9] Melnyk V. et al.: Implementation of the simplified communication mechanism in the cloud of high performance computations. Eastern-European Journal of Enterprise Technologies 2(86), 2017, 24–32, [http://doi.org/10.15587/1729-4061.2017.98896].

[10] Pekh P. et al.: Generators of Some Kinds Random Erlang Numbers and Estimation of Their Complexity. 10th International Conference on Advanced Computer Information Technologies, ACIT 2020, 306–310, [http://doi.org/10.1109/ACIT49673.2020.9208831].

[11] Pham Q. et al.: A Survey of Multi-Access Edge Computing in 5G and Beyond: Fundamentals, Technology Integration, and State-of-the-Art. IEEE Access 8, 2020, 116974–117017, [http://doi.org/10.1109/ACCESS.2020.3001277].

[12] Ren Y. et al.: Dynamic Auto Scaling Algorithm (DASA) for 5G Mobile Networks. 2016 IEEE Global Communications Conference (GLOBECOM), 2016, 1–6, [http://doi.org/10.1109/GLOCOM.2016.7841759].

[13] Růzičková M. et al.: The estimation of the dynamics of indirect control switching systems. Tatra Mountains Mathematical Publications 48(1), 2011, 197–213, [http://doi.org/10.2478/v10127-011-0018-0].

[14] Saad W. et al.: A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems. IEEE Network 34(3), 2020, 134–142, [http://doi.org/10.1109/MNET.001.1900287].

[15] Satsyk, V. et al.: Reduction of Server Load by Means of CMS Drupal. 10th International Conference on Advanced Computer Information Technologies ACIT 2020, 523–528, [http://doi.org/10.1109/ACIT49673.2020.9208874].

[16] Shafi M. et al.: 5G: A Tutorial Overview of Standards, Trials, Challenges, Deployment, and Practice. IEEE Journal on Selected Areas in Communications 35(6), 2017, 1201–1221, [http://doi.org/10.1109/JSAC.2017.2692307].

[17] Tkachuk A. et al.: Basic Stations Work Optimization in Cellular Communication Network. D. Cagánová et al. (eds.), Advances in Industrial Internet of Things, Engineering and Management. EAI, Springer Innovations in Communication and Computing 2021, 1–19 [http://doi.org/10.1007/978-3-030-69705-1_1].

[18] Toroshanko Y. et al.: Control of Traffic Streams with the Multi-Rate Token Bucket. International Conference on Advanced Information and Communications Technologies – AICT 2019, 352–355, [http://doi.org/10.1109/AIACT.2019.8847860].

[19] Wu L., Lin Y.: Authentication Vector Management for UMTS. IEEE Transactions on Wireless Communications 6(11), 2007, 4101–4107, [http://doi.org/10.1109/TWC.2007.060245].

[20] Xu L. et al.: A Comprehensive Operation and Revenue Analysis Algorithm for LTE/5G Wireless System Based on Telecom Operator Data. IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation, 2019, 1521–1524, [http://doi.org/10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00274].

[21] Zhang Y., Fujise M.: An improvement for authentication protocol in third-generation wireless networks. IEEE Transactions on Wireless Communications 5(9), 2006, 2348–2352, [http://doi.org/10.1109/TWC.2006.1687756].

Ph.D. Serhii Moroz
e-mail: s.moroz@lntu.edu.ua

Author of more than 100 scientific papers, 7 patents for utility models. Research interests: Methodological bases of construction of sensors of physical quantities and measuring modules, Research of info-communication systems in the context of the development of the concept of the Internet of Things.

http://orcid.org/0000-0003-4677-5170

Ph.D. Anatolii Tkachuk
e-mail: a.tkachuk@lntu.edu.ua

Vice-dean for R&D Faculty of Computer and Information Technologies, Lutsk National Technical University.
Member of European Alliance for Innovation (EAI), International Association for Technological Development & Innovations (IATDI).
http://www.researchgate.net/profile/Anatolii-Tkachuk

http://orcid.org/0000-0001-9085-7777

Ph.D. Mykola Khvyshchun
e-mail: m.khvyshchun@lntu.edu.ua

Author of more than 100 scientific papers, 2 patents for utility models. Research interests: Investigation of semiconductor materials, Research of radiation-protective epoxy composite coating for n-Ge and n-Si single crystals.

http://orcid.org/0000-0002-3918-4527

Ph.D. Stanislav Prystupa
e-mail: s.prystupa@lntu.edu.ua

Author of more than 60 scientific papers, 5 patents for utility models. Research interests: design of microelectronic intelligent sensors of physical quantities.

http://orcid.org/0000-0003-3705-1541

Ph.D. Mykola Yevsiuk
e-mail: m.yevsiuk@lutsk-ntu.com.ua

Research interests: telecommunication networks, radio engineering devices, power supply systems of radio engineering devices and systems. Author and co-author of about 40 scientific articles, two textbooks, two monographs

http://orcid.org/0000-0002-3768-8959

http://doi.org/10.35784/iapgos.2892 *received: 10.03.2022 | revised: 18.03.2022 | accepted: 29.03.2022 | available online: 31.03.2022*

HYBRID NAVIGATION SYSTEM FOR INDOOR USE

Michał Styła[1], Przemysław Adamkiewicz[1,2]
[1]Information Technology Research & Development Center (CBRTI sp. z o.o.), Lublin, Poland, [2]University of Economics and Innovation, Faculty of Transport and Information Technology, Lublin, Poland

Abstract. *This article describes the design and implementation of a hybrid in-building navigation system. The word hybrid has a twofold meaning in this case. On the one hand, it refers to the use of two tracking methods: demanding (beacons) and not requiring an electronic device (radio tomography imaging). On the other hand, it specifies several commercial wireless communication protocols that make up the presented system. Ultimately, the network created in this way will be designed to provide the user with location and navigation services with increased accuracy and reliability. The text describes both the topology of created networks, methods of communication between devices and their hardware layer, as well as the effects of work resulting from the actual test object.*

Keywords: bluetooth, electromagnetic propagation, indoor radio communication, radiofrequency integrated circuits, radio navigation, tomography

HYBRYDOWY SYSTEM NAWIGACJI DO UŻYTKU WEWNĄTRZ POMIESZCZEŃ

Streszczenie. *Artykuł opisuje projekt i sposób realizacji hybrydowego systemu nawigacji wewnątrzbudynkowej. Słowo hybrydowy ma w tym przypadku dwojakie znaczenie. Z jednej strony odnosi się do zastosowania dwóch metod namierzania: wymagającej (radiolatarnie) i nie wymagającej posiadania urządzenia elektronicznego (obrazowanie radio-tomograficzne). Z drugiej wyszczególnia kilka komercyjnych protokołów komunikacji bezprzewodowej składającej się na przedstawiony system. Docelowo utworzone w ten sposób sieć będzie miała za zadanie świadczyć użytkownikowi usługi lokalizacyjne i nawigacyjne o zwiększonej dokładności i niezawodności. Treść tekstu opisuje zarówno topologię tworzonych sieci, metody komunikacji między urządzeniami oraz ich warstwę sprzętową jak i efekty prac wynikłych na podstawie rzeczywistego obiektu testowego.*

Słowa kluczowe: bluetooth, propagacja fal elektromagnetycznych, systemy radiowe do użytku wewnętrznego, układy scalone o częstotliwości radiowej, nawigacja radiowa, tomografia

Introduction

Hybrid navigation system is a wireless network consisting of a given number of radio measuring sensors and associated monitoring devices. It combines the features of three radio technologies: Bluetooth, Wi-Fi (IEEE 802.11) and IEEE 802.15.4 protocol (e.g. XBee modules). The system covers the ISM 2.4 GHz radio band (Industrial, Scientific, Medical), which works well for the purpose of detecting living beings in closed rooms due to the high percentage of water content in organisms. To a lesser extent, the system can also be used to detect inanimate objects depending on what material they are made of. This is due to the nature of the electromagnetic wave and its tendency to be reflected and absorbed depending on the type of matter it encounters. The basic task of the presented system is to guarantee all users within the range the data necessary to determine their location relative to the monitored object (coarse radio beacon navigation). If the user enters the room, the system can additionally increase its measuring accuracy by means of radio tomography, whose algorithms have been implemented in electronic systems (diminutive navigation).

1. Beacon system structure

The beacon network is a set of highly energy-efficient radio signal transmitters with which the user's device (equipped with proprietary software) is able to determine its own position and plot it on a planar model of the building. The main point of reference here is the calculation performed on RSSIs, which are reference radio signal values expressed in decibels per milliwatt (dBm). The user's device, against a comparison of signal sources from different transmitting stations, becomes capable of providing location services to the user. While this method works well in large-area facilities and is easily scalable, it is unfortunately subject to relatively high inaccuracy. For this reason, it plays the role of navigation along main traffic routes in the system. The RTI system (radio tomographic imaging) was designed and intended for precise tracking.

The radio beacon is a circular PCB with a diameter of 30 mm. This device is designed to work with battery power supply, which in combination with the small size of the laminate allows for installation even in hard to reach places of the building. Due to the intermittent nature of the transmitter's work, the device goes into sleep mode most of the time. Thanks to this procedure, an operating time of several months has been achieved. This statement applies to a CR2077 lithium battery with a capacity of 960 mAh. Together with the basket the overall dimensions of the device increase up to 20 mm in thickness giving approximately a cylindrical shape.

The visualisation of the device can be traced in figure 1, while the block diagram is presented in Figure 4 but reduced from the stationary power supply (power conditioning system and USB port) and the IEEE 802.15.4 communication module section together with the antenna.

The heart of the device is the nRF52832 microcontroller, which, due to its internal architecture and the presence of a radio layer capable of transmitting in the 2.4 GHz ISM band (modulation with phase keying and approximation using a Gauss curve), enables the implementation of such wireless communication technologies as Bluetooth LE or ZigBee. The rest of the electronic structure consists of the components necessary for the radio layer and the CPU (central processing unit) to function properly. These include two quartz resonators: 32 MHz (performance) and 32.768 kHz (power saving), signalling, interference filters, antenna path, programmer port and RGB LED signalling.

It should be emphasized that due to the nature of communication between the beacons and the user's device (one-to-many communication), no permanent connections are created between them, as e.g. in the case of Bluetooth Classic and streaming. Beacons will use the advertising offered by the fifth version of Bluetooth LE technology. It causes that each of the radio transmitters behaves like an individual, miniature broadcasting station that uses only cyclic transmissions of data packets without the need to acknowledge their receipt.

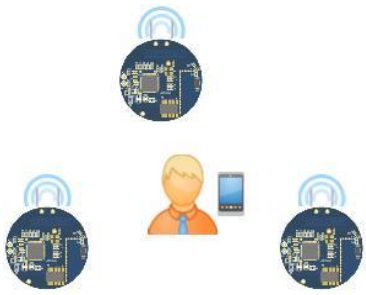


Fig. 1. Graphical presentation of the concept of beacon navigation

2. Structure of the radio tomography system

The presented RTI system consists of a set of sixteen hybrid measurement probes, data acquisition device, Apache Kafka data bus and a computational unit performing reconstruction for the needs of radio tomography. The dependencies exist between the radio probes are presented in figure 2.

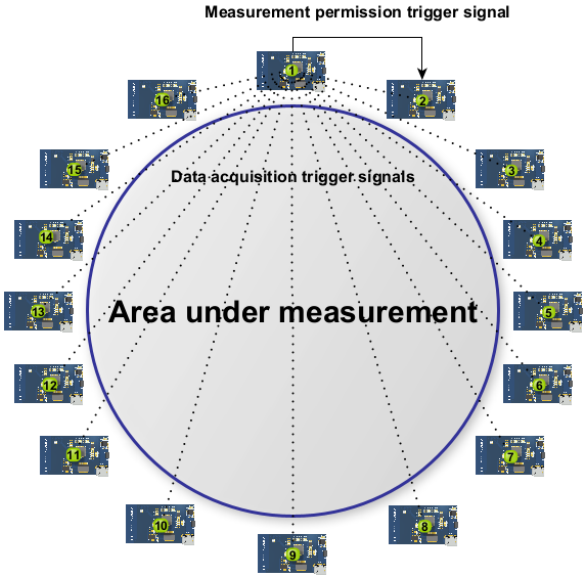


Fig. 2. The structure of the hybrid radio tomography system inside the single test room with a resolution of 16×16 (measuring sequence for probe number 1)

The structure of RTI navigation system also does not maintain permanent network connections between devices (in particular by means of probes), and its operation is based on Bluetooth technology (in one-to-many radio transmission variant) and the IEEE 802.15.4 (in one-to-one radio transmission variant) [3]. In practice, this means that the hybrid radio sensor is able to act independently and engage in various tracking processes without having to cooperate with other elements of the system. However, despite the relative independence of the units and the abandonment of the coordination of transmitters by means of a central unit, the synchronisation burden was transferred to the radio probes [1].

After the exchange of information between probes on RSSIs, the trigger signal is sent to the next probe by sequence number. The resulting data (contained in a Bluetooth LE version 5 compatible communication package) is acquired by the central control unit and treated as a single row of the measurement matrix needed to perform the reconstruction. After completing all rows of the matrix (one from each radio probe), they are properly segregated and sent via the Apache Kafka data bus to the reconstruction server (network connection – via twisted-pair Ethernet or Wi-Fi wireless connection), and then returned via the network in the form of reconstruction to a designated device equipped with a display [5].

In order to better understand the processes taking place between devices, one should start with the structure of the communication frame and the types of data carried by the devices.

As mentioned, the RSSI indicator was used for determining the user's position in both radio beacons and RTI, but it is processed and compiled differently. In communication by beacons, only the signal value at a given time and at least three beacons are enough. To enable human detection in radio tomography, differential analysis of two arrays was used. The first matrix is a set of projection angles (i.e. mutual relations of signal strengths received alternately by all RTI probes) of the monitored area in a rest state (devoid of a living factor). Such a matrix is called a background matrix. Then, measurements of a normally used room are made and the obtained measurement matrix is compared with the background matrix using appropriate algorithms [2].

The source of the signals can be both the Bluetooth 5 (speed) and IEEE 802.15.4 transmitter (smaller standard deviations of the RSSI results in the system equilibrium). In the presented solution, IEEE 802.15.4 was used as a reference signal source, and Bluetooth 5 as a wireless transport layer.

Microcontrollers of radio probes, by means of a serial port connection with a radio module compatible with the IEEE 802.15.4 standard, issue cyclical commands to him to obtain information on RSSI. A single measurement cycle is presented in figure 2 for the probe case number one. After the values are obtained, they go to the serial bus buffer, which is the same route as the commands are issued. Based on the unique addresses, the values are recognized and sorted in such a way as to form a matrix row. The matrix row (depending on its length) is placed in the data frame shown in figure 3.

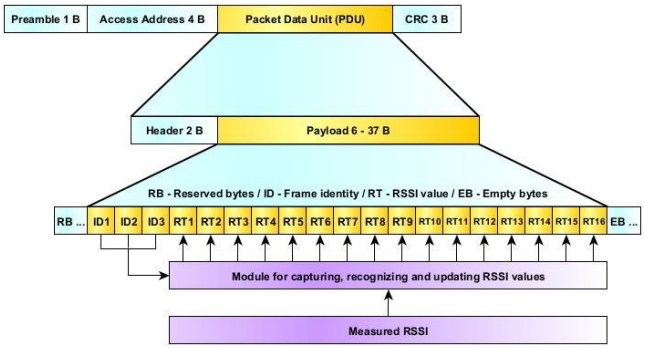


Fig. 3. Communication package created for the transport of RSSI indicators to the central control unit

The RSSI indicators in number depending on the system size are sequentially folded together with the identification bytes in the container for user data (Bluetooth LE advertising mode). The ID bytes allow the CPU to recognize which probe the row of the matrix comes from and where it should place it in relation to the others. The number of ID bytes may vary depending on the need (to reduce the probability of transmission errors or system size changes).

The data frame prepared in this way is cyclically broadcast every 100 milliseconds. The BLE scanner implemented in the central unit scans the environment in search of packets compatible with the conditions written in the code and only identical to the content offered by RTI probes. This process occurs every 50 milliseconds and lasts 50 milliseconds (measurement window) [3].

The RSSIs inserted in the data frame are calculated according to the algorithm implemented in the IEEE 802.15.4 protocol and are sent in a ready-to-use form via the serial port from the XBee radio module to the microcontroller. These are values stored on a single byte in U2 code (with sign). This is due to the fact that an electromagnetic wave is not able to pass through a medium such as air, water or building materials without loss. This makes the real values oscillate between -20 dBm down to about -100 dBm, which is the so-called sensitivity, characteristic for modern commercial radio systems. Further development of data structures was therefore not logically justified. The only exception is the desire to increase the number of identification bytes or to increase the number of transmitters. Then it is possible to expand a single frame to 20 bytes of RSSI value. Further scaling of the system required the use of the Bluetooth LE scan response mode (allows a second, additional data frame to be transmitted in response to a scanner request), splitting the matrix row into an older and a younger part and transferring them alternately, or using extended Bluetooth LE data packets.

The data collected by the CPU is collected using a proprietary adapter connected to the peripherals of the single chip computer via a serial port. The adapter is a technical variation of the radio beacon adapted primarily for scanning, pre-formatting data and continuous transfer to the system. Before this can happen, however, they must be received by the Jetson Nano or Raspberry

Pi computer, then formatted and nested in JSON. Only then are they ready to be put into the network and via the Apache Kafka data bus to the so-called cloud. From there, the reconstruction server retrieves them and starts computational sequences aimed at synthesising a heat map of the room under study.

3. RTI hardware solutions

According to the description of the system structure, each of the two parts of the hybrid navigation system comprises a certain pool of equipment necessary for operation. While in the case of the radio beacon it is one type of device, in the case of the developed RTI system it is a radio probe and a central unit, which consists of several components. In addition, all the communication devices that are not original technical solutions, but mediate in the transmission of data after entering the Internet network, should be specified.

3.1. Radio probes

The radio probe is a development of the radio beacon base design. The core of the device is still the nRF52832 microcontroller. The block diagram of the device is shown in figure 4, while the PCB appearance is presented in figure 5. It is distinguished from the radio beacon by the presence of a stationary power supply system realised in the form of a USB port and a voltage stabilizer. The stabilizer reduces the voltage of 5 V to the level required for proper operation of the nRF52832 microcontroller system, i.e. 3.3 V. This makes it possible to realise an RTI network using a regular cable for mobile device chargers. This procedure was intended to simplify the testing process. At the same time, for the purposes of building installations, two types of proprietary power supplies have been designed to work with 230 VAC electrical installation and PoE (Power over Ethernet), respectively. The power supplies not exceeding the dimensions of a matchbox in combination with a 30 mm radio transmitter make it possible to install it in building junction boxes or lighting fixtures. During the commissioning tests in the lighting fixture, the operating conditions of the electronics were also examined in terms of temperature. Based on the readings from the thermal imaging camera, the maximum was recorded on the body of the voltage converter, but it did not exceed 32°C due to the high efficiency reaching 90% and very low load energy consumption [4].

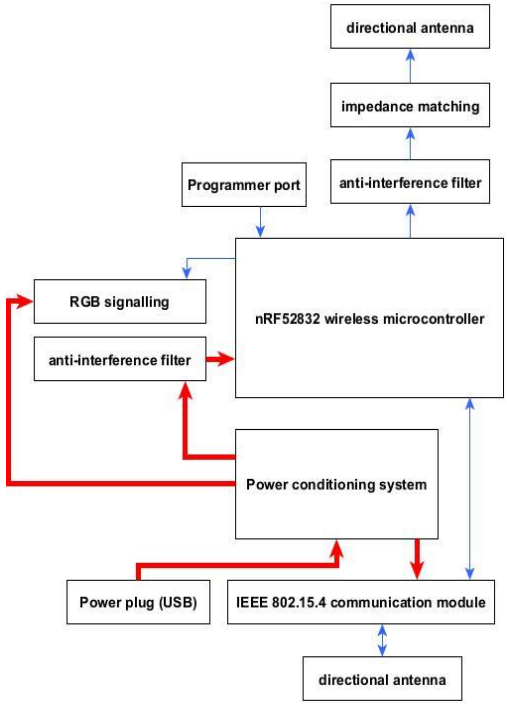


Fig. 4. Block diagram of the radio probe, where bold arrows denote power supply lines and thin arrows denote signal lines

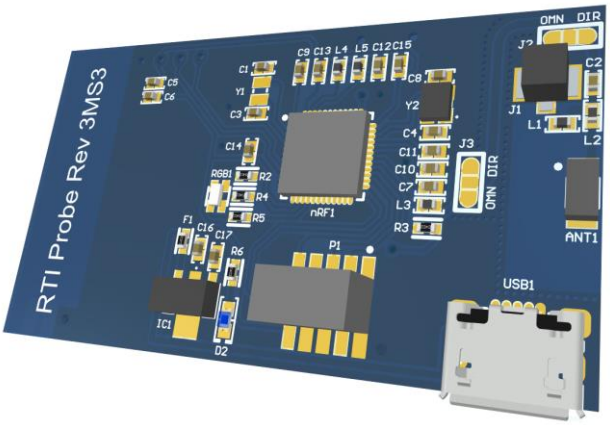


Fig. 5. Three-dimensional visualization of a complex radio probe PCB

Another significant difference is the presence of two radio modules, one of which (Bluetooth) acts as a transmitter, and the other (IEEE 802.15.4) as a transmitter and a receiver. It required a collision-free installation of two antennas in one housing. that is why the Bluetooth transmitter has a ceramic antenna located on one edge of the PCB, and the antenna lines of the XBee module were pulled away from the board using the u.FL connector and the antenna equipped with a 10 cm shielded cable. This gives a greater possibility of avoiding electromagnetic wave interference from both sources. It should be emphasized that both transmitters use the same band and modulation. The differences result from the division of the band into channels, of which in the case of Bluetooth LE, the channel number 37 (2402 MHz), 38 (2426 MHz) and 39 (2480 MHz) should be specified, which are used alternately before the broadcast mode. The BLE channels used by the RTI system are shown on the RSSI monitor in figure 6. In the case of IEEE 802.15.4 and XBee modules, the channel number is automatically determined by the pair of data exchange devices. It is possible to manually change the communication channel, but it has no performance justification. Efficiency of the XBee module buffer and the measured query and response time at the level from 50 milliseconds to 80 milliseconds (time measured in the system connected to the manufacturer's software used for testing and updating the software) [1].

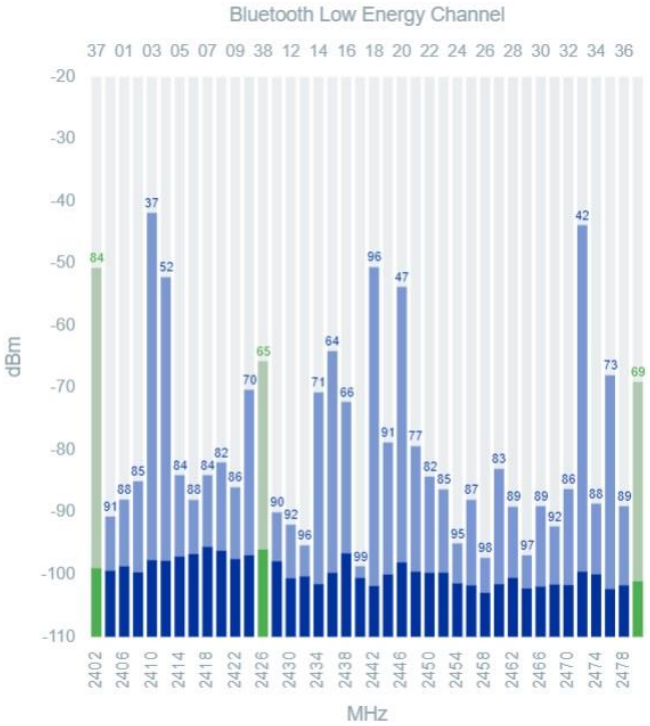


Fig. 6. BLE channels used for data transmission from radio probes to the central unit (specified in a different color with the frequencies 2402 MHz, 2426 MHz and 2480 MHz)

Due to the high sensitivity of radio systems, appropriate filtration of the supply voltage and unnecessary noise from the antenna system was taken care of. To communicate with the user, the probe uses a programmable RGB diode, which marks the processes it carries out (stimulation, measurement and triggering another probe) with an appropriate colour. Thanks to this, it is possible to estimate the speed of the matrix acquisition by visual observation of the activation of successive radio transmitters (e.g. with the help of a stopwatch). Typically, data acquisition and transfer over the PAN does not take more than a few seconds. The measurement time is strictly dependent on the software parameters of the broadcast frequency, scanning frequency, measurement window width, and also within a certain range of the power level. Insufficient power can result in more frequent transmission errors and an increase in retransmission requests. Too high power of the transmitter, in turn, increases the reflection of EM waves in the room, which translates into the level of interference and noise in the image. Noises are manifested in the form of hard to justify, the so-called artifacts in the heat map. The above-mentioned facts make the RTI system implementation for each room different and it is necessary to properly calibrate it (including the appropriate arrangement of the probes so as to maintain the maximum density of the grid made of projection angles) [7].

3.2. Control unit

While the radio probes are responsible for data acquisition, mutual coordination and data transport through the PAN (Personal Area Network) network, the first hardware unit that can be considered as a kind of "operator panel" is the central unit. This unit is entrusted with the operations related to the acquisition and formatting of the entire array, as well as the processes of data transfer and maintaining the connection with the reconstruction server. A Jetson Nano single-board computer was selected to provide the appropriate programming tools and performance (included). The block structure of the device built on the basis of this computer is presented in figure 7.

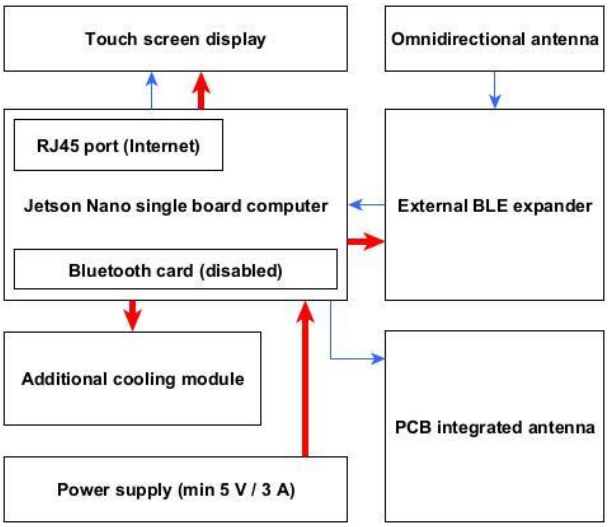


Fig. 7. Block diagram of the central unit, where bold arrows denote power supply lines and thin arrows denote signal lines

The CU consists of components such as a 7 inch TFT touchscreen display for convenient presentation of created arrays. It is connected via a dedicated video connector on the Jetson Nano computer. Although the selected platform already has an integrated 2.4 GHz antenna and the ability to communicate with the world using BLE and Wi-Fi, after performance and stability tests it was decided to introduce modifications. Tests were carried out on a real zone and measurement facility and found frequent hang-ups and connection drops with the built-in Bluetooth hardware layer. This necessitated the need to design and test a custom expander that would be more susceptible to modification and allow more interference with the software layer. It was connected to the computer via a serial port. The external BLE expander is actually a miniature 2.4 GHz receiver (built using an nRF series microcontroller). It offers the possibility of connecting both a directional ceramic antenna and an omnidirectional antenna (this is another advantage of not using the built-in BLE module).

The entire design is powered by a dedicated 5 V / 3 A power supply. Due to the increased temperature of the processor system and during intensive work, the PCB was enriched with an aluminum heatsink with a fan. The reasonableness of the cooling assembly is also supported by the intended use of the CU to work in a sealed enclosure. The first test case was made with the help of a 3D printer and plastic. It included holes for the screen and a thread for mounting an omnidirectional antenna of any length. The device with its surface area does not exceed the size of an A4 sheet of paper and, depending on the needs, can be quadrupled in size provided that the headless version (i.e. screenless, without GUI) is used. A version with a panel was created for the tests. The system managing the CU was Linux OS, while Qt Creator and C++ language was used as a programming environment. The complete prototype is shown in figure 8.

Fig. 8. Complete prototype of central processing unit during RSSI matrix acquisition

4. Research

In order to carry out the image reconstruction process, it is necessary to start with the so-called influence matrix. It is obtained using the Fresnel zone model (a model describing the relationship between the radio signal transmitter and receiver) in calculations. Then the inverse problem must be solved using Tychonoff's regularisation with a lambda of 19 thousand. This parameter was set arbitrarily. The mentioned processes lead to an inverse matrix that can be used to determine the vector image. In order for the RTI system to cover areas of any solid shape, the computational algorithm had to be enhanced with a convolution filter based on the Marr wavelet. After the image matrix is created, this filter takes it to itself to configure its value to 60 per cent of the maximum.

The tests were carried out in a gradual and successive manner. After combining the hardware and software layers, a test area was created in the form of an office room (which was in normal use) and then 4, then 8 and finally 16 radio probes were distributed around its perimeter. The central unit was positioned close to the centre of the room as intended. The shape of the measurement zone and the distribution of transmitters are presented in figure 9. The system reactions to the introduced disturbances were observed on the basis of changes in the heat map obtained by reconstruction. Example maps are also presented within figure 9. The test scenarios included different types of disturbance configurations. These ranged from a single stationary user to a moving group of people. It can therefore be concluded that the calibration of the RTI system was carried out experimentally, however, based on the research carried out, it was possible to formulate several assumptions according to which the system should be scaled.

By correctly defining the room model and accurately marking the positions of the transmitters on it, it was possible to conclude that the response of the heat map to disturbances occurring in the measurement zone was adequate as far as space was concerned. The only limitation in this respect was the density

of the grid created from the available projection angles. In some areas of the zone it was simply too small, which could cause slight shifts of maximum disturbance in the planar heat map in relation to the area in reality. This could have been counteracted by increasing the number of transmitters or by modelling the grid (changing the position of the transmitters).

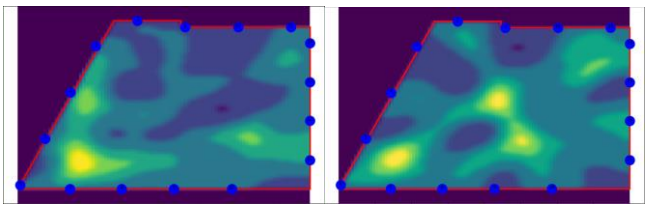


Fig. 9. Reconstructions of the room and the presence registered in it combined with the heat map (bright yellow spots indicate persons): one person in the bottom left corner of the room (on the left) and a total of three people in the corner and in the middle of the room (on the right)

To reconstruct an image like the one in figure 9 the equation must be solved:

$$J*e = m \quad (1)$$

where J represents the sensitivity matrix and the parameter m denotes the vector of measurements. All the resulting finite elements of the computational grid are denoted as e and treated as a vector.

In the case of the sensitivity matrix, it is built up by rows and columns, where each value contained in them corresponds to the sensitivity achieved by a selected reconstruction pixel per measurement for a given transmitter-receiver pair (in the case of the presented system, sixteen radio devices will form a 256-element matrix). The indexing of the fields of the sensitivity matrix J follows each possible combination of radio devices (transmitter and receiver).

$$J^{-1}*m = e \quad (2)$$

The study encountered a significant computational problem arising from the non-squaring of the sensitivity matrix J. This was true for almost all observed cases. This forced a search for a solution and eventually led to the identification of the reason as a bad conditioning of Tychonoff's regularization (Tychonoff's 1998). As mentioned, a configuration of sixteen transmitters should result in a vector of length equal to 256 elements. Nevertheless, there have been incidents where the number of values was able to exceed 1000. However, this did not trigger the abandonment of the search for inverse matrices with the help of Tychonoff's regularization. As far as the sensitivity matrix is concerned, it is determined assuming that the zone formed by the flow of electromagnetic waves between the transmitter and the receiver assumes an ellipsoidal shape (according to Fresnel's theorem). Of course, it is also assumed that inside the determined zone fluctuations of the RSSI value may occur [3, 6].

Entails huge computational resources. This is especially true when the sensitivity matrix J has not yet been determined because it is a reusable element. Once determined, it provides a benchmark for the room but under the condition that the hardware layer associated with the radio probes is not displaced. It is therefore necessary to keep this matrix in permanent memory in case the calculation is resumed. Once all computational processes have been established, the method achieves performance in terms of reconstruction speed in the order of 1 millisecond to 2 milliseconds per process.

5. Summary

Research on the system has shown that commercial radio technologies used in the construction of the assumed navigation system present the potential for expansion of this type of structure in the future. The efficiency of the system increased as the weaknesses of individual technologies were eliminated by the advantages of others (BLE, 802.15.4 – XBee, 802.11 – Wi-Fi) An office building was used as the object of research of the navigation system, and one of its rooms as a test area for a radio tomography. Beacons and probes have been located in characteristic floor points throughout their area. In the case of probes due to irregularities in the room, the devices were placed on special rotary hooks, which allowed correction of the antenna radiation angles. Thanks to the appropriate calibration of the transmitters and their placement at the appropriate height, at some moments it was possible to determine the circumference in the belt of the individual violating the test area with an accuracy of several cm.

However, it should be noted that the potential for ambiguity increased as the disturbance moved from the edge of the test area towards the centre of the room. It is also worth noting that continually increasing the number of transmitters in the same working area brings the system to a point where too close proximity of antennas will lead to a sharp increase in the level of wave interference, i.e. mutual interference. While in the case of radio beacons this is not so critical, in the case of radio probes moderation must be exercised. However, it is necessary to eliminate the so-called dead zones resulting from locally occurring too low grid densities. This type of event occurs when objects are crossed by only 1-2 angles of projection / beam. From a mathematical point of view, too few measurement points result in low spatial resolution.

References

[1] Bluetooth SIG. Bluetooth Core Specification, version 4.0. Bluetooth SIG: Kirkland, WA, USA, 2010.
[2] Kaltiokallio O., Bocca M.: Real-time intrusion detection and tracking in indoor environment through distributed RSSI processing. Proceedings of the 2011 IEEE 17th International Conference on Embedded and Real-Time Computing Systems and Applications, 2011, 61–70.
[3] Maj M., Rymarczyk T., Kania K., Niderla K., Styła M., Adamkiewicz P.: Application of the Fresnel zone and Free-space Path for image reconstruction in radio tomography. International Interdisciplinary PhD Workshop 2019 – IIPhDW 2019, Wismar, Germany.
[4] Rymarczyk T., Styła M., Oleszek M., Maj M., Kania K., Adamkiewicz P.: Object detection using radio imaging tomography and tomographic sensors, Przegląd Elektrotechniczny 96(1), 2020, 182–185.
[5] Styła M., Oleszek M., Rymarczyk T., Maj M., Adamkiewicz P.: Hybrid sensor for detection of objects using radio tomography. Proc. Applications of Electromagnetics in Modern Engineering and Medicine – PTZE 2019, 219–223.
[6] Vauhkonen M., Vadasz D., Karjalainen P. A., Somersalo E., Kaipio J. P.: Tikhonov regularization and prior information in electrical impedance tomography. IEEE Trans. Med. Imaging 17, 1998, 285–293.
[7] Wilson J., Patwari N.: See-through walls: Motion tracking using variance-based radio tomography networks. IEEE Trans. Mob. Comput. 10, 2010, 612–621.

M.Sc. Eng. Michał Styła
e-mail: michal.styla@cbrti.pl

Michał Styła received the M.Sc. Eng. degree in electrical engineering from the Lublin University of Technology, in 2018. Since 2020, he fulfills a role of specialist in the field of wireless communication solutions at CBRTI sp. z o.o. His field of activity encompasses design and programming of electronic circuits connected with radio communication techniques and their subsequent implementation in tracking, navigation and radio tomographic imaging systems.

http://orcid.org/0000-0002-1141-0887

Ph.D. Przemysław Adamkiewicz
e-mail: przemyslaw.adamkiewicz@cbrti.pl

Przemysław Adamkiewicz received the Ph.D. degree in physics from the Maria Curie-Sklodowska University, in 2013. Since 2021, he has been a CEO at CBRTI sp. z o.o.. Leader and participant in several international projects. His research area focuses on the application of embedded IoT solutions, electrical tomography, image reconstruction, process tomography, radio tomography imaging, artificial intelligence and computer measurement systems.

http://orcid.org/0000-0003-3425-9566

http://doi.org/10.35784/iapgos.2895 received: 10.03.2022 | revised: 18.03.2022 | accepted: 29.03.2022 | available online: 31.03.2022

INDOOR LOCALIZATION SYSTEM USING UWB

Dominik Gnaś[1], Przemysław Adamkiewicz[2,3]
[1]Netrix S.A. Research & Development Center, Lublin, Poland, [2]Information Technology Research & Development Center (CBRTI sp. z o.o.), Lublin, Poland, [3]University of Economics and Innovation, Faculty of Transport and Information Technology, Lublin, Poland

Abstract*. This paper discusses two ways of measuring the distance between the transmitter and the receiver using UWB technology, then identifies their advantages and disadvantages. The method of calculating the position is presented together with the method of predicting errors based on room geometry. The hardware configuration of the transmitter and receiver systems included in a location system based on UWB technology is explained. Bluetooth technology is discussed, which is used to connect the location system with the environmental monitoring system.*

Keywords: bluetooth, identity management systems, indoor radio communication, logistics, navigation, ultra-wideband technology

WEWNĘTRZNY SYSTEM LOKALIZACJI Z WYKORZYSTANIEM UWB

Streszczenie. *W niniejszym artykule omówiono dwa sposoby pomiaru odległości między nadajnikiem, a odbiornikiem wykonanymi w technologii UWB, następnie określono ich wady oraz zalety. Przedstawiono sposób obliczania pozycji wraz z metodą przewidywania błędów w oparciu o geometrie pomieszczenia. Wyjaśniono konfigurację sprzętową nadajnika oraz odbiornika układów wchodzących w skład systemu lokalizacji opartych o technologię UWB. Omówiono technologię Bluetooth, która wykorzystana została do połączenia ze sobą systemu lokalizacji wraz z systemem nadzorowania warunków środowiskowych.*

Słowa kluczowe: bluetooth, komunikacja radiowa wewnątrz pomieszczeń, logistyka, nawigacja, systemy zarządzania tożsamością, technologia ultra szerokopasmowa

Introduction

Accurate determination of the position of objects and people in closed spaces with high accuracy is something desired by the consumer and industrial market. GPS technology that works properly in open space and provides accuracy on the level of several dozen meters, which is sufficient to locate the building. For a simple location inside the building, it is possible to use Bluetooth beacons that calculate the distance based on the RSSI, which gives the measurement accuracy within one to two meters. However, in confined spaces such as warehouses, much higher accuracy is required. Therefore, for indoor locations where high accuracy of location measurement is required, a technology that ensures accuracy at the level of several to several dozen centimetres should be used. The best solution to the problem of indoor location is to use UWB technology with the calculation of the location of the object based on the response time of the transmitter. In UWB technology, it is also possible to calculate the position based on the RSSI. RTLS (Real-Time Location System) is based mainly on UWB (Ultra-Wide Band) technology, which works in high-frequency radio technology with a wide signal band [1].

1. The methods of measuring distance

In UWB technology, two distance measurement methods can be distinguished, such as RSSI (Received Signal Strength Indication) and ToF (Time of Flight) measurement. The first method discussed is the localization of objects using RSSI. It allows determining the signal strength between the sensor and individual anchors, in this method one transmitter and one receiver is used. This method is presented in figure 1. The accuracy of object localization is in the range of 0.5 m – 1 m. Equation 1 presents the method of calculating RSSI:

$$RSSI = 10log_{10}\left(\frac{C2^{17}}{n^2}\right) - A[dBm] \quad (1)$$

where:
C – channel impulse response power,
N – number of preamble accumulation,
A – constant name of the transmission frequency.

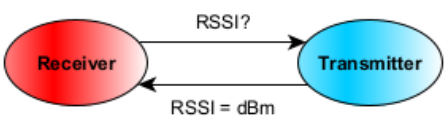


Fig. 1. Method of distance measurement by asking for RSSI

The individual values can be read from the internal registers of the device. The calculated value is expressed in dBm. It is the unit of absolute level related to the power of 1 mW. The method used is an estimation method, i.e. it gives an approximate value of signal power.

The second discussed the method of determining the distance is ToF (Time of Flight). The simplified principle of operation is shown in figure 2. The operation of this method consists in determining the time of flight of the frame between the transmitting and receiving device. Knowing the time of transmission and the speed at which radio waves move (the speed of light) we can determine the distance with high accuracy. This method allows for tracking many objects at the same time and makes it possible to locate them with the accuracy of several dozen centimetres. The disadvantage of this method is the necessity of having at least three receivers so that the measuring system could determine the position in space in an unprecedented way [2].

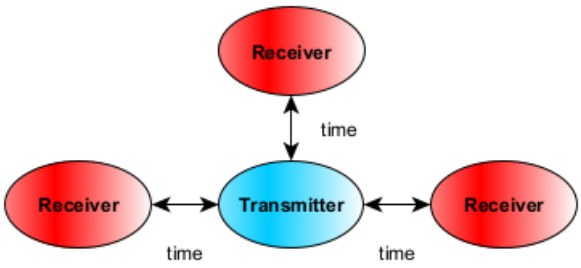


Fig. 2. Method of measuring distance by measuring the response time

Figure 3 shows the principle of trilateration, which is used to locate objects, a similar localization method is used by GPS. Equations 2, 3, and 4 show how to calculate the location of an object in two-dimensional space. The distances d1, d2, d3 are calculated based on the response time of the transmitter. Furthermore, in this localization method, the position of the anchors in the room under study must be considered. To obtain the transit time, the sensor sends a frame while specifying the send time. After receiving the frame the particular anchor determines the time of receiving the frame and the expected time of sending the frame, which are included in the return frame [7]. After receiving the return message, the sensor determines the transit time taking into account the time needed for operations performed by the anchor. In order to obtain the most accurate result, the software also takes into account the antenna delay i.e. the time needed for the antenna to propagate the received signal. The obtained value determines the round trip time, so it should be

divided into two. In the final stage the obtained value should be multiplied by the speed of light, which gives the final value of the distance between the sensor and the anchor expressed in centimetres. With proper calibration, it is possible to achieve an accuracy of ±10 cm [4].

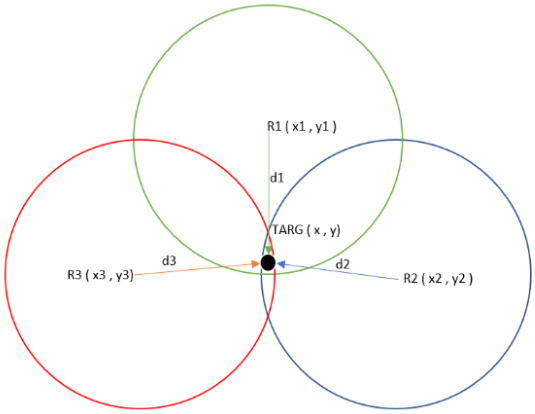


Fig. 3. Trilateration estimation

The distances d_i are given by the following equation:

$$\begin{cases} d_1 = (x_1 - x)^2 + (y_1 - y)^2 \\ d_2 = (x_2 - x)^2 + (y_2 - y)^2 \\ d_3 = (x_3 - x)^2 + (y_3 - y)^2 \end{cases} \quad (2)$$

Solving this system of equation reveals the tag position (x, y) as follows:

$$x = \frac{Ay_{32} + By_{13} + Cy_{21}}{2(x_1 y_{32} + x_2 y_{13} + x_3 y_{21})} \quad (3)$$

$$y = \frac{Ax_{32} + Bx_{13} + Cx_{21}}{2(y_1 x_{32} + y_2 x_{13} + y_3 x_{21})} \quad (4)$$

where:

$$\begin{cases} A = x_1^2 + y_1^2 - d_1^2 \\ B = x_2^2 + y_2^2 - d_2^2 \\ C = x_3^2 + y_3^2 - d_3^2 \end{cases} \quad (5)$$

$$\begin{cases} x_{32} = x_3 - x_2 \\ x_{13} = x_1 - x_3 \\ x_{21} = x_2 - x_1 \end{cases} \quad (6)$$

$$\begin{cases} y_{32} = y_3 - y_2 \\ y_{13} = y_1 - y_3 \\ y_{21} = y_2 - y_1 \end{cases} \quad (7)$$

The highest measurement accuracy can be achieved in rooms with simple geometry, in which there are no objects blocking or attenuating the electromagnetic wave (laboratory conditions). In order to improve the accuracy of the measurements, measurement errors should be determined. The formula for calculating the mean square error was used to determine the measurement errors (equation 7). In statistics, the mean squared error of an estimator measures the mean squared error, i.e. the mean squared difference between the estimated values and the true values an example is shown in figure 4 [3] .

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(y_t - y_y^P)^2 \quad (8)$$

where:
n – number of data points,
t = 1, 2, 3, …, n,
y_t – observed value,
y_t^P – predicted value.

Fig. 4. Average squared error

Table 1. Differences between RSSI and ToF measurement method

Method of measuring	RSSI	Response time
Accuracy of measurement	1-2 m	10-40 cm
Minimal number of receivers (anchors)	1	3
Data acquisition time	Fast	Slow
The complexity of the measurement system	Easy	Complicated

2. Hardware

The design of the device was based on the following assumptions:

- possibility of battery and external power supply,
- small size,
- control unit with Bluetooth communication module,
- sensors for environmental conditions,
- as simple as possible design.

The designed device consists of three main sections:

- power supply section,
- section with the NRF52832 control circuit,
- section with the radio module,
- section with the sensor environmental conditions.

The power supply system has been designed to be powered from batteries (Li-lon and Li-Pol) or from an external power source. The external power supply has been separated from the measurement system by an isolated DC/DC converter due to the high sensitivity of radio systems to conducted interference. Additionally, an external power source is used to charge the electrochemical cell. For this purpose, the BQ24092DGQT chip is used, which manages the working of the battery and enables its charging. To be able to switch between power sources, an additional P-MOSFET transistor is used to switch on the power from the battery if the power cable is not connected.

For the NRF52832 chip, an additional external antenna was required. It was decided to use the 2450AT18A100E single-chip antenna. This antenna was designed for 2400–2500 MHz, the frequencies most suitable for Bluetooth applications. The UWB chip used has a built-in omnidirectional antenna which guarantees similar output signal gain in different directions.

In addition, the project used environmental condition sensors such as a temperature sensor, a humidity sensor, and a pressure sensor. All sensors used the I2C data bus. The anchor and sensor used the same schematic and PCB. In the case of anchor, sensors were not soldered as well as batteries were not connected as they were powered from the external power source. Figure 5 shows a simplified schematic diagram of the transmitter and receiver (differences in the internal structures of the devices are visible).

The device also uses additional elements such as diodes, buttons, and connectors. The diodes used allowed, among other things, to verify the correctness of the device's power supply or battery charging. The use of buttons allowed to reset the device and influence the program operation.

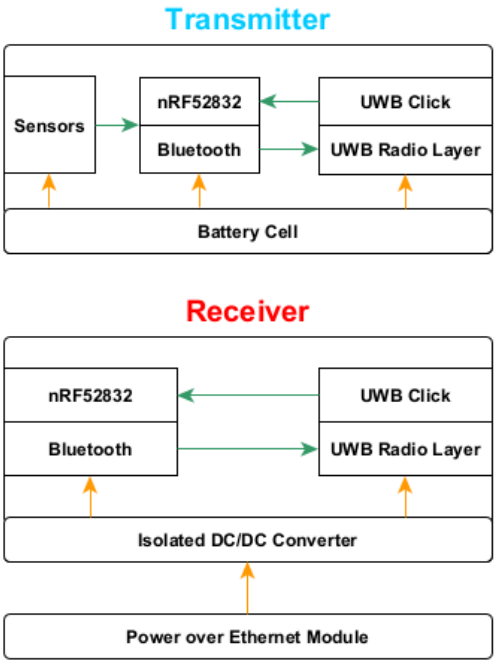


Fig. 5. Simplified circuit diagram

The following figure shows the electrical diagram of the building location system. When designing the scheme, an attempt was made to reduce its complexity, therefore a ready-made UWB module was used.

Based on the schematic diagram, a PCB design was created. It is a double-sided PCB with a thickness of 1.6mm. The external dimensions of the board are approximately 35 mm × 45 mm. Elements responsible for communication, i.e. UWB module and Bluetooth antenna are placed in the upper part of the device. This solution was decided upon in order to avoid signal interference and to achieve the most stable microwave and Bluetooth communication.

The DW1000 microwave module only requires a power supply and an SPI communication bus for basic operation. To ensure its proper operation, it was decided to use paths of the same length for each of the bus signals. A suitable design environment tool was used for this purpose. To equalize the length of the paths, it increases the length of the individual tracks by wrapping them in a small area.

For Bluetooth communication to work properly, it was necessary to adjust the impedance of the antenna path to a value of 50 Ω. To do this, a built-in Altium Designer environment tool was used to adjust the width of a given path to achieve the required impedance[5].

Mounting holes are also provided in the PCB design. This allows the device to be easily fixed in the enclosure. Figure 7 shows a 3D visualization of the developed PCB forming the in-building navigation system.

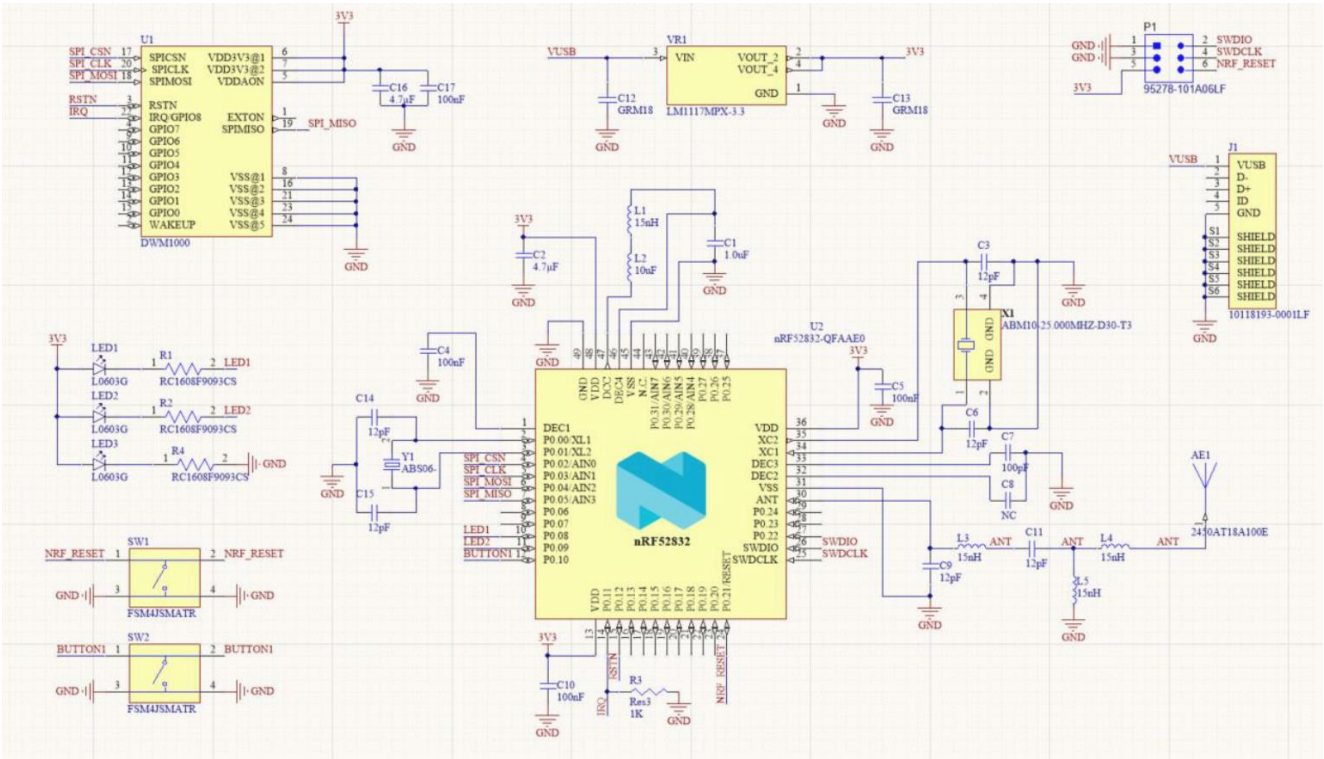


Fig. 6. Electrical diagram of the device

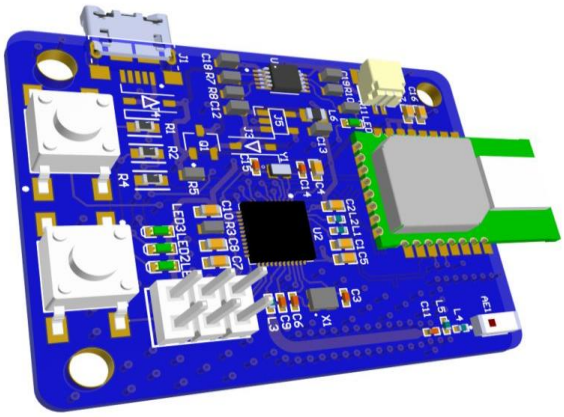


Fig. 7. 3D visualization of the device

2.1. Findings

Tests of the designed distance measuring system were carried out in laboratory conditions, i.e. in a 5 m × 5 m room with anchors placed in the corners of the room. Figure 8 shows a view of the room described. There were no objects in the room that could interfere with the signal emitted by the transmitters or receivers. The measurement procedure was started remotely using Bluetooth technology. The measurement started when a value of 1 was sent to a given anchor and then a return message of 0 was awaited. The message was sent via the Bluetooth UART service [5, 7].

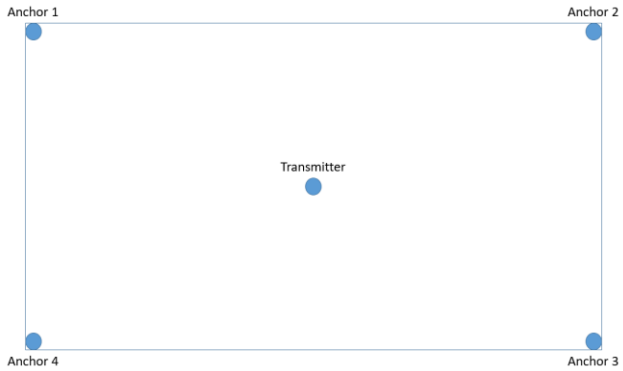


Fig. 8. Simplified diagram of the room

The mounts were designed to be attached to the wall with double-sided tape. The bracket has an articulated joint that allows the tilt of the mounted device to be adjusted. A view of the holder is shown in figure 9.

Fig. 9. Handle made with 3D printing technology

Before the measurements were made, the way in which the tests were to be carried out was defined, so the tests were carried out as follows:

- measuring the distance of the marker from the anchor with a measuring device,
- the transmitting and receiving devices shall be at the same height,
- return of antennas in the same position.

During the distance measurement with the UWB devices, the distance between the transmitting and receiving devices was measured simultaneously with a tape measure. This measurement allowed to verification the results obtained from the developed device. In addition, to make the measurements reliable and repeatable for each device, they were positioned so that they were at the same height. Additionally, the transmitting devices were directed so that the UWB antenna in the transmitter was facing the anchor. The reason for such an arrangement was to reduce interference resulting from wave reflections from walls as well as to shield the signal power in the form of printed circuit boards present on the PCB. Positioning the antenna in an appropriate way was very important in the case of testing the measurement method based on RSSI.

The graph below shows the signal measurement series. The sensor was moved closer to each anchor in turn, which is well reflected in the graph. It can be seen that as the distance increases, there is a greater scatter in the values received from the measuring system. The scatter of the measured values may be due to overlapping of the waves reflected from the walls of the room or incorrect positioning of the devices in relation to each other.

Fig. 10. Graph of the distance between individual anchors and the sensor

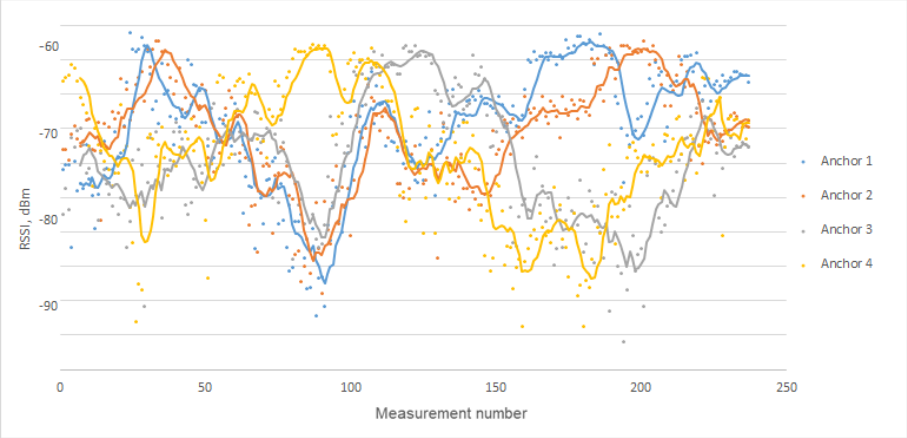


Fig. 11. Graph of RSSI values for individual anchors

The distortion in the RSSI method is much higher compared to the ToF method (this is shown in the figure 10). It can be observed that the greater the distance between the transmitter and the receiver, the less stable the signal strength was.

To make the graph more readable, a trend line based on the moving average of the values has been added to each series. This makes it clear when the market has moved away from a given anchor and when it has moved towards it. The distortions shown in figure 10 are most likely due to disturbances caused by electromagnetic waves reflecting off walls or imperfections in the antenna radiating the signal.

3. Summary

This paper presents a complete demonstration of UWB's indoor location. Two indoor localization methods are described, namely, a distance measurement method using signal strength determination (RSSI) and a frame transit time (ToF) measurement method. The research shows that the RSSI method is less accurate than the ToF method. After analysing the results obtained from RSSI measurement it was concluded that this method will not be taken into account in further works on the localization system due to easy interference and fluctuations of signal strength for unknown reasons. The measurement accuracy for the ToF method ranged from 10 cm – 50 cm. Some of the measurement errors could be due to imperfections related to and manual setting of the transmitter position. The ToF method, despite the longer data acquisition time, achieved much higher accuracy and was more resistant to interferences, which can be observed in the previously presented graphs.

References

[1] Fischer G. et al.: An impulse radio UWB transceiver with high-precision TOA measurement unit. International Conference on Indoor Positioning and Indoor Navigation (IPIN '10).
[2] Garbaruk M.: Time and frequency analysis of signals used in pulsed antenna ultra-wideband radiocommunication systems. 6th International Conference on Antenna Theory and Techniques, 2007, 276–277, [http://doi.org/10.1109/ICATT.2007.4425183].
[3] Kang M. et al.: NLOS mitigation for low-cost IR-UWB RTLS. IEEE International Conference on Ultra-Wideband (ICUWB), 2011, 96–100, [http://doi.org/10.1109/ICUWB.2011.6058931].
[4] Kolakowski J.: Application of ultra-fast comparator for UWB pulse time of arrival measurement. IEEE International Conference on Ultra-Wideband ICUWB '11.
[5] Maj M. et al.: Application o the Fresnel zone and Free-space Path for image reconstruction in radio tomography. International Interdisciplinary PhD Workshop – IIPhDW 2019, Wismar, Germany.
[6] Slovák J. et al.: RTLS tracking of material flow in order to reveal weak spots in production process. 22nd International Conference on Process Control – PC19, 2019, 234–238, [http://doi.org/10.1109/PC.2019.8815220].
[7] Styła M., Oleszek M., Rymarczyk T., Maj M., Adamkiewicz P.: Hybrid sensor for detection of objects using radio tomography. Applications of Electromagnetics in Modern Engineering and Medicine – PTZE, 2019, 219–223.

M.Sc. Eng. Dominik Gnaś
e-mail: dominik.gnas@netrix.com.pl

Dominik Gnaś graduated from Lublin University of Technology with a Master of Engineering degree in 2019 and is currently working at Netrix S.A. as an electronics programmer and designer. The area of research work includes wireless technologies, using known solutions as well as searching for new ones.

http://orcid.org/0000-0002-8169-9700

Ph.D. Przemysław Adamkiewicz
e-mail: przemyslaw.adamkiewicz@cbrti.pl

Przemysław Adamkiewicz received the Ph.D. degree in physics from the Maria Curie-Sklodowska University, in 2013. Since 2021, he has been a CEO at CBRTI sp. z o.o.. Leader and participant in several international projects. His research area focuses on the application of embedded IoT solutions, electrical tomography, image reconstruction, process tomography, radio tomography imaging, artificial intelligence and computer measurement systems.

http://orcid.org/0000-0003-3425-9566

http://doi.org/10.35784/iapgos.2883 *received: 20.02.2022 | revised: 16.03.2022 | accepted: 29.03.2022 | available online: 31.03.2022*

APPLICATION OF CONVOLUTIONAL NEURAL NETWORKS IN WALL MOISTURE IDENTIFICATION BY EIT METHOD

Grzegorz Kłosowski[1], Tomasz Rymarczyk[2]
[1]Lublin University of Technology, Faculty of Management, Lublin, Poland, [2]University of Economics and Innovation in Lublin, Institute of Computer Science and Innovative Technologies, Lublin, Poland

Abstract. *The article presents the results of research in the area of using deep neural networks to identify moisture inside the walls of buildings using electrical impedance tomography. Two deep neural networks were used to transform the input measurements into images of damp places - convolutional neural networks (CNN) and recurrent long short-term memory networks LSTM. After training both models, a comparative assessment of the results obtained thanks to them was made. The conclusions show that both models are highly utilitarian in the analyzed problem. However, slightly better results were obtained with the LSTM method.*

Keywords: machine learning, deep learning, electrical impedance tomography, moisture detection in walls

ZASTOSOWANIE KONWOLUCYJNYCH SIECI NEURONOWYCH W IDENTYFIKACJI ZAWILGOCEŃ ŚCIAN BUDYNKÓW METODĄ EIT

Streszczenie. *W artykule przedstawiono rezultaty badań w obszarze wykorzystania głębokich sieci neuronowych do identyfikacji zawilgoceń wewnątrz ścian budynków przy użyciu elektrycznej tomografii impedancyjnej. Do przekształcenia pomiarów wejściowych na obrazy przedstawiające zawilgocone miejsca użyto dwóch rodzajów głębokich sieci neuronowych – konwolucyjne sieci neuronowe (CNN) i rekurencyjne sieci typu LSTM. Po wytrenowaniu obu modeli dokonano oceny porównawczej uzyskanych dzięki nim rezultatów. Wnioski wskazują na dużą utylitarność obu modeli w badanej problematyce, jednak nieco lepsze rezultaty uzyskano dzięki metodzie LSTM.*

Słowa kluczowe: uczenie maszynowe, głębokie uczenie, tomografia impedancyjna, wykrywanie wilgoci w ścianach

Introduction

The presence of moisture inside the walls of buildings is the cause of many unfavourable phenomena. One of them is the chemical and physical degradation of walls, which leads to the weakening of the structure of buildings, reducing their strength. Wet walls crack at low temperatures due to freezing. Chemical compounds, especially aggressive salts and chlorides, which penetrate deep into the walls through faulty foundation insulation, accelerate the erosion of porous materials such as bricks, cement and plaster. The phenomenon of capillary leakage contributes to the spread of moisture areas inside the walls. Moisture destroys plasters and paints coats, worsening the aesthetics. It, in turn, necessitates more frequent renovations, which raises the operating costs of buildings [1]. A separate aspect is the negative impact of moisture on the health and comfort of people staying inside damp rooms. Due to favourable conditions, fungi and microorganisms often breed and multiply inside wet walls. Penetrating humans' breathe causes allergies and other respiratory illnesses [2].

This article presents a non-invasive method to identify moisture inside walls using electrical impedance tomography. Particular attention was paid to the issue of converting electrical measurements generated by tomograph electrodes into spatial images (reconstructions) visualizing moisture areas. The transformation of 96 input measurements into 6215 pixels of the output image resolves an inverse problem that is undefined [1]. For this purpose, a convolutional neural network (CNN) and a deep recurrent long short-term memory network (LSTM) with regressive outputs were used [2].

The novelty of the presented solution is the adaptation of the deep neural network structure, which as a rule is designed for image classification problems to solve the regression problem in which the input is not an image but a vector consisting of 96 voltage measurements. Furthermore, the layers' types, amounts, and parameters were finetuned such that the neural network was effective in learning. As a result, reconstruction images were obtained, the quality of which is adequate to identify areas of moisture inside the walls with sufficient precision.

1. Materials and methods

In order to verify the effectiveness of CNN in the problem of identifying moisture inside the walls of buildings using electrical impedance tomography (EIT) [3, 4], a set of simulation cases was generated. The data set included 40,000 measurement vectors (inputs) and pattern images (outputs). Each measurement vector consisted of 96 measurements correlated with the voltages measured between the different electrode pairs. In order to generate the training data, the Eidors toolbox was used, which cooperates with the Matlab software. Eidors works based on the finite element method [5]. A set of 16 linearly positioned electrodes was used in the tests. The scheme of the test stand is shown in Fig. 1.

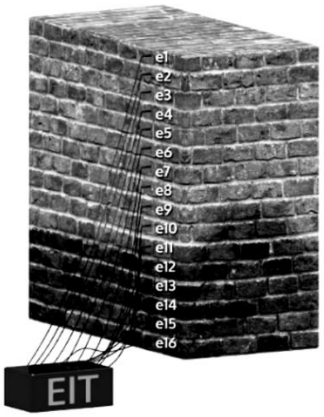


Fig. 1. Scheme of the measuring stand [6]

The forward problem is solved by determining the potential distribution within the region given the boundary conditions and complete information about the region, i.e. by solving Laplace's equation:

$$-\boldsymbol{\nabla} \cdot (\boldsymbol{\sigma}\, \boldsymbol{\nabla} \mathbf{u}) = 0 \tag{1}$$

where $\boldsymbol{\sigma}$ denotes conductivity, symbol $\mathbf{u}$ represents electrical potential. The optimized fitness function is constructed in the following way:

$$F_{fit} = 0.5 \sum_{i=1}^{n} (\boldsymbol{U} - \boldsymbol{U}_m)^T (\boldsymbol{U} - \boldsymbol{U}_m) \tag{2}$$

where n is a projection angle (the number of measurement sequences: 12 measurements × 8 angle projections = 96 independent measurements), $\boldsymbol{U}_m$ – the measured voltage, $\boldsymbol{U}$ – the calculated voltage by solving the equation (1). Fig. 2 shows the way of transforming measurements into tomographic images.

On the surface of the damp brick wall, there are 16 electrodes arranged vertically in a straight line. The electrodes were connected to the EIT tomograph, which manages the measurement

process [5, 7]. Built-in multiplexer determines the frequency of measurements and the sequence of changes of electrode pairs between which the measurement is made. The recorded measurements are then converted into images using CNN and LSTM. The resolution of the spatial image mesh is 6215 finite elements (pixels).

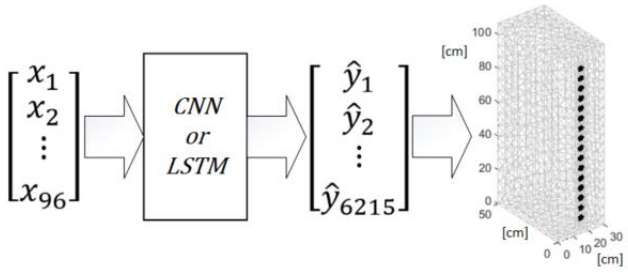


Fig. 2. The method of transforming measurements into tomographic images

In order to better compare the usefulness of deep learning methods in electrical tomography, two types of networks were trained: CNN and LSTM. The structure of the layers of the CNN network is shown in Fig. 3. All used in the CNN fully connected layers model and the regression output layer have the same number of outputs, amounting to 6215.

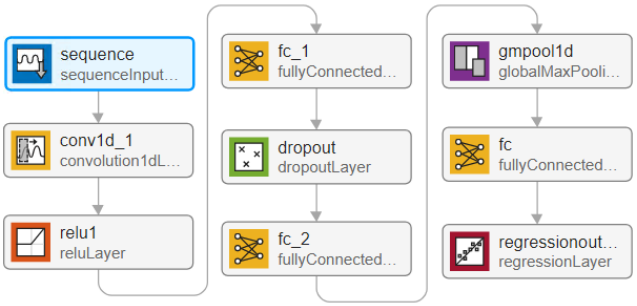


Fig. 3. Convolutional neural network (CNN) layers

The CNN network consists of 9 layers. The first layer has a single sequence structure. It is a vector of 96 measurements. The second layer is a 1-D convolutional layer that applies sliding convolutional filters to 1-D input. It contains 192 filters of size 4. Next is the Rectified Linear Unit (ReLU) layer. The next layer is a fully connected layer that precedes the dropout layer with a probability of 0.3. The sixth layer is connected again, followed by the global max-pooling layer. A fully connected layer multiplies the input by a weight matrix and adds a bias vector. The eighth layer is the third fully connected layer that precedes the regression output layer that computes the half-mean-squared-error loss for regression tasks.

The LSTM model had only 4 layers as shown in Fig. 4. First, the sequence input layer injects sequence data into the network.

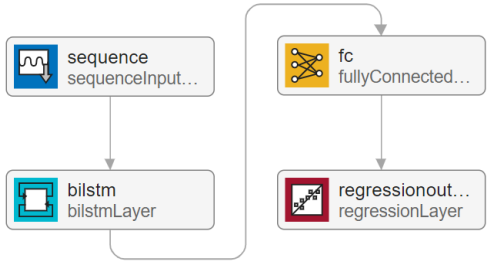


Fig. 4. LSTM network layers

A bidirectional LSTM (BiLSTM) layer learns long-term bidirectional dependencies between time steps of time series or sequence data. These dependencies are advantageous when the network is required to learn from the entire time series at each time step. The last two layers are the fully connected and regression output layers.

Fig. 5 shows the CNN learning process based on Root Mean Square Error (RMSE). RMSE is calculated according to formula (3)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}}, \tag{3}$$

where N is the number of responses, y_i is the target output, and $\hat{y}_i$ is the model's prediction for response i. In addition, Fig. 6 shows the same CNN learning flow but based on the *Loss* value. The *Loss* is calculated as (4)

$$Loss = \frac{1}{2S}\sum_{i=1}^{S}\sum_{j=1}^{N}\left(y_{ij} - \hat{y}_{ij}\right)^2, \tag{4}$$

where S is the sequence length. In our case $S = 96$.

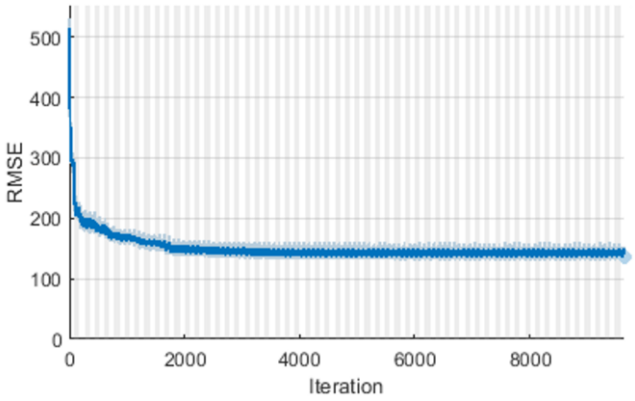


Fig. 5. Training progress of the CNN through the RMSE indicator

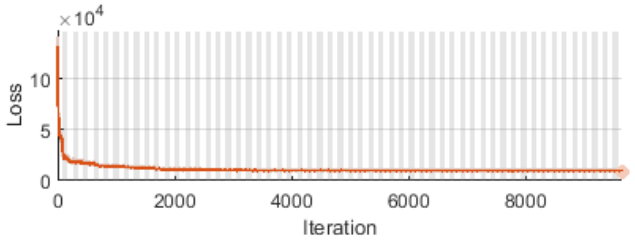


Fig. 6. Training progress of the CNN through the Loss indicator

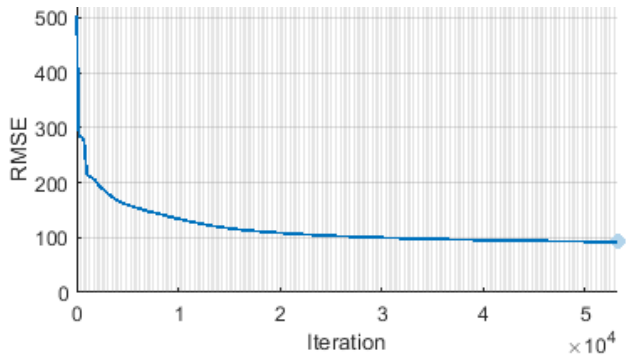


Fig. 7. Training progress of the LSTM through the RMSE indicator

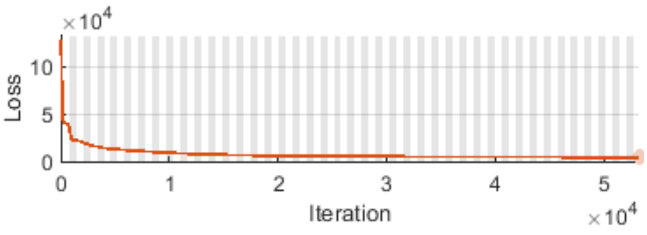


Fig. 8. Training progress of the LSTM through the Loss indicator

As can be seen from the figures above, CNN RMSE is around 145 and Loss around 10^4. In Figs. 7 and 8 show analogous indicators of the learning quality of the LSTM network. RMSE and Loss here are respectively 95 and $5{\cdot}10^3$. It proves the advantage of

LSTM over CNN. The RMSE and Loss values result from the assumptions made, according to which the background value (dry areas) of the tomographic image is 1, and the moist areas are 10.

2. Results and discussion

Fig. 9 shows the results obtained using CNN and LSTM methods for selected 3 cases. The first case (Fig. 9 a-c) shows the moisture located in the rear part of the examined area. The LSTM reconstructions seem to be more precise because CNN shows the moisture area larger than in the reference image.

The second case (Fig. 9 d-f) shows the moisture located in the front part of the examined area. Although the LSTM reconstructions seem to be better again, CNN shows areas with different moisture levels also in other parts of the section of the wall, which makes it difficult to identify wet areas correctly.

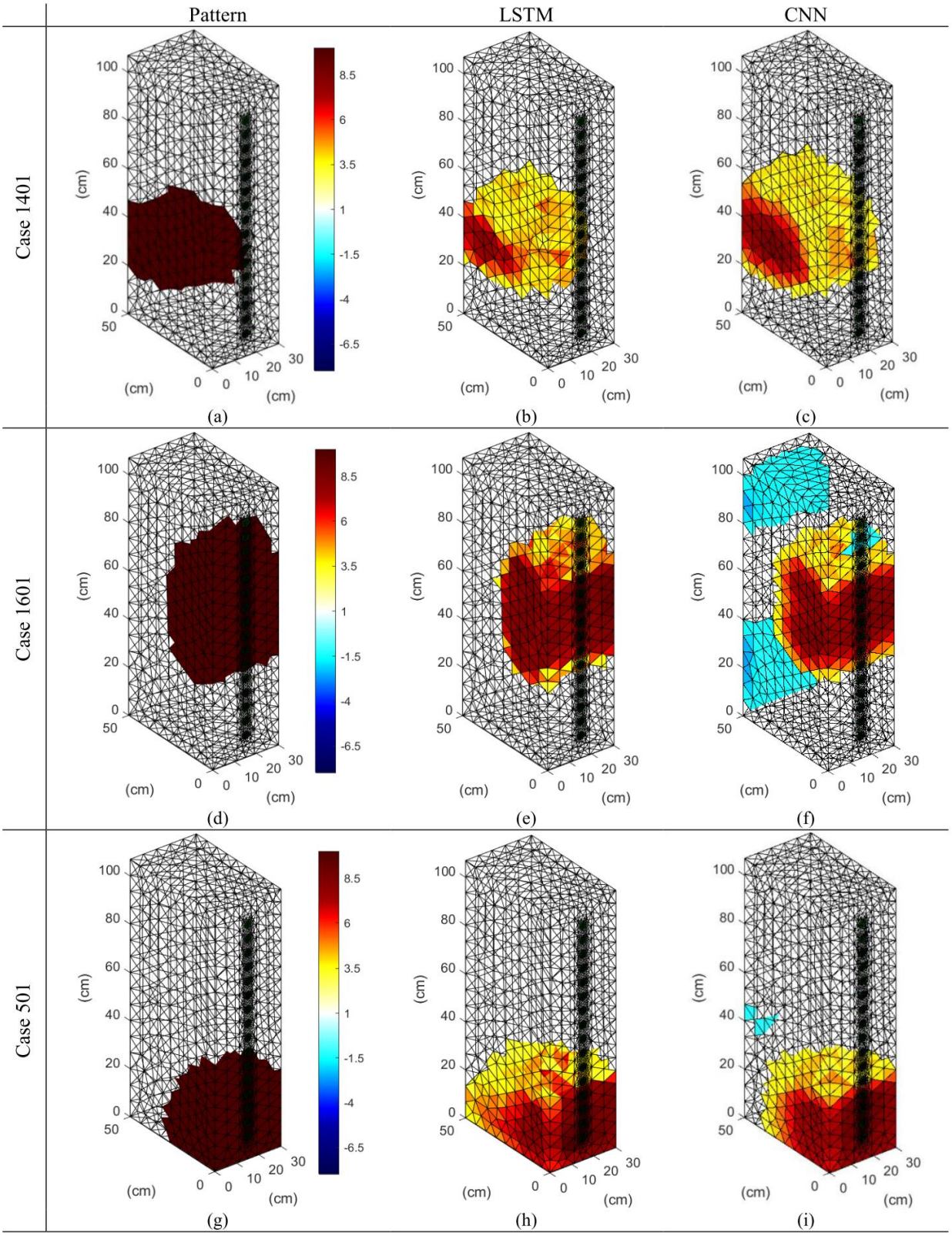


Fig. 9. Comparison of selected reconstructions. Images (a,d,g) are the patterns. Images (b,e,h) were obtained with LSTM algorithm. Images (c,f,i) were created with the CNN algorithm

The third case (Fig. 9 g-i) shows the moisture located in the front lower part of the examined area. Of all the presented cases, this one seems to be the closest for both methods (CNN and LSTM). The shape and contour of the moisture better reflect the CNN, but at the same time, there are minor disturbances in the rear part of the tested section, approximately 40 cm above the ground level.

Four widely used metrics were used to evaluate the quality of tomographic reconstructions objectively: root mean square error (RMSE), normalized mean square error (NMSE), relative image error (RIE), and image correlation coefficient (ICC). The root mean square error is calculated according to the previously presented formula (3). NMSE is calculated by (5)

$$\mathrm{NMSE} = \frac{\|\boldsymbol{y}-\hat{\boldsymbol{y}}\|^2}{\|\boldsymbol{y}-\bar{\boldsymbol{y}}\|^2}, \tag{5}$$

where $\boldsymbol{y}$ is the reference (ground-truth) conductivity distribution, $\bar{\boldsymbol{y}}$ is the average reference ground-truth conductivity distribution, $\hat{\boldsymbol{y}}$ denotes the reconstructed conductivity distribution, and $\|\cdot\|$ is the L2–norm set [8,9]. RIE is calculated according to formula (6)

$$\mathrm{RIE} = \frac{\|\boldsymbol{y}-\hat{\boldsymbol{y}}\|}{\|\boldsymbol{y}\|}, \tag{6}$$

and ICC is described by equation (7)

$$\mathrm{ICC} = \frac{\sum_{i=1}^{n}(y_i-\bar{y})(\hat{y}_i-\bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2\sum_{i=1}^{n}(\hat{y}_i-\bar{\hat{y}})^2}}, \tag{7}$$

where $\bar{\hat{\boldsymbol{y}}}$ is the mean reconstruction conductivity distribution. The lower the RMSE, NMSE, and RIE, and the greater the ICC, the higher the tomographic image quality. ICC = 1 indicates ideal reconstruction, whereas ICC = 0 signifies the worst one.

The Table 1 compares the LSTM and CNN methods. Four criteria defined as indicators were used for this purpose: RMSE, NMSE, RIE, and ICC.

Table 1. Reconstruction quality indicators for the three cases compared

Methods of reconstruction	Indicator	Reconstruction cases		
		Case 1	Case 2	Case 3
CNN	RMSE	1.572	1.781	1.323
	NMSE	0.379	0.244	0.158
	RIE	0.587	0.325	0.275
	ICC	0.722	0.900	0.935
LSTM	RMSE	1.159	1.027	1.261
	NMSE	0.206	0.081	0.143
	RIE	0.433	0.187	0.262
	ICC	0.849	0.968	0.941
Winning method:		LSTM	LSTM	LSTM
		LSTM	LSTM	LSTM
		LSTM	LSTM	LSTM
		LSTM	LSTM	LSTM

Three distinct moisture content cases were evaluated. The table's final four lines contain information about the method that produces the best result for a given case when each measure is considered. As can be seen, the LSTM method produced superior results in all cases evaluated. It should be stated objectively that some of the differences are quite minor. For instance, the difference is very small when comparing the ICC values for Case 3. The delta (difference) ICC is only 0.006. When comparing the two methods, one additional consideration should be made: the reconstruction time. The model's ability to generate images quickly is critical, even more so when performing measurements during dynamic industrial processes. The time required for the LSTM reconstruction was 0.006390 seconds, while the time required for the CNN reconstruction was 0.035113 seconds. In practice, the difference favours the LSTM more than 5 times.

3. Conclusions

This article introduces an innovative algorithmic concept for solving the static problem of tomographic image reconstruction using a recurrent deep LSTM network and convolutional neural network. Electrical impedance tomography was used to image moisture within a brick wall. The LSTM and CNN networks were successfully trained by treating the measurement vector as a single time step sequence signal. The reconstruction's high quality was confirmed by comparing it to images generated using another high-efficiency method, LSTM. Both methods described here allow for spatial visualization of the moisture distribution within a wall. It is significantly different from traditional indirect methods, which only test the humidity at selected wall points. In addition, the LSTM method has a significant advantage in terms of reconstruction speed, which opens up new application possibilities, particularly in the area of automated, dynamic industrial processes. Future research will focus on deciphering the moisture expansion processes occurring within porous materials. To fully exploit the LSTM and CNN networks' potential, more sophisticated modifications to the input vector and data preprocessing are planned.

References

[1] Fabijańska A., Banasiak R.: Graph Convolutional Networks for Enhanced Resolution 3D Electrical Capacitance Tomography Image Reconstruction. Applied Soft Computing 110, 2021, 107608, [http://doi.org/10.1016/J.ASOC.2021.107608].

[2] Hola A.: Measuring of the Moisture Content in Brick Walls of Historical Buildings-the Overview of Methods. IOP Conference Series: Materials Science and Engineering 251(1), 2017, [http://doi.org/10.1088/1757-899X/251/1/012067].

[3] Kłosowski G. et al.: Quality Assessment of the Neural Algorithms on the Example of EIT-UST Hybrid Tomography. Sensors 20(11), 2020, [http://doi.org/10.3390/s20113324].

[4] Kłosowski G. et al.: The Concept of Using Lstm to Detect Moisture in Brick Walls by Means of Electrical Impedance Tomography. Energies 14(22), 2021, [http://doi.org/10.3390/en14227617].

[5] Litti G. et al.: Hygrothermal Performance Evaluation of Traditional Brick Masonry in Historic Buildings. Energy and Buildings 105, 2015, 393–411, [http://doi.org/10.1016/j.enbuild.2015.07.049].

[6] Porzuczek J.: Assessment of the Spatial Distribution of Moisture Content in Granular Material Using Electrical Impedance Tomography. Sensors 19(12), 2019, 2807, [http://doi.org/10.3390/s19122807].

[7] Romanowski A. et al.: X-Ray Imaging Analysis of Silo Flow Parameters Based on Trace Particles Using Targeted Crowdsourcing. Sensors 19(15), 2019, 3317, [http://doi.org/10.3390/s19153317].

[8] Rymarczyk T. et al.: Area Monitoring Using the ERT Method with Multisensor Electrodes. Przegląd Elektrotechniczny 95(1), 2019, [http://doi.org/10.15199/48.2019.01.39].

[9] Rymarczyk T., Adamkiewicz P.: Nondestructive Method to Determine Moisture Area in Historical Building. Informatics Control Measurement in Economy and Environment Protection 7(1), 2017, [http://doi.org/10.5604/01.3001.0010.4586].

Ph.D. Eng. Grzegorz Kłosowski
e-mail: g.klosowski@pollub.pl

Assistant Professor in Department of Organization of Enterprise at the Faculty of Management of Lublin University of Technology. The author's research interests include artificial intelligence, simulation, and modelling of engineering and business processes – leader and participant in several implementation projects.

http://orcid.org/0000-0001-7927-3674

D.Sc. Tomasz Rymarczyk
e-mail: tomasz.rymarczyk@netrix.com.pl

Director in Research and Development Center Netrix S.A. His research area focuses on applying non-invasive imaging techniques, electrical tomography, image reconstruction, numerical modelling, image processing and analysis, process tomography, software engineering, knowledge engineering, artificial intelligence and computer measurement systems.

http://orcid.org/0000-0002-3524-9151

http://doi.org/10.35784/iapgos.2882 *received: 22.02.2022 | revised: 11.03.2022 | accepted: 28.03.2022 | available online: 31.03.2022*

TILTED FIBER BRAGG GRATING SENSORS FOR REFRACTIVE INDEX MEASUREMENTS OF LIQUID SOLUTIONS

Damian Harasim
Lublin University of Technology, Faculty of Electrical Engineering and Computer Science, Department of Electronics and Information Technology, Lublin, Poland

Abstract. *This publication presents the basic assumptions about the light guiding mechanisms in tilted fiber Bragg gratings, transmission spectra of the fiber with such structures and their sensor properties related to the occurrence of the so-called cladding modes. The light in the form of modes coupled to the optical fiber cladding causes their spectral properties to depend on the difference in the refractive index of the surrounding medium and the fiber cladding. With the introduction of a small inclination angle of the diffraction fringes forming the Bragg grating, the cladding modes show sensitivity to changes in the refractive index of the environment of aqueous solutions, which makes the spectrum of their applications broad. This publication presents changes in the spectra of selected modes measured for solutions with a specific refractive index, shift sensitivity, and changes in the mode transmission coefficient and processing characteristics at a selected concentration range of solutions. Experimental results show that high order cladding modes respond both by shifting the central wavelength as well as by changing the transmission minimum. In selected ranges, these parameters show a linear characteristic as a function of changes in the concentration of the cane sugar aqueous solution. In the case of TFBG with a tilt angle of 8°, the sensitivity of wavelength changes is 0.012 nm/RIU for solutions with concentrations ranging from 0% to 10% by weight.*

Keywords: fiber Bragg gratings, tilted fiber Bragg gratings, refractive index measurement, surrounding refractive index

ŚWIATŁOWODOWE SKOŚNE SIATKI BRAGGA JAKO CZUJNIKI W POMIARACH WSPÓŁCZYNNIKA ZAŁAMANIA CIECZY

Streszczenie. *Niniejsza publikacja prezentuje podstawowe założenia dotyczące mechanizmów prowadzenia światła w światłowodowych skośnych siatkach Bragga (ang. tilted fiber Bragg grating), widma transmisyjne światłowodu z wytworzonymi takimi strukturami oraz ich właściwości czujnikowe związane z występowaniem tzw. modów płaszczowych. Prowadzenie światła w postaci modów sprzęganych do płaszcza światłowodu sprawia, że ich właściwości spektralne zależne są od różnicy współczynników załamania ośrodka otaczającego oraz płaszcza włókna. Przy wprowadzeniu niewielkiego kąta pochylenia prążków dyfrakcyjnych tworzących siatkę Bragga, mody płaszczowe wykazują wrażliwość na zmiany współczynnika załamania otoczenia roztworów wodnych, przez co spektrum ich zastosowań jest szerokie. W niniejszej publikacji przedstawiono zmiany widm wybranych modów mierzonych dla roztworów o określonym współczynniku załamania, czułości przesunięcia oraz zmiany współczynnika transmisji modu oraz charakterystyki przetwarzania przy wybranym zakresie stężenia roztworów. Wyniki eksperymentalne wskazują, że mody płaszczowe wysokich rzędów reagują zarówno przesunięciem centralnej długości fali jak również zmianą minimum transmisji. Parametry te w wybranych zakresach wykazują charakterystykę liniową w funkcji zmian stężenia roztworu wodnego cukru trzcinowego. W przypadku TFBG o kącie pochylenia 8°, czułość zmian długości fali wynosi 0,012 nm/RIU dla roztworów o stężeniach od 0% do 10% stężenia wagowego.*

Słowa kluczowe: światłowodowe siatki Bragga, skośne siatki Bragga, pomiar współczynnika załamania, współczynnik załamania ośrodka

Introduction

Optical fiber technology was developed as an efficient way for transmission of large amount of data. On the other hand, optical fiber properties have become an object of interest for applications as physical quantities sensors [1, 4, 8]. Development of a novel areas in industry and a challenges of maintaining the quality of existing structures creates a new approaches for sensing applications. According to the advantages such as extremely small size, electromagnetic interference immunity and capability for the possibility of embedding them inside the measured structure, Fiber Bragg Gratings (FBGs) has emerged as a promising transducers for sensing applications [12–14]. In case of sensor systems based on FBGs important advantages are also the independence of measurement accuracy from fluctuations in the light source and the ability to build more complex measurement systems by placing many sensors on a single fiber optic fiber. The basic idea of this structures is to create a periodic zones of increased refractive index in single mode fiber core. This kind of structures found their application in temperature, strain etc. measurements. Consecutive fringes of grating perturbations cause a the ability to reflect light radiation of a strictly defined wavelength λ_B, while being transparent to light of other wavelengths. Bragg wavelength is defined by:

$$\lambda_B = 2n_{eff} \cdot \Lambda \tag{1}$$

where n_{eff} is the effective refractive index of a fiber core and Λ is a grating period. By basing the measurement on determining the relative shift of the central wavelength, fluctuations in the optical power of the light source do not affect its accuracy. There are many methods of determining the Bragg wavelength, also allowing for its determination on the basis of spectra containing strong noise [9]. Linear conversion of the measured quantity to the shift of the characteristic wavelength called the central Bragg wavelength makes them natural converters of physical quantities.

Two basic technical parameters which determines optical properties of manufactured structures are apodization and chirp. An important parameter determining the usefulness of a given periodic structure is minimizing the side lobes. Apodization is one way to achieve this effect by changing the modulation depth of the refractive index changes in the fiber core along its axis. The multitude of special applications of fiber-optic periodic structures exact the implementation of apodization functions into the techniques of their production. The profile of apodization is commonly given by the power distribution of inscribing laser beam. Usually the profile is defined as Gaussian function [6]:

$$f(x) = a \cdot \exp\left(-\left(\frac{x-b}{c}\right)^2\right) \tag{2}$$

The conventional gratings are created with a constant distance between following zones with internal refractive index perturbations. Another possibility for adjusting a spectral properties for particular application is creating a chirp by differentiation of distances between consecutive refractive index perturbations [3]. Most commonly used is linear chirp which allows to extend the width of reflected light spectrum by creating grating which operates as a series of short structures with increasing Bragg wavelength.

Another group of periodic structures are Tilted Fiber Bragg Gratings (TFBGs) which are manufactured by creating a tilt angle Θ_{TFBG} between refractive index fringes and fiber cross-section plane [7, 11]. Figure 1 shows schematically internal structure of tilted increased refractive index zones pattern and a coupling of backward propagating core mode and cladding modes guided by outer optical fiber layer.

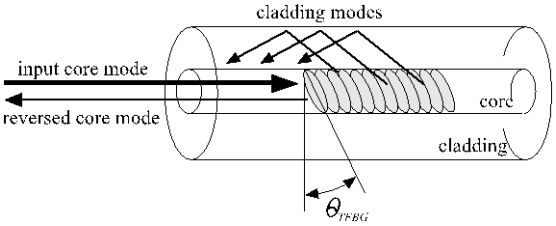


Fig. 1. Scheme of internal structure of TFBG grating and coupling mechanism of core-cladding resonances

1. Principles of TFBG operation and RI sensing

Fiber modification presented in Introduction cause a coupling of backward propagating modes in cladding of fiber. Occurrence of cladding modes reveals as series of dips in transmission spectrum in wavelengths shorter than Bragg wavelength. The resonant wavelengths of the cladding modes, corresponding to the backwards-propagating coupling between the core mode and the cladding modes, are given by:

$$\lambda_{clad\_i} = \left(n_{eff\_clad} + n_{eff}\right)\frac{\Lambda}{\cos(\Theta_{TFBG})} \tag{3}$$

where n_{eff} is the effective refractive index of a core, $n_{eff—clad}$ is the effective refractive index of the i-th cladding mode, $i = 1…m$ where m is the total number of cladding modes. Figure 2 shows schematically transmission spectra of a) conventional FBG and b) 2° tilt TFBG. Temperature sensitivity of this kind of structures equals about 10 pm/°C and is similar to traditional FBGs sensitivity. Measurement of bending and surrounding refractive index (SRI) requires more advanced method of transmission spectra analysis.

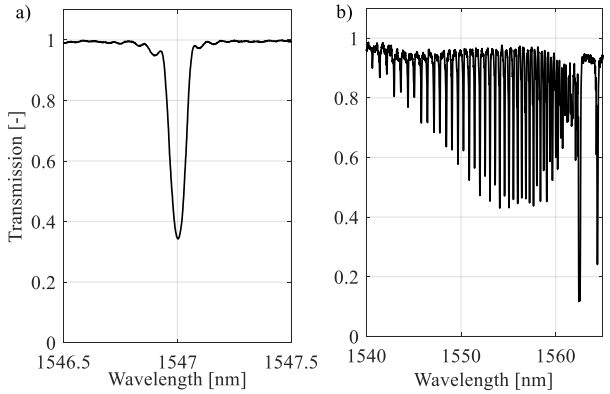


Fig. 2. Transmission spectra of: a) conventional FBG with Gaussian apodization, b) TBFG with 2° tilt angle and the same apodization profile

The propagation of light in the close proximity of the outer border of the optical fiber makes the spectral parameters of the minima observed in transmission dependent on the refractive index of the medium outside the optical fiber. Tilt angle has strong influence on grating possibilities and conditions the optical properties of the spectrum and ddetermines which modes will be visible in spectrum and their intensity. Transmission spectra of 2°, 4° and 6° tilt structures are presented on figure 3. Tilting of refractive index pattern and cross-section plane of fiber cause a widening of wavelength range of cladding-coupled modes. The dips with wavelengths closest to the Bragg resonance are called low-order modes and are guided in close proximity to fiber core. The cladding modes with shortest wavelengths are high-order modes. It is visible that increasing of non-parallelism between High order modes are stronger coupled when tilt angle is greater and are guided closer to the external surface of the optical fiber. This makes them sensitive to changes in the refractive index of the medium surrounding the optical fiber with TFBG written in the core [2, 10]. Together with increasing of surrounding refractive index value, particular cladding modes disappears – the coupling is weakening.

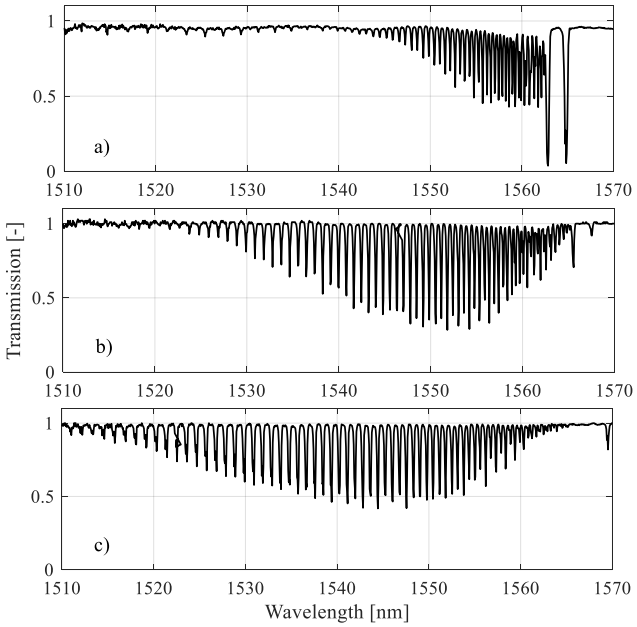


Fig. 3. Transmission spectra presented for three TFBG structures with a): 2° tilt, b) 4° tilt and c) 6° tilt

Spectrum is getting smoother, amplitude of dips related with individual modes is reduced, and then the peaks disappears. Increasing SRI reach the value of effective refractive index of specified cladding mode which results with extraction of this mode outside fiber structure. At further decreasing wavelengths there comes a point which the grating couples light to modes that are no longer guided by the cladding. These modes have resonance positions that do not shift in wavelength in response to SRI changes, but only in amplitude. Therefore, cut-off cladding mode could be use as indicator of surrounding refractive index changes. In most applications, refractive index of examined medium has value close to a water which is 1.3333.

2. TFBG transmission spectra under various refractive index of surrounding medium

The possible application with expected range of possible refractive index is determines the tilt angle of manufactured TFBG transducer. Solutions with RI close to the refractive index of fiber cladding requires week tilted structures which leads to coupling of small amount of low order modes. When examined solution has lower value of forseen RI, the tilt angle have to be greater to ensure coupling of higher order modes guided for shorter wavelengths. For the purposes of this article 10 solutions of glucose in water were prepared. Table 1 presents a refractive indexes of water solutions with 0% to 10% glucose concentrations.

Table 1. Refractive index of proposed glucose solutions with different concentrations

No	Concentration	RI
1	0%	1.3333
2	1%	1.3344
3	2%	1..3359
4	3%	1.3374
5	4%	1.3388
6	5%	1.3403
7	6%	1.3418
8	7%	1.3433
9	8%	1.3448
10	9%	1.3463
11	10%	1.3478

According to refractive indexes of prepared solutions, the chosen tilt angle of internal TFBG structure was 8°. The grating was inscribed on Ge-doped single mode photosensitive fiber by using excimer UV laser and phase mask technique. Transmission spectra were measured by using Optical Spectrum Analyzer (OSA). TFBG was illuminated by broadband light source which allows

to observe a slight changes of fiber transducer optical properties. Figure 4. presents a scheme of equipped system for measuring transmission characteristics of 8° TFBG immersed in water solutions with differ refractive indexes according to table 1.

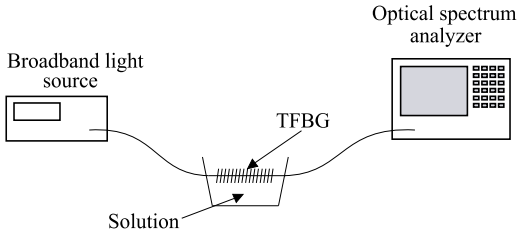


Fig. 4. Scheme of measurement system for characterization transmission spectra of TFBG immersed in solutions with different RI

Changes in transmission spectrum could be divided into two types: global and local. Global changes are related with the variations of shape of spectrum measured in wide range of wavelengths containing number of cladding modes. Analysis of many modes could give an improvement of accuracy or extension of possible refractive index range. However it is related with a need for measurement of spectrum which requires a specialized equipment – Optical Spectrum Analyzer (OSA). Methods which are based on analysis of selected cladding mode are so-called local methods. Surveys about the properties of cladding modes should be conducted with using OSA to obtain a detailed sensitivities of selected resonance. Properties and sensitivities of one selected cladding mode designated with OSA could be also used with measurement systems build by optical filters. In this kind of systems, expensive laboratory equipment could be replaced by passive optical filters (usually conventional FBGs) and photodiode with electrical voltage converter. On figure 5 are presented transmission spectra of four cladding modes in two ranges of wavelengths: a) 1536–1541 nm and b) 1546–1551 nm. As it could be seen, changes of surrounding medium refractive index induces different changes in spectral parameters of following cladding modes. Inset a presents modes with longer wavelengths which are presenting strong sensitivity of transmission coefficient to SRI. Cladding resonances presented on inset b) of Fig. 5. stands out with stability of optical power amplitude and shift of center wavelengths.

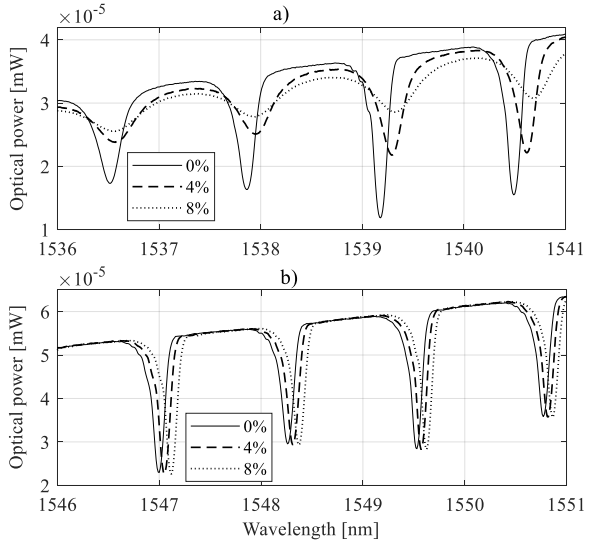


Fig. 5. Transmission spectra of 8° angle TFBG measured for grating immersed in 0%, 4% and 8% sugar solutions presented for two ranges of cladding modes: a) 1536–1541 nm and b) 1546–1551 nm

Figure 6. presents a detailed view of selected cladding modes of TFBG with 8° internal angle, which have greatest sensitivity of spectral parameters related to changes in sugar concentration (from 0 to 10%) in water solution. Figure 6 presents higher order mode which shows changes of transmission coefficient.

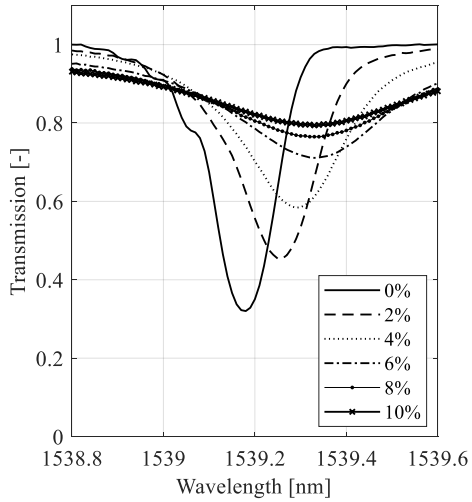


Fig. 6. Transmission spectra of selected cladding mode, when TFBG immersed in solutions with different refractive index

Transmission spectra presented on figure 6. were used to determinate a characteristic of shorter wavelength mode transmission to changes of cane sugar concentration in water solution, which is presented in figure 7.

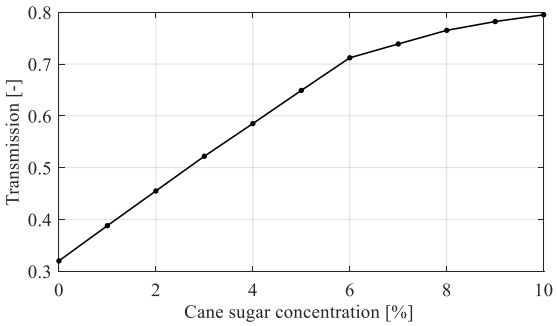


Fig. 7. Scheme of measurement system for characterization transmission spectra of TFBG immersed in solutions with different RI

As it is visible, changes of refractive index of surrounding medium have a quasi-linear influence on transmission. In 0–6% range, sensitivity is greater, and when the refractive index of surrounding medium is getting close to the effective index of selected cladding mode, transmission is changing weaker. It have to be noted, that the angle of internal tilt should be adjusted to assumed possible values of SRI according to different sensitivities of following resonances.

3. Surrounding refractive index measurements with designed system

As it was mentioned before, analysis of selected cladding mode spectral parameters allows to create measurement system which are not using optical spectrum analyzer. Results presented in this chapter are related with measurement setup which is using FBG as passive optical filer for determination of common part of selected cladding mode and filter spectrum. Configuration with optical bandpass filter most often requires use of optical circulator. High selectivity of signals between channels, provides ability for measurements of signal reflected from FBG filter. Central wavelength of filtering grating have to be adjusted to wavelength of TFBG cladding mode. Changes of transmission coefficient affects the shape of reflected light spectrum which is also detectable a change of the intensity. For measurement of spectral characteristics it is necessary to provide broadband source of input light. In presented measurements was used SLED Thorlabs S5FC1550P-A2 source. Figure 8. presents reflection spectrum of FBG filter measured for sensor head immersed in water solutions with different concentrations of cane sugar.

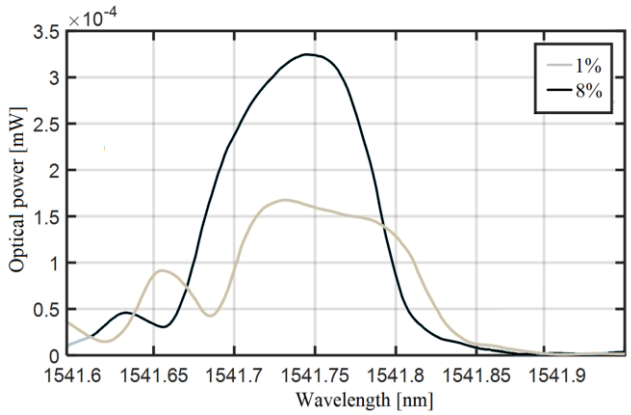


Fig. 8. Spectra of light reflected by FBG filter with TFBG sensor immersed in water solutions with different values of cane sugar concentrations

Analysis of spectra presented in figure 8 shows, that the area under a spectrum is getting bigger while a concentration of sugar is greater. This is related with changing of transmission of cladding mode of TFBG which has wavelength matched to filtering FBG. When sugar concentration is higher, transmission of cladding resonance sensitive to SRI is bigger so more of the light reflected from filter is transferred to analyzer. This property makes the measurement setup able, to use photodetector instead of OSA. When the area under the spectrum is greater, it means that the integral value will be greater could be detected as a intensity of radiation. Optical spectrum analyzer was replaced by Thorlabs- PDA30B-EC detector, which gives output signal as a value of direct voltage. Figure 9 presents a value of output voltage which was measured, for TFBG immersed in water solutions with increasing concentration of sugar.

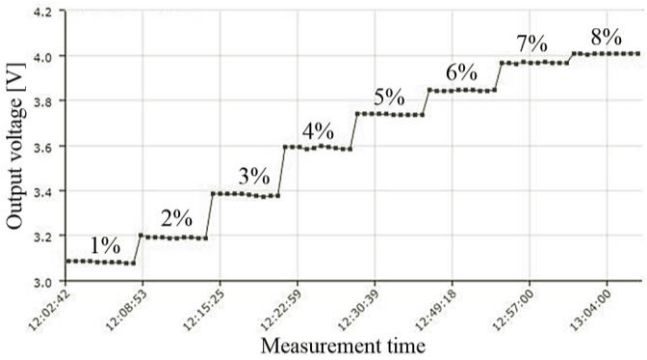


Fig. 9. Output voltage of measurement setup measured for different refractive index of surrounding medium

For each concentration, value of voltage was measured ten times at intervals of 10 seconds. It is visible, that voltage has constant value when SRI is not changing. Transformation of spectral changes into voltage value, makes results much easier to process than in case of spectral characteristics. It has to be noted that measurement results will be different with application of different optical components.

4. Conclusions

This paper presents an approach for using a TFBGs as a transducers of refractive index of medium surrounding optical fiber. Introducing a tilt angle between planes of grating fringes and fiber cross-section provides strong coupling of resonances guided by cladding. According to this property, spectral characteristics of cladding modes are related to refractive index of surrounding medium. Creating of measurement setup with FBG works as a reflective filter, allows to measure SRI in range related with few-percentage changes of cane sugar concentration in water solution. Application of photodetector for light intensity measurement allows to get output signal in form of votage, which is easy to process for example in microprocessor drivers. Exemplary results presented in this paper obtained for usage of TFBG with 8° tilt of internal structure, shows that changes of slight changes of concentration could be measured.

This work was supported by Project RPLU.01.02.00-06-0020/16

References

[1] Bakaic M., Hanna M., Hnatovsky C., Grobnic D., Mihailov S., Zeisler S., Hoehr C.: Fiber-Optic Bragg Gratings for Temperature and Pressure Measurements in Isotope Production Targets for Nuclear Medicine. Applied Sciences 10, 2020, 4610 [http://doi.org/10.3390/app10134610].
[2] Caucheteur C., Mégret P.: Demodulation technique for weakly tilted fiber Bragg grating refractometer. IEEE Photonics Technology Letters 17(12), 2005, 2703–2705 [http://doi.org/10.1109/LPT.2005.859411].
[3] Chettouh S., El-Akrmi A., Triki H., Hamaizi Y.: Spectral properties of nonlinearly chirped fiber Bragg gratings for optical communications. Optik 147, 2017, 163–169 [http://doi.org/10.1016/j.ijleo.2017.08.08].
[4] Erdogan T.: Fiber grating spectra. Journal of Lightwave Technology 15, 1997, 1277–1294 [http://doi.org/10.1109/50.618322].
[5] Fazzi L., Groves R. M.: Demodulation of a tilted fibre Bragg grating transmission signal using α-shape modified Delaunay triangulation. Measurement 166, 2020, 108197 [http://doi.org/10.1016/j.measurement.2020.108197].
[6] Gong J. M., Chan C. C., Jin W., MacAlpine J. M. K., Zhang M., Liao Y. B.: Enchancement of wavelength detection accuracy in fiber Bragg grating sensors by using spectrum correlation technique. Optics Communications 212, 2002, 29–33 [http://doi.org/10.1109/OFS.2002.1000525].
[7] Guo T., Liu F., Guan B., Albert J.: Tilted fiber grating mechanical and biochemical sensors. Optics & Laser Technology 78B, 2016, 19–33 [http://doi.org/10.1016/j.optlastec.2015.10.007].
[8] Guo T., Tam H. Y., Albert J.: Chirped and tilted fiber Bragg grating edge filter for in-fiber sensor interrogation. Optical Society of America Science and Innovations 2011 [http://doi.org/10.1364/CLEO_SI.2011.CThL3].
[9] Harasim D., Yussupova G.: mprovement of FBG peak wavelength demodulation using digital signal processing algorithms. Proc. SPIE 966212, 2015, 270–276 [http://doi.org/10.1117/12.2205547].
[10] Liu F., Guo T., Liu J., Zhu X., Liu Y., Guan B., Albert J.: High-sensitive and temperature-self-calibrated tilted fiber grating biological sensing probe. Chinese Sciences Bulletin 58, 2013, 2607–2611 [http://doi.org/10.1007/s11434-013-5724-3].
[11] Lu Y., Shen C., Chen D., Chu J., Wang Q., Dong X.: Highly sensitive twist sensor based on tilted fiber Bragg grating of polarization-dependent properties. Optical Fiber Technology 20(5), 2014, 491–494 [http://doi.org/10.1016/j.yofte.2014.05.011].
[12] Mohammed N. A., Ali T. A., Aly M. H.: Evaluation and performance enhancement for accurate FBG temperature sensor measurement with different apodization profiles in single and quasi-distributed DWDM systems. Optics and Lasers in Engineering 55, 2014, 22–34 [http://doi.org/10.1016/j.optlaseng.2013.10.013].
[13] Ren L., Jia Z., Li H., Song G.: Design and experimental study on FBG hoop-strain sensor in pipeline monitoring. Optical Fiber Technology 20, 2014, 15–23 [http://doi.org/10.1016/j.yofte.2013.11.004].
[14] Wydra M., Kisała P., Harasim D., Kacejko P.: Overhead Transmission Line Sag Estimation Using a Simple Optomechanical System with Chirped Fiber Bragg Gratings. Part 1: Preliminary Measurements. Sensors 18(1), 2018, 309 [http://doi.org/ 10.3390/s18010309].

M.Sc. Eng. Damian Harasim
e-mail: d.harasim@pollub.pl

Damian Harasim has received the MSc degree in mechatronics from Lublin University of Technology, Poland in 2014. He is currently working toward the Ph.D. degree in Automation, Electronics and Electrical Engineering at Faculty of Electrical Engineering and Computer Science in Lublin, Poland. From 2015 he is employed as assistant in Department of Electronics and Information Technology at Lublin Univ. of Technology. His current research interest lie in fabrication and characterization of optical sensing systems, especially based on unconventional Bragg structures.

http://orcid.org/0000-0002-9859-5879

http://doi.org/10.35784/iapgos.2894 *received: 10.03.2022 | revised: 18.03.2022 | accepted: 29.03.2022 | available online: 31.03.2022*

OPTIMISATION OF COMMERCIAL BUILDING MANAGEMENT PROCESSES USING USER BEHAVIOUR ANALYSIS SYSTEMS SUPPORTED BY COMPUTATIONAL INTELLIGENCE AND RTI

Michał Styła[1], Przemysław Adamkiewicz[1,2]
[1]Information Technology Research & Development Center (CBRTI sp. z o.o.), Lublin, Poland, [2]University of Economics and Innovation, Faculty of Transport and Information Technology, Lublin, Poland

Abstract. *The aim of the presented project was to create a comprehensive building management system equipped with a network of wireless and energy-efficient sensors that collect data about users and on their basis control final devices such as lighting, ventilation, air conditioning and heating. In the presented system, end devices can be both products offered by the market (commercial) and proprietary solutions (own). This is to allow the adaptation of commercial radio communication protocols with high integration capabilities and common occurrence. In addition, the system has been enriched with an innovative system of tracking and building navigation and access control, which are supported by a network of radio beacons and radio-tomographic imaging technology (RTI). The whole system is to be supervised by computational intelligence learned from scratch.*

Keywords: building automation, building management systems, computational intelligence, indoor radio communication, radio navigation, tomography

OPTYMALIZACJA PROCESÓW ZARZĄDZANIA BUDYNKAMI KOMERCYJNYMI Z WYKORZYSTANIEM SYSTEMÓW ANALIZY ZACHOWAŃ UŻYTKOWNIKÓW WSPOMAGANYCH INTELIGENCJĄ OBLICZENIOWĄ I RTI

Streszczenie. *Celem prezentowanego projektu było stworzenie kompleksowego systemu zarządzania budynkiem wyposażonego w sieć bezprzewodowych i energooszczędnych czujników, które zbierają dane o użytkownikach i na ich podstawie sterują urządzeniami końcowymi, takimi jak oświetlenie, wentylacja, klimatyzacja i ogrzewanie. W prezentowanym systemie urządzeniami końcowymi mogą być zarówno produkty oferowane przez rynek (komercyjne), jak i rozwiązania autorskie (własne). Ma to na celu umożliwienie adaptacji komercyjnych protokołów komunikacji radiowej o dużych możliwościach integracyjnych i powszechnym występowaniu. Dodatkowo system został wzbogacony o innowacyjny system śledzenia i nawigacji po budynkach oraz kontroli dostępu, które są wspomagane przez sieć radiolatarni oraz technologię obrazowania radiowo-tomograficznego (RTI). Nad całością systemu ma czuwać inteligencja obliczeniowa wyuczona od podstaw.*

Słowa kluczowe: automatyka budynkowa, inteligencja obliczeniowa, systemy radiowe do użytku wewnętrznego, systemy zarządzania budynkiem, nawigacja radiowa, tomografia

Introduction

The development of the field of automation in recent years has resulted in high market demands for the degree of integration of individual subsystems in commercial and residential buildings. These activities are not only aimed at in-creasing the comfort of use, but also at seeking savings and ecology. This is the reason for the increasing emphasis on the degree of integrity of individual building components. Although manufacturers of such components are trying to unify products under one established communication protocol (wired or wireless), this still does not allow for free mixing of devices and locks them into groups with a specific supplier.

The idea of the system, realised on a real building and then described in this article, is to integrate all the comfort modules of a typical office building: light, heating, ventilation and air conditioning in such a way as to simultaneously meet the requirements and then independently adjust the environmental parameters of individual rooms according to the habits of users detected and located by the building. This also involves linking the previously mentioned actuators to a network of sensors, and then collecting the results in a storage medium that can be accessed by a service and computational intelligence. Combined with security aspects such as access control or time recording for people and devices, the number of devices working on a centralized control system will increase exponentially.

1. Structure of building subsystems

The system developed consists of many subsystems of different types. To help understand the complexity of the resulting network, Figure 1 has been used.

The core of the system is the Home Assistant service and the Apache Kafka data bus located on the main server. The main server in turn is connected to the clients. The Home Assistant service is supported by a number of additional programmes and plug-ins, such as a module for data collection and archiving or a statistical module, which presents building parameters in a user-friendly graphic form [9].

Behind the client platforms, there are PAN (Personal Area Networks) designed to maintain communication with individual building subsystems. These networks consist of such wireless communication technologies as Bluetooth, Wi-Fi, ZigBee or Z-Wave. The use of all of them is obviously not obligatory and depends on the solutions selected for installation and those found in the building. All subsystems that have been used in the building can be traced in figure 1. Their presence depends on the needs and does not have to include all items, however, it should be remembered that some cannot function without the others, e.g. the use of comfort module control without the installation of sensor networks will prevent learning and effective operation of computational intelligence [5].

The computational intelligence (algorithms) are located together with the Home Assistant service on the customer's server. For the computational intelligence being created, the learning system uses three algorithms: linear regression, logistic regression and support vector machine (SVC). Based on the tests carried out (during normal working days, which are the source of data for learning), the following links were selected: linear regression with air conditioning, logistic regression with ventilation and heating, and light with SVC [8, 13, 15].

Computational intelligence algorithms are designed to retrieve the settings of individual devices belonging to the actuator group and data from devices recording environmental parameters. This concerns not only the conditions inside the building but also the weather outside. Taking such data into account is possible by means of a network plug-in downloading data from a local meteorological station.

The presented mechanism of machine learning as a whole is supposed to lead to elimination of cases of wasting electrical energy and life of operating elements by carelessness of a human factor, while maintaining operating parameters of individual parts of the building in accordance with the requirements of users. In the long run, it also has a positive effect on the rate of wear and tear and service costs of building equipment.

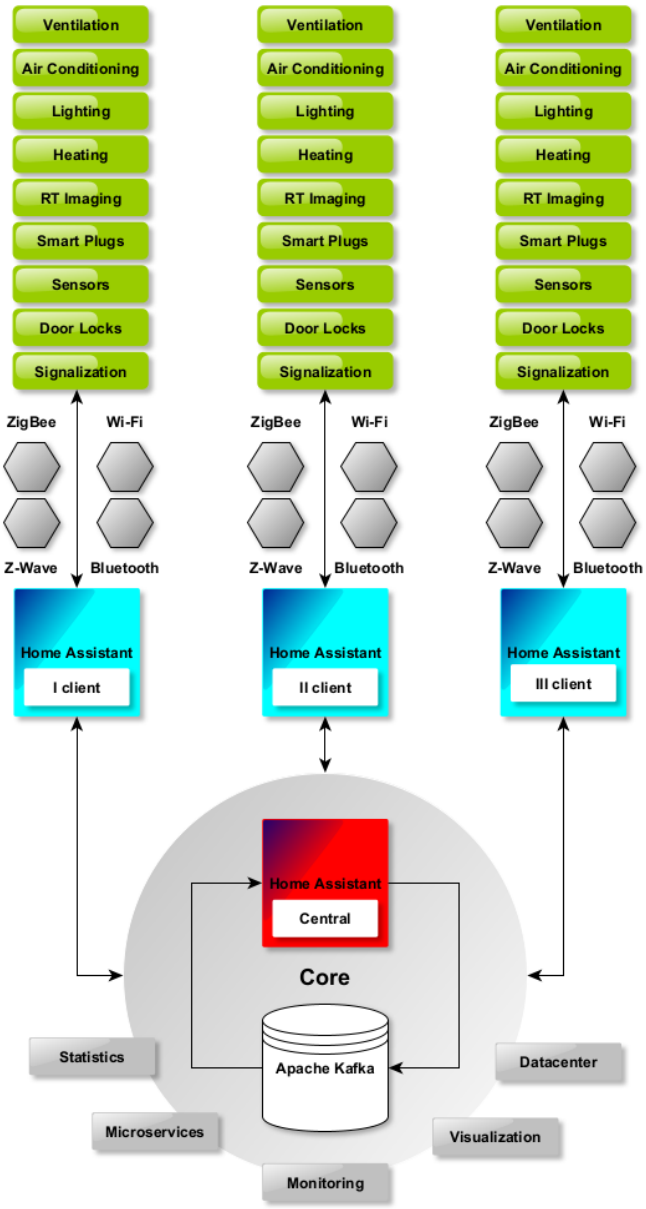


Fig. 1. General structure of created building system management

2. Comfort management systems

The comfort modules cover all the energy consumers that provide the user with the right level of light intensity, air temperature and air circulation. Each of these characteristics is reciprocated in a separate device adapted to its operation.

2.1. Comfort management – hardware

The comfort modules designed for the project fall into two categories: devices with actuator and controller in one housing, and devices with actuator and controller mounted in separate housings.

An example of an actuator and controller in one is the air conditioning controller. It is shown in figure 2 and consists of a PCB (Printed Circuit Board) equipped with a microcontroller for Bluetooth version 5 and ZigBee wireless communication, as well as necessary peripheral elements such as quartz resonators, antenna path, signal filters, communication ports and other minor RLC elements. Due to the fact that it is common to control air conditioning using an infrared beam, one of the PWM ports of the microcontroller was equipped with a high-power IR emitter. Based on the code table of the air conditioner, a sequence of instructions was created to enable automatic and manual control from the Home Assistant console. In order to make it possible to connect the widest possible range of devices to the service, the PCB was designed in such a way that emitters of different wavelengths and different powers could be used (even several at the same time). A schematic of the IrDA port-controlled actuator and controller is shown in figure 3.

The remaining hardware layer serving ventilation, heating and light was divided into two parts. First, the parts whose design is the same, i.e. the controllers, will be presented. These are PCB boards without IrDA port and equipped instead with other circuits useful for controlling a given comfort module. The appearance of the controllers separated from the actuator is shown in figure 4.

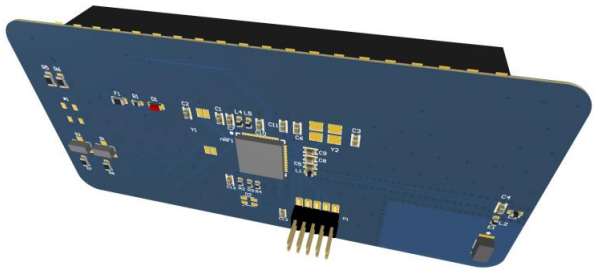


Fig. 2. Appearance of the actuator and controller of an IrDA-enabled air conditioner

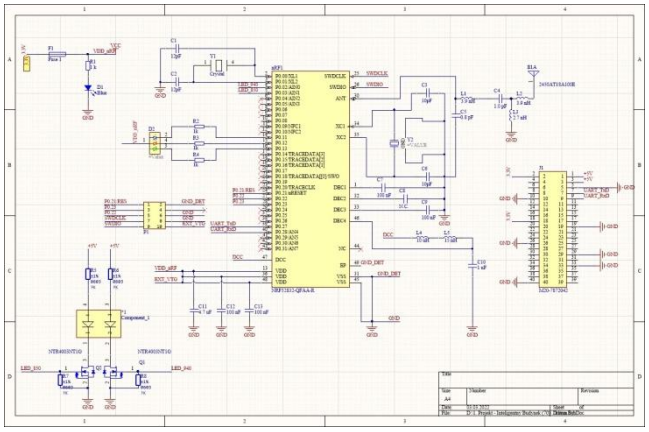


Fig. 3. Schematic diagram of an actuator and controller of an air conditioner equipped with an IrDA port

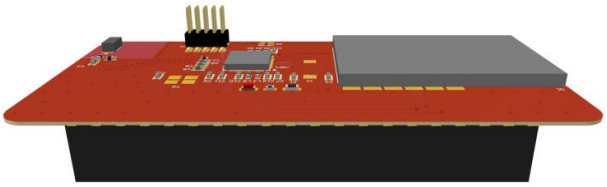


Fig. 4. Appearance of the heating controller

The common feature of this group is the appearance of a wireless communication module using the Z-Wave protocol. This treatment was intended to give the platform versatility in terms of actuator selection.

Due to the presence of two radio transmitters (Bluetooth/Zigbee and Z-Wave), the transmitting and receiving parts were moved away from each other by the maximum possible distance. Therefore, on each of the boards they are located at two opposite ends of the PCB. This is all the more important due to the fact that these protocols (including Wi-Fi) operate on the same band (ISM 2.4 GHz) only with a different band distribution. Additionally, the same type of modulation should be taken into account here (modulation with phase-shift keying and approximation with Gaussian curve) [16].

The electronics part using Wi-Fi technology is not directly on the boards shown, but on the Raspberry Pi Zero W single-board computers, which are connected to the designed PCBs via a 40-pin connector. This makes the complete controller, when placed in the enclosure, take the form shown in figure 5.

Fig. 5. The prototype of the comfort controller: the parts of the controller responsible for communication with the actuators and the single board computer managing the network communication enclosed in a single housing

The primary tasks of the Raspberry Pi Zero W single board computers are to maintain a network connection to the server service, support the Apache Kafka data bus, parse and process the data and then interpret it correctly and pass it on to the part responsible for communicating directly with the actuator. It is worth noting that the controller does not have to support just one actuator. Just bear in mind that too many slaves may cause minor delays in communication.

2.2. Comfort management – software

Programmatically, the upper layer of the controller in figure 5 represents (in the realised case) a simple operating system designed for embedded systems with radio stack support. All the code managing this hardware part was realised using C language (embedded). In addition to processing incoming and outgoing data from the controller, the BLE stack was used to cyclically transmit radio signals with a range of up to 30 metres. These signals are compatible with the Bluetooth version 5 radio transmission standard and advertising mode. This means that the devices do not force each other to connect permanently, i.e. "pairing", but work on the principle of miniature radio stations, where each message can be received by any device of interest in the system [1].

A single board computer running the Linux OS family is responsible for delivering instructions from the Home Assistant service to the WSoC (Wireless System on a Chip). It is connected to the local area network of the building and using a script that runs automatically at system start-up, it initiates the Apache Kafka producer and consumer module written in Qt Creator with the help of C++ language. Thanks to this the device has obligatory two-way communication up to the Linux system level. The connection between the computer and the wireless microcontroller is realised using a serial port. Although it provides two-way communication, it is used or not depending on building subsystem. In the case of comfort controllers, two-way communication is used.

In the case of the GUI (Graphical User Interface), the creation of a control panel that allows the user to remotely manipulate each of the settings offered by the selected actuators should be emphasised. Its appearance is presented in figure 6.

Fig. 6. View of the Home Assistant control panel for controlling the designed comfort modules

3. Environmental data acquisition system

Environmental data acquisition inside the building is provided by proprietary electronic modules equipped with sensors for temperature, humidity, light and carbon dioxide content in the air.

3.1. Environmental data acquisition – hardware

The sensors can be connected to the microcontroller using the communication interfaces it offers, such as UART, I2C or SPI. The WSoC nRF52832 microcontroller offers them in sufficient numbers to connect all the necessary sensors and to leave a reserve in case the design evolves. For analogue sensors, the platform offers A/D converters with 12-bit resolution, which are quite sufficient for the application presented. The appearance of the sensor assembly design is presented in figure 7, while the schematic diagram is presented in figure 8.

Fig. 7. Appearance of the actual environmental measuring module for indoor mounting (in the upper left corner of the drawing a 3D visualization of the PCB reverse side has been added, showing a cage for mounting a CR2032 battery with ~320 mA capacity)

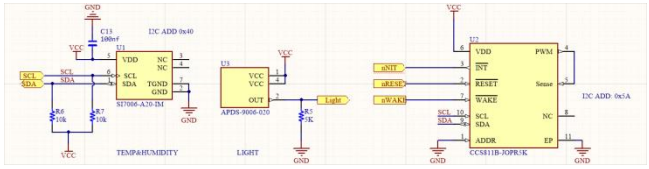


Fig. 8. Schematic diagram of the sensors and their electronic circuits used in the design of the module for testing environmental parameters

The environmental monitoring module includes temperature and relative humidity sensor SI7006-A20-IM (for use with heating and air-conditioning control modules), light intensity sensor APDS-9006-020 (for use with luminaire control modules) and carbon dioxide sensor CCS811B-JOPR5K (for use with ventilation control module), which is also capable of recording the content of organic particles in the air (VOC – Volatile Organic Compounds) expressed in ppm (particles per million) or ppb (particles per billion). Taking into account the completed set, it can be concluded that the sensor is suitable for basic tests of air quality in a building [4, 11].

The data obtained from this device is pre-formatted by it and then transmitted to the service and displayed on the control panel. Properly archived, they also provide material for machine learning algorithms.

3.2. Environmental data acquisition – software

As with the comfort modules, communication between the sensors and the system was implemented using Bluetooth 5. A 31 byte user data container was used for data transport. It is nested in a frame compatible with BLE advertising packet data transmission. Data from individual communication interfaces and analogue-to-digital converters are collected and then formatted to fit into the specified container size. The way the data is formatted is not the same for all sensors, because in the case of temperature the result is stored on two bytes, while for the light sensor four are reserved. They also differ in the way of coding or in the byte stacking weight (LSB/MSB). As an example, we can give a temperature sensor, which, despite the nature of its application (eliminating the need to work in negative temperatures) takes into account the sign of the returned value in the results. All these discrepancies had to be taken into account both in the comfort controllers and their scripts handling communication for the Home Assistant, and in the service located on the server (Node-RED).

The participation of the comfort modules in the communication is dictated by the fact that the sensors do not have their own unit providing connections to the local network, so they use neighbouring computers belonging to the light, ventilation, etc. controllers. This makes the controllers a kind of hub. This surplus of computing power means that this has been achieved without compromising the working conditions of the other components of the system. In addition, even if the sensors were equipped with a Wi-Fi module (a wired connection was rejected due to installation constraints), it would not guarantee a long operation time per single cell due to battery power.

The data collected by the comfort controller is reformatted to transmission standards using the JSON standard and sent to the service and then presented in a clear form in the operator panel. Its appearance is presented in figure 9.

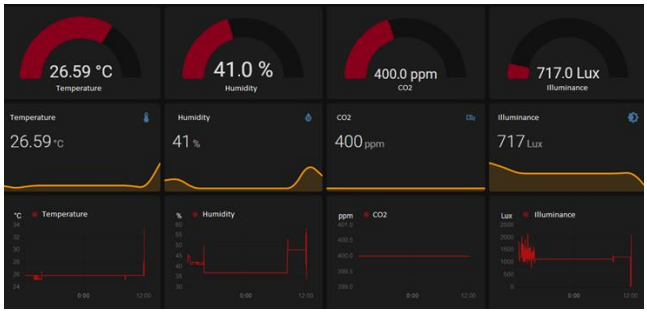


Fig. 9. View of the Home Assistant control panel for monitoring the environmental parameters of a building at a selected point

4. Indoor navigation and RTI system

The indoor navigation system is divided into two parts: classical by means of beacons and RTI (Radio Tomographic Imaging) supported. The beacon part uses location techniques based on the user's device and triangulation or related methods. It is characterised by good scalability, relatively low production costs, and low failure rates, but it should not be expected to have a high tracking accuracy (it achieves an accuracy of 1-2 metres) [6].

RTI, on the other hand, does not require any direct electronic device to track the user and is characterised by an increased level of positioning accuracy (up to several tens of centimetres). However, it is characterised by a higher degree of system complexity.

The purpose of the beacon navigation system is to guide the user from point A to point B through large, open spaces and main communication routes. After reaching a specific room and leaving the corridor, the RTI system takes the initiative, which is able to identify the presence of the user in the room, and then put his signature on the heat map.

The principle of RTI operation is based on the area of the room surrounded by networks of radio probes, which receive from other transmitters, apart from themselves, information about the received signal strength. Then, after conversion and conversion, this signal takes the form of the RSSI (Received Singal Strength Indicator). Each connection of the transmitter and receiver can be treated as a projection angle, i.e. one thread of the mesh, the examination of which (and more precisely, examination of the signal

fluctuations occurring in them) gives information about the presence or movement of users. The mathematical representation of this grid is the matrix, which is divided into two categories: the background matrix (i.e. the representation of the room at rest) and the measurement matrix (i.e. the representation of the room with recorded disturbances). In order to extract disturbances from the measurement matrix, it is necessary to perform a differential analysis with the background matrix. Then, on the basis of appropriate algorithms from the matrix, the heat map is reconstructed.

4.1. Indoor navigation and RTI – hardware

The hardware layer of the beacon system is represented by the so-called beacon, i.e. a miniature, energy-saving (most often battery-powered) and cheap to produce radio signal transmitter [2]. In the system offered, a proprietary structure of this type was created, the dimensions of which do not exceed 30 mm in diameter (wheel shape) and are powered by a single CR2477 lithium cell with a capacity of 960 mAh. The appearance of the beacons is presented in figure 10.

A cell with the indicated capacity and voltage of 3 V in combination with energy saving techniques implemented in beacons (standby, put to sleep) led to a working time counted in months. The maximally simplified structure in relation to other devices makes this structure perfect for production in large quantities. After the implementation of the microcontroller from the nRF family, only the source of the clock signal and a correctly designed antenna path were necessary to start the WSoC, but to ensure the correct start-up at each start-up and stable operation, it was necessary to use additional RLC elements.

Fig. 10. The appearance of the designed beacons

The hardware layer responsible for the RTI system includes two devices: a radio probe and a central unit. The radio probe is a BLE radio transmitter enriched with a wireless communication module compatible with the IEEE 802.15.4 standard. It is used alternatively for BLE due to the increased stability of the sent signals. The central unit (CU), in turn, is a Raspberry Pi 4 or Jetson Nano single-board computer equipped with a proprietary Bluetooth expander (it was replaced with a factory network card due to instability in the presented solution), a 7-inch TFT touch screen and an omnidirectional 30 cm long antenna. The appearance of the devices is presented in figures 11 and 12.

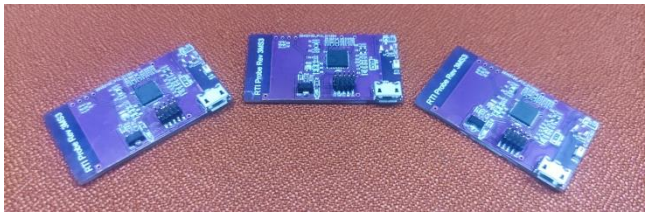


Fig. 11. Appearance of the designed radio probes

For both beacons and RTIs, a sufficient number of units are required for operation. The number of beacons depends on the floor area of the building and its arrangement plan. In RTI, an average system cell spanning one room contains sixteen radio probes. This value is sufficient to cover medium-sized office spaces. Otherwise the number can be increased or decreased (usually times / by 2). The central unit consists of one device per RTI cell, i.e. it covers one room. It is possible to cover a larger number, but the location should be considered in such a way that it receives a set of signals from each room [14].

Fig. 12. The appearance of the prototype of the central unit during operation (visible outline of the measurement matrix)

4.2. Indoor navigation and RTI – software

In terms of software, the beacons were implemented in accordance with the iBeacon standard. Using the advertising designed for this purpose, they periodically inform all interested devices about their RSSI, minor and major (these are a kind of identifiers of belonging to specific groups of devices serving a common area), Tx Power and a unique UUID identifier [2]. All data in the beacon frame are constant (as macro-definitions), therefore their cyclical update is not needed. This makes communication one-way by definition. If necessary, a subprogram for remote software update has been implemented in all transmitters so that the operator with the help of a smartphone with an application and a data package containing a complicated program can (without having to search for, get in or remove the device) introduce changes to the system without any problems [3]. A similar system has been used in the case of radio probes, but the device, unlike a beacon, often changes the content of its data packet. Often, these changes cannot even be captured with the unaided eye due to the limited refresh times of scanners, and their effects are only visible in the image reconstruction. The application (apart from handling the Bluetooth layer) has three tasks. Get information on the strength of signals from other probes, preformat the measurement results and activate subsequent devices to work. The entire code of the beacons and radio probes was implemented using the C (embedded) language [7]. The central unit software is a system from the Linux OS family. The software supporting the RTI system was realized with the help of the C ++ language in the Qt Creator environment. The generated script is run at the system start-up and can automatically start the measurement or wait for the user's command. Figure 13 shows a typical heat map realized by the built RTI system.

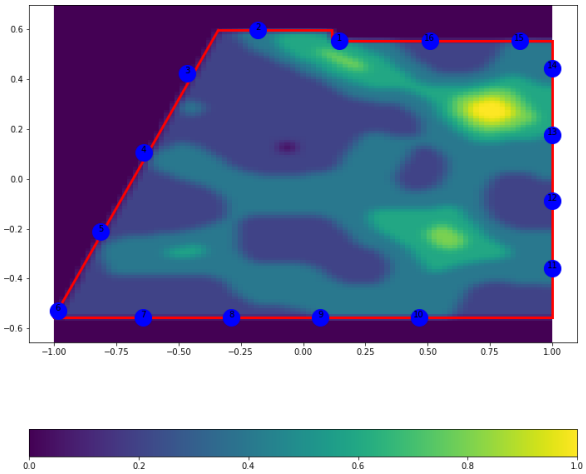


Fig. 13. Heatmap of the test room with one person inside (visible in the upper right corner of the room)

5. Access control

Access control is a derivative of navigation with beacons and differs from it mainly in the way of placing this type of transmitters and the software. In terms of hardware, there are no significant differences to be specified. At least as far as the radio signal transmitter is concerned. To the sum of devices, you need to add a wirelessly controlled door lock module and hardware compatible with it. The control is carried out using the user's telephone. The signal received from the directional antenna mounted on the door frame, after exceeding a certain threshold, sends reports to the system and releases the lock. After the user enters the room, in order to avoid the cyclical opening and closing of the door, the RTI module will not lock the lock back when detecting an animal, however, it can be done by manual change by the user. The view of the door management panel of one of the sections of the building is shown in figure 14.

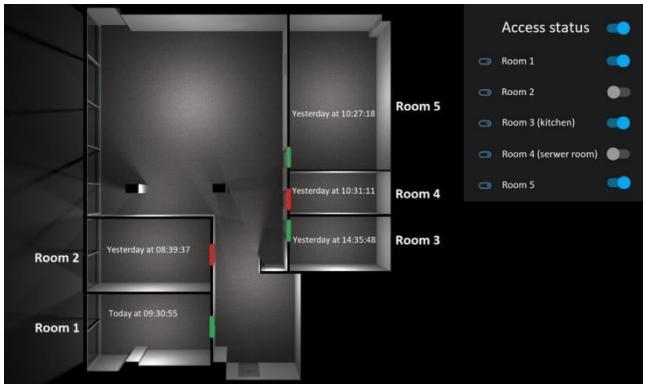


Fig. 14. View of the control panel for door access control in the Home Assistant service

6. Other subsystems

The remaining subsystem includes minor building functionalities, such as a panel with 3D visualizations of the building and active marking of users' positions, recording the working time of people and devices, and a panel for monitoring weather conditions outside the building. The method of implementation of the panels for monitoring the working time of devices and people is presented in figures 15 and 16.

In order to enable the observation of the operating time of the devices, the so-called smart plug with a built-in electricity meter and the possibility of wireless connection to the network via Wi-Fi. Thanks to them, it is also possible to estimate the expenses related to powering the building.

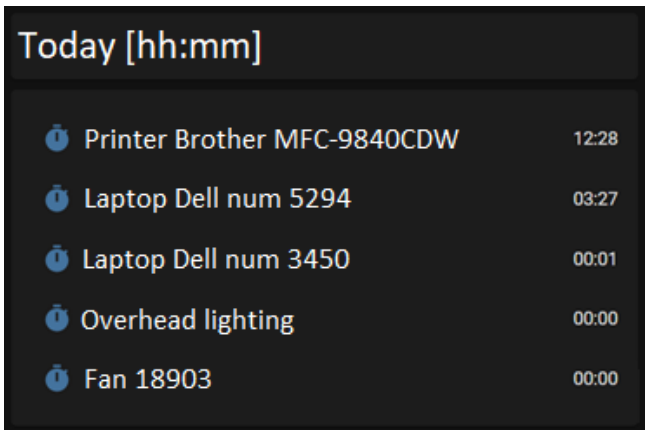


Fig. 15. Method of implementation of the panel for assembling employee working time

Fig. 16. The method of implementation of the panel for assembling the working time of devices

The method of recording the working time of employees was carried out with the help of the same application that is used for building navigation. After the system is registered on a given day, a device with a unique UUID assigned to the employee confirms his presence in the system until the time provided for in the supervision timeout has elapsed. This time can be modified in a very wide range.

The 3D visualization system is designed to present the building in a three-dimensional form, and then the coordinates obtained through the operation of beacon navigation and RTI are applied to the building map.

A plug that allows you to download data from weather services allows computational intelligence to better adjust the operating parameters inside the building. The energy input for summer / winter operation can vary considerably and must be taken into account in machine learning.

7. Computational intelligence

Most of today's devices dedicated to building automation are controlled with simple dependencies using information provided by sensors. The rules created usually refer to continuous control of a set parameter and starting or stopping a device depending on the current deviation [10].

In the project presented here, an attempt has been made not to rely solely on current data but also to archive historical data and use it in machine learning to act in a more personalised way towards users. This requires adapting the sensory infrastructure and actuators in such a way that any changes in them are registered by the system. The result is increased overall energy efficiency and a longer lifespan for individual building components without compromising working conditions [12]. The designed system integrates several original solutions with those already available on the market. In addition, the most up-to-date communications occurring in IoT (Internet of Things) can be included. The adopted structure of the system allows easy scalability and simple adaptation to different types of existing and future solutions on the market.

To achieve the assumed effect, proprietary algorithms supporting effective building management and machine learning mechanisms were used. The model is trained using previously collected data. This data relates primarily to the way individual rooms are used by the personnel assigned to them. Data acquisition and transport is carried out on the basis of mechanisms described in the previous subsections.

Building machine learning is carried out with the use of a strictly defined algorithm. Three different computational methods were used to solve the estimation problems, such as linear regression, statistical regression and the support vector machine. However, before any calculations can be made, each data must follow the path presented in figure 16. Based on the algorithm from Figure 16, the system's response to the user's presence is adjusted, and then the comfort modules are properly controlled.

In addition, figure 17 shows the control of the comfort modules taking into account selected elements of computational intelligence and their relationship to inputs and outputs.

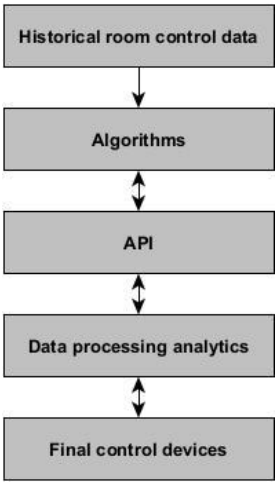


Fig. 17. Prediction and control of end devices

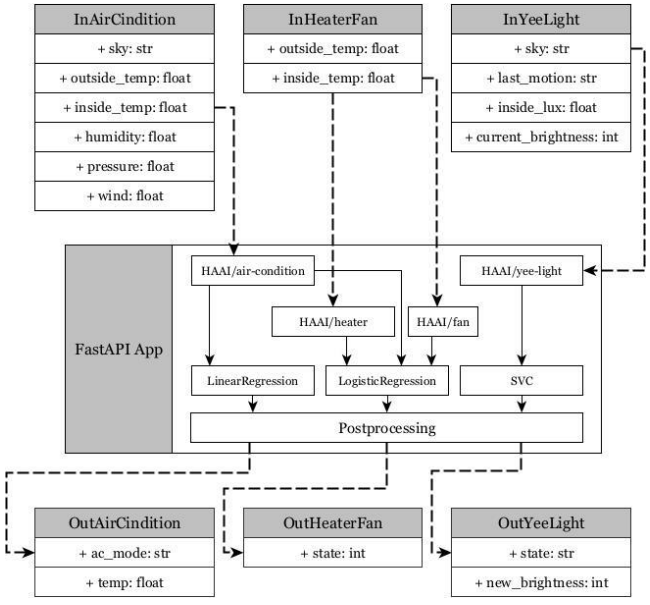


Fig. 18. Diagram of machine learning activities of the test building

As shown in figure 18 the simplest computational intelligence element used is linear regression. This makes it affordable to implement and relatively user-friendly. This mechanism is based on the assumption that a certain set of characteristics $N\,(x_1, \ldots, x_N)$, is known by which a continuous value of y is predicted. Then if a proportional relationship is found between the elements of the set N and the predicted values of y then the whole relationship can be approximated by the equation [8]:

$$\tilde{y} = c_0 + c_1 x_1 + \cdots + c_N x_N \tag{1}$$

or otherwise:

$$\tilde{y} = c_0 + \sum_{i=1}^{N} c_i x_i \tag{2}$$

where $\tilde{y}$ is assumed to be a linear combination of the characteristics denoted by x_i and c_i are the coefficients of the assumed model. The next requirement is to find the objective function, which can be calculated for n data points. The case described leads to the calculation of the mean square error using equation:

$$L = \frac{1}{n}\sum_{i=1}^{n}(\tilde{y} - y)^2 \tag{3}$$

In this way, the defined objective function is minimised. This makes it possible to determine the optimal value for the $\boldsymbol{c_i}$ variables while maintaining their minimum fitting error.

Another element used is logistic regression. To some extent, it can be thought of as a derivative of linear regression. The main difference here was treating the variable y not as a continuous element, but as a variable with a two-point distribution. This makes this mechanism work better in cases where we are dealing with two states, or in other words, a binary response. This causes the mechanism to operate on the values in the range [0,1]. This is called the conditional probability for the value $y = 1$ given the value of x [13].

$$P(y = 1|x) = p = \frac{1}{1+e^{-f(x)}} \tag{4}$$

where $f(x)$ should be treated as linear combination of independent variables (in other words, features). It can be expressed in the same way as the value $\tilde{y}$ in linear regression, that is:

$$f(x) = c_0 + \sum_{i=1}^{N} c_i x_i \tag{5}$$

It should be added that in such a case the occurrence of the opposite event will be complementary and expressed by the equation:

$$P(y = 0|x) = 1 - p \tag{6}$$

Determination of the objective function that will enable the determination of c_i coefficients is possible after defining the model. The c_i coefficients should be treated as parameters of the chosen model. Thus, in the case of a two-point distribution, the square of the error should not be used this time, but the likelihood function given by the formula:

$$L(C|y;x) = \prod_i p(x_i)^{y_i}\left(1 - p(x_i)\right)^{(1-y_i)} \tag{7}$$

Ultimately, the objective function is the logarithm of formula 7, which is then maximized to find optimal parameter sets. Simple gradients are used for this process.

The last of the three elements is SVC that is support vector classifier. It is especially useful when working with data broken down into binary classes. Belongs to support vector machines (SVM). Compared to the two previous methods, it can be used to make accurate indications even when the classes do not show a tendency in proportion to the changes in features (in other words, when there is a non-linearity) [15].

The method aims to find a hypersurface that does not belong to either of the two binary classes, separates them from each other, and at the same time is as far away from both of them as possible. Moving on, any of the outputs marked with X can be represented as a point in a multidimensional space. Each of these points can be described as follows:

$$Y = \{0,1\} \tag{8}$$

In the process, the SVC mechanism tries to find an area in space that separates the two sets. Typically, hyperplanes are used to split data apart.

If, however, the separation process becomes ineffective, then the SVC can extend the model with additional dimensions until the effect is achieved, i.e. until the place where the classification is more optimal is found. In some cases, it is possible to use nonlinear nuclei incapable of generating hyperplanes but surfaces with a greater degree of shape complexity.

The selection of optimal parameters for the SVC method can be performed by minimizing the expression:

$$\left[\frac{1}{n}\sum_{i=1}^{n} max\left(0,1-y_i(W^T x_i-b)\right)\right]+\lambda||W||^2 \qquad (9)$$

where the number of points in the data set should be taken as *n*, and the matrix of coefficients in W. The λ parameter should be treated as a regularization factor, the lack of which will make it impossible to determine the objective function. In this case, an initial regularization parameter of 0.0001 was used. It should only be added that the base radial function was used as the nucleus. Finally, the presented transformations lead to the receipt of a model that needs to be checked in terms of the accuracy of its predictions. If so, you can start its implementation in the developed solutions.

8. Research and summary

During the research, a huge amount of data related to the functioning of the building was recorded, such as environmental parameters, both internal and external (weather station), through user location data, record of employees and equipment working time, non-contact access control (ten-meter standard of Bluetooth range), detection of equipment idleness and neglect resulting from not exclusive of receivers.

In the case of the comfort modules and their actuators, manipulating the parameters in the service panel revealed a delay response of less than 1 second. What is taken into account here is the full path that the signal takes from the server through the network to the controller module and then through the PAN to the actuator. The only exception here was the actuator for the heating system, which was a DC motor for the valve control head. Its response time varied within 5 seconds. Given that the heating system has the highest inertia this delay was imperceptible in practice. The sensor network acquiring environmental data was similarly efficient.

Research conducted on combined detection and prediction systems has eliminated instances of wasted resources due to staff negligence. Together with other aspects of work optimization resulting from the use of machine learning, energy savings of 20% were achieved. Confirmation was provided by a series of test scenarios consisting of two parts. The first involved user behavior recorded during learning, the second was a decision automatically made by the system when a request was made. As an example, the cooling in a conference room was started as soon as work began when the room was scheduled to be used after 12:00 p.m. Allowing the computational intelligence to automatically select the start time and air conditioner settings saved approximately 350 Wh of energy. The management mechanism has done this on the basis of data on the structure of the building, the weather forecast and the technical parameters of the equipment it controls. In another case, the service carried out an automatic shutdown of the air conditioner due to the absence of detected identifiers (radio beacons) and human signatures (RTI) in the room left behind. The cooperation between the radio beacon system and the RTI has proved to be important, as in the absence of one of them the certainty of the choice is reduced. During the observation, for example, cases were registered in which the room remained empty for a long time, but had an identification signal from the telephone locked in it. The absence of an employee in such a case can only be confirmed by the absence of a signature on the heat map. Based on the observations of users in the room, the optimum height of the radio probe belt of 120 cm was worked out (analysis of heat maps and the reaction of the system to the introduced disturbances). This was due to the fact that there is the greatest amount of matter at this height. For an average height European, this is the range of the abdomen and waist.

Future plans include increasing the level of integration of individual building elements (adding new actuators) and extending the areas covered by the RTI system, i.e. expansion from 16 to 32 and more radio transmitters to cover not only offices but also warehouses.

References

[1] Bluetooth SIG. Bluetooth Core Specification, version 4.0.; Bluetooth SIG: Kirkland, WA, USA, 2010.
[2] Cavallini A.: iBeacons Bible, 2015.
[3] Corna A., Fontana L., Nacci A. A., Sciuto D.: Occupancy detection via ibeacon on android devices for smart building management. Proceedings of the Design, Automation & Test in Europe Conference & Exhibition (DATE), 2015, 629–632.
[4] Havard N., McGrath S., Flanagan C., MacNamee C.: Smart Building Based on Internet of Thing Technology. Proceedings of the 12th International Conference on Sensing Technology (ICST), 2018.
[5] Kozłowski E., Mazurkiewicz D., Żabiński T., Prucnal S., Sęp J.: Machining sensor data management for operation-level predictive model. Expert Syst. Appl. 159, 2020, 1–22.
[6] Liu H., Darabi H., Banerjee P., Liu J.: Survey of wireless indoor positioning techniques and systems. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. 37, 2007, 1067–1080.
[7] Matteuchi M.: An Adaptive Indoor Positioning System Based on Bluetooth Low Energy RSSI. Politecnico di Milano, Milano 2012.
[8] Montgomery D. C., Peck E. A., Vining G. G.: Introduction to Linear Regression Analysis. World Scientific Publishing, Singapore 2012.
[9] Node-RED Guide. Available online: http://noderedguide.com/ (accessed on 19.02.2022).
[10] Peng Y., Rysanek A., Nagy Z., Schlter A.: Using machine learning techniques for occupancy-prediction-based cooling control in office buildings. Appl. Energy 211, 2018, 1343–1358.
[11] Postolache O. A., Dias Pereira J. M., Silva Girao P. M. B.: Smart Sensors Network for Air Quality Monitoring Applications. IEEE Trans. Instrum. Meas. 58, 2009, 3253–3262.
[12] Rivera-Illingworth F., Callaghan V., Hagras H.: Automated Discovery of Human Activites inside Pervasive Living Spaces. Proceedings of the International Symposium on Pervasive Computing and Applications, 2006, 77–82.
[13] Rymarczyk T., Kozłowski, E., Kłosowski G., Niderla, K.: Logistic Regression for Machine Learning in Process Tomography. Sensors 19, 2019, 3400.
[14] Styła M., Oleszek M., Rymarczyk T., Maj M., Adamkiewicz P.: Hybrid sensor for detection of objects using radio tomography. Applications of Electromagnetics in Modern Engineering and Medicine, PTZE, 2019, 219–233.
[15] Suykens, J.A., Vandewalle J.: Least squares support vector machine classifiers. Neural Process. Lett. 9, 1999, 293–300.
[16] Tragos E. Z., Foti M., Surligas M., Lambropoulos G., Pournaras S., Papadakis S., Angelakis V.: An IoT based intelligent building management system for ambient assisted living. Proceedings of the IEEE International Conference on Communication Workshop (ICCW), 2015, 246–252.

M.Eng. Michał Styła
e-mail: michal.styla@cbrti.pl

Michał Styła received the M.Eng. degree in electrical engineering from the Lublin University of Technology, in 2018. Since 2020, he fulfills a role of specialist in the field of wireless communication solutions at CBRTI sp. z o.o.. His field of activity encompasses design and programming of electronic circuits connected with radio communication techniques and their subsequent implementation in tracking, navigation and radio tomographic imaging systems.

http://orcid.org/0000-0002-1141-0887

Ph.D. Przemysław Adamkiewicz
e-mail: przemyslaw.adamkiewicz@cbrti.pl

Przemysław Adamkiewicz received the Ph.D. degree in physics from the Maria Curie-Sklodowska University, in 2013. Since 2021, he has been a CEO at CBRTI sp. z o.o.. Leader and participant in several international projects. His research area focuses on the application of embedded IoT solutions, electrical tomography, image reconstruction, process tomography, radio tomography imaging, artificial intelligence and computer measurement systems.

http://orcid.org/0000-0003-3425-9566

http://doi.org/10.35784/iapgos.2888 *received: 02.03.2022 | revised: 11.03.2022 | accepted: 29.03.2022 | available online: 31.03.2022*

DEVELOPMENT OF DEPOSITION TECHNOLOGY AND AC MEASUREMENT OF COPPER ULTRATHIN LAYERS

Aleksandra Wilczyńska[1], Karolina Czarnacka[2], Andrzej Kociubiński[3], Tomasz N. Kołtunowicz[1]
[1]Lublin University of Technology, Department of Electrical Devices and High Voltage Technology, Lublin, Poland, [2]University of Life Sciences in Lublin, Department of Technology Fundamentals, Lublin, Poland, [3]Lublin University of Technology, Department of Electronic and Information Technology, Lublin, Poland

Abstract. *In this paper, the transport properties of discontinuous 4 nm copper layers obtained by dual-source non-reactive magnetron sputtering in the presence of argon are presented. The value of resistance and capacitance of the current parallel to the plane of these layers can be adjusted independently by changing the nominal thickness of the metallization. The influence of frequency on the conductivity of the obtained structures in the range from 4 Hz to 8 MHz was studied. Additionally, in order to compare the non-oxidized and oxidized layers, some of them were heated at 500°C. Based on the results obtained, the mechanism of electric charge transfer was determined, the knowledge of which is essential for planning further experiments based on this sputtering method and potential selection of future application of the structures. Statistical measurements at room temperature will serve as a reference for the conductivity and resistivity values obtained by mathematical calculations from measurements of resistance, capacitance, phase shift angle, and dielectric loss tangent as a function of temperature from 20 K to 375 K, which are expected in further studies on the obtained structures. The work is an introduction to the technology of obtaining multi-layer metal-dielectric structures.*

Keywords: magnetron sputtering, ultrathin layers, AC measurement, conductivity measurement, charge transport, Cu films

OPRACOWANIE TECHNOLOGII OSADZANIA I POMIARÓW ZMIENNOPRĄDOWYCH ULTRACIENKICH WARSTW MIEDZI

Streszczenie. *W niniejszej pracy przedstawione zostały właściwości transportowe nieciągłych 4 nm warstw miedzi otrzymanych metodą dwuźródłowego niereaktywnego rozpylania magnetronowego w obecności argonu. Wartość rezystancji i pojemności prądu równoległego do płaszczyzny tych warstw można dostrajać niezależnie poprzez zmianę nominalnej grubości metalizacji. Przebadano wpływ częstotliwości na konduktywność otrzymanych struktur w zakresie od 4 Hz do 8 MHz. Dodatkowo, w celu porównania nieutlenionych i utlenionych warstw niektóre z nich zostały wygrzane w temperaturze 500°C. Na podstawie otrzymanych wyników określono mechanizm przenoszenia ładunków elektrycznych, którego znajomość jest niezbędna do planowania kolejnych eksperymentów bazujących na tej metodzie napylania oraz potencjalnym doborze przyszłego zastosowania struktur. Statystyczne pomiary w temperaturze pokojowej posłużą za punkt odniesienia dla wartości konduktywności i rezystywności otrzymanych na drodze obliczeń matematycznych z pomiarów rezystancji, pojemności, kąta przesunięcia fazowego oraz tangensa strat dielektrycznych w funkcji temperatury od 20 K do 375 K, które przewidywane są w dalszej części badań nad otrzymanymi strukturami. Praca stanowi wstęp do technologii otrzymywania wielowarstwowych struktur typu metal-dielektryk.*

Słowa kluczowe: rozpylanie magnetronowe, ultra cienkie warstwy, pomiary zmiennoprądowe, pomiary konduktywności, transport ładunków, warstwy miedzi

Introduction

In order to meet the continuous demand of the modern semiconductor industry for embedded condensers with smaller dimensions and unchanged or better properties, nanocomposites composed of a matrix of non-conductive materials have become increasingly popular due to their excellent dielectric properties [16]. The essence of conducting research on material properties in an electrical context is to understand the physical phenomena occurring in them. Of particular interest is the possibility of tuning the resistance, capacitance and inductance over a wide range at the stage of obtaining composite structures by various methods [7].

Microprocessor technologies use many conductive materials to build electrical components. One of them is copper, which is used particularly often due to its conductive properties. There are many methods for the deposition of thin copper layers, such as: DC magnetron sputtering (DCMS) [18], high power impulse magnetron sputtering (HiPMS) [3], ion beam deposition [15], quenching [22], chemical vapour deposition (CVD) or more environmentally friendly supercritical fluid chemical deposition (SFCD) [8].

Nowadays, nanotechnology is one of the fastest developing engineering fields in the world due to its possibility of obtaining precisely nanometric structures for many applications [9]. Unfortunately, nanoscale particles don't always have the same properties as micro-scale particles of the same material. X-ray fluorescence (XRF) studies have shown that the measured value of copper nanoparticles in dispersed systems is less than micro-particles [5]. Moreover, grain grinding has a beneficial effect on the electrochemical properties of pure copper. With the reduction of the grain size, the passive and corrosive current density decreases, and the acceptor density decreases. However, the grain size doesn't affect the type of semiconductor in thin films [11]. On the basis of the available structural studies, it can be concluded that the layers ≥ 130 nm are uniform and have an amorphous character, and with the increase in thickness, arched grain clusters appear, which, getting larger and larger, smoothing the surface. In contrast, increasing the metallization thickness reduces the resistivity of the structure to a value of 500 nm. Beyond this level, no effect of film thickness on resistivity is observed. This correlation can be defined as the inverse linear dependence of resistivity on grain size caused by reduced scattering at their boundaries by charge carriers [12].

In the case of ultra-thin copper layers below 10 nm, it has been noticed that they show strong increases in resistivity with decreasing film thickness, which is related to electron scattering by phonons, point defects, impurities, grain boundaries, or substrate / layer boundaries. [12, 17]. It is assumed that copper layers below 5 nm will be discontinuous, therefore they will show quite different electrical properties compared to continuous layers [2, 6, 10].

Using this information, it seems reasonable to use copper as one of the phases of metal-dielectric nanocomposites. Based on the available AC studies, it is assumed that the obtained structures will exhibit a voltage resonance phenomenon at the resonance frequency characteristic of a series RLC system, which was produced by a stepped charge transfer mechanism [7, 13, 21]. An important element in selecting the potential application of nanocomposite structures in which metal grains are embedded in a dielectric matrix is determining the value of the percolation threshold (x_c). In the case of metallic phase content (x) below this threshold ($x < x_c$) or in its vicinity ($x \approx x_c$) it is possible to determine the mechanism of electric charge transfer between the neutrally charged potential wells [9].

The aim of the work was to conduct the process of deposition of ultra-thin discontinuous copper layer, which is an introduction to the development of the technology of multilayer metal-dielectric composites.

1. Materials and technology

Thin-layer structures will be obtained by sputtering the material in the process of magnetron sputtering. This technique is based on the explosion of single particles of a material source under the influence of energy ionized by a strong electric field of a noble gas, in this case argon. The ions with an energy of 100 V to 1 kV formed in the plasma area bombard the target surface and breaking the bonds between individual atoms. High frequency of direct current or its discharges make it possible to obtain sputtering ions. A fluorescent current excitation is created between the anode of the grounded substrate and the cathode of the sputtered material. Magnetron sputtering is an extremely efficient sputtering method because it allows for the extension of the free electron path [1].

The first stage of work on obtaining an ultra-thin layer of copper metallization was mounting laboratory slides with dimensions of approx. 15 mm × 10 mm × 2 mm on a rotating plate. The deposition process was carried out by the technique of non-reactive magnetron sputtering in a sputtering machine Kurt J. Lesker Nano 36^{TM}, which previously had a vacuum of 1×10^{-7} Torr. During spraying, no high temperatures were used that could damage the source material. A 99.999 % purity target purchased from Kurt J. Lesker offer was used to obtain the copper layer. The process was initiated by a DC magnetron in the sputtering parameters: argon flow 75 sccm, plasma power 100 W for about 2 minutes. Thin film structures with dimensions of 10 mm × 5 mm × 4 nm were obtained. Contacts were applied to them using silver paste, as schematically shown in Fig. 1. Then some of them were heated at the temperature of about 500°C in order to oxidize the metallization.

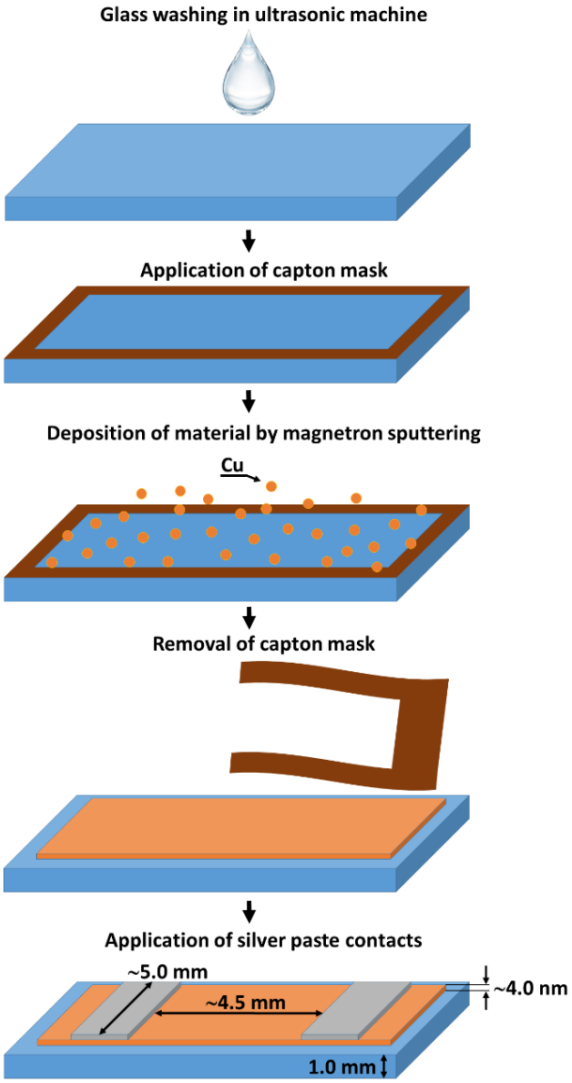


Fig. 1. Scheme and geometrical dimensions of the tested samples

2. Experimental

After applying the silver contacts, measurements were made using the four-point method using a Hioki IM3536 impedance meter. Before starting the measurements, the meter was calibrated using a dedicated test fixture Hioki 9261. The LCR meter program, available on the manufacturer's website for download, was used for the first measurements. The measurement parameters were resistance, conductance, phase angle and capacitance as a function of frequency in the range from 4 Hz to 8 MHz. Measurements were carried out at room temperature 10 times for each frequency and then subjected to statistical calculations. From the obtained measurements by mathematical calculation, the conductivity was determined using the geometric dimensions of the sample between contacts using the electrical conductivity formula of the material:

$$\sigma = \frac{G \cdot l}{S} \tag{1}$$

where: σ – conductivity $\left[\frac{S}{m}\right]$, G – conductance $[S]$, l – side length $[m]$, S – element cross-sectional area $[m^2]$.

In order to visualize the obtained measurement results and parameters from the mathematical calculations, code was written in Matlab.

3. Results and discussion

Fig. 2 and 3 show the dependence of conductivity on the measurement frequency of the deposited copper layer before and after annealing. The graphs show that the conductivity value in the low frequency range up to approx. 10 Hz is practically constant and then increases linearly by approx. 5 orders of magnitude. Based on the AC model of electric charge transfer presented in the works [25, 26, 27], it can be concluded that the layers aren't continuous, and the dominant mechanism is electron hopping or electron tunneling between copper grains. In the case of a continuous copper layer, the conductivity should show a practically constant value with a frequency, as in the case of a thin-film resistor [20].

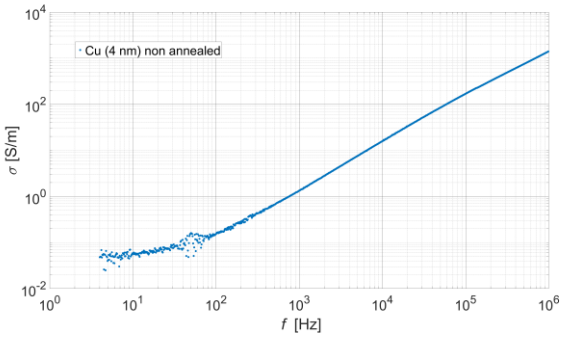


Fig. 2. Dependence of conductivity σ on the measuring frequency f of the copper layer immediately after deposition

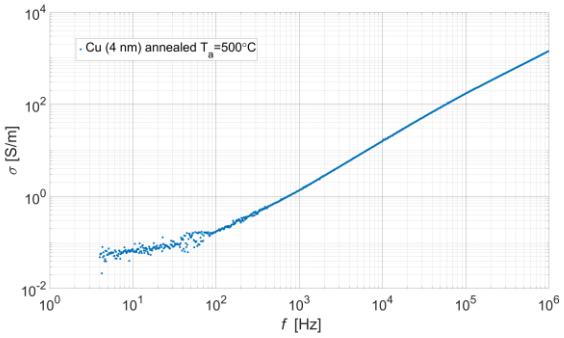


Fig. 3. Dependence of conductivity σ on the measuring frequency f of the copper layer heated at the temperature T_a=500℃

Based on the proposed model, the frequency coefficient m was determined, which is the derivative of lgσ after lg$f$ and $\sigma \sim f^m$ [19]. The values of the m(f) layer relationship before and after annealing for different frequency ranges are shown in Table 1. As a rule, the value of m(f) at the low frequency range should be close to 0, which indicates DC conductivity. For both low-frequency and high-frequency values of m(f), it can be seen that they are slightly different from each other. Thus, it can be concluded that the annealing of the structures will not have a significant effect on their electrical properties.

Table 1. Dependence of the frequency coefficient m on the measurement frequency f of the copper layer immediately after deposition and the copper layer heated at temperature T_a=500°C

	Nonannealed copper layer	Annealed Ta=500°C copper layer
m [a.u.] for 4 Hz–80 Hz	0.296865856	0.244488744
m [a.u.] for 1 kHz–100 kHz	1.051019326	1.047038941

Fig. 4–5 show the frequency dependence of the phase shift angle of the samples before and after annealing, respectively. For an ideal resistor, that is, a continuous metallic layer, the angle should be constant at 0° [20]. In the case of the obtained copper layer, in both cases, the angle is negative and decreases with frequency reaching values close to -90°. This is characteristic of a conventional RC parallel system.

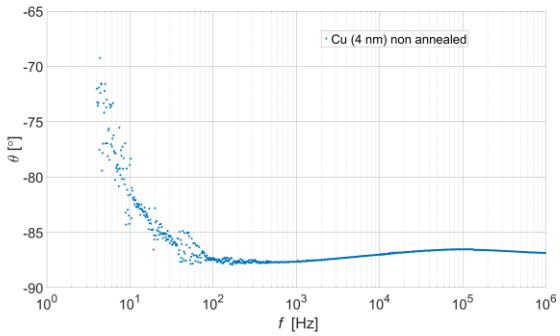


Fig. 4. Dependence of the phase shift angle θ on the measurement frequency f of the copper layer immediately after deposition

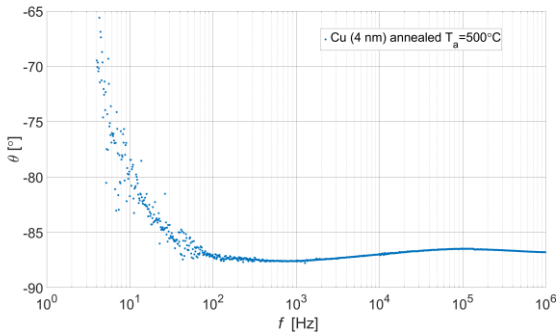


Fig. 5. Dependence of the phase shift angle θ on the measurement frequency f of the copper layer heated at temperature T_a=500℃

Literature data describe that the DC resistance of ultra-thin copper films increases significantly when the films are less than 10 nm thick [3, 14, 23], by up to two orders of magnitude compared to micrometric films [24]. It has also been shown that layers with thicknesses below 5 nm are discontinuous and consist of small islands [10, 17]. Therefore, it is justified to occur abrupt exchange of charges or tunneling between the formed islands and the percolative nature of the obtained structures [2, 6, 10].

4. Conclusions

A technology for deposition of discontinuous thin copper layers using DC magneto-sputtering onto glass substrates was developed. The process was carried out in a vacuum at the level of $1 \cdot 10^{-7}$ Torr, in an argon atmosphere with a flow of 75 sccm and a plasma power of 100 W for about 2 minutes. In this way, 4 nm of copper was deposited. Then, AC measurements of the obtained structures were carried out before and after annealing at 500°C. On the basis of the obtained results, it can be concluded that the obtained layer shows a percolation character and the dominant mechanism for transferring electric charges is electron hopping or their tunneling between the formed copper islands.

Comparing the obtained results of measurements of non annealed and annealed (oxidized) structures, it can be concluded that the values of the parameters are slightly different from each other, therefore, no influence of the appearance of oxygen particles on nanocomposites exhibiting the previously mentioned charge transfer mechanisms is expected.

Based on these observations, further research will be carried out on the preparation of nanocomposites with the structure of alternately arranged discontinuous metal and dielectric layers by magnetron sputtering. The comparison of results in this paper allows for annealing experiments of the planned structures.

References

[1] Beck R.: Technologia krzemowa. PWN, Warszawa, 1991.
[2] Biegański P., Dobierzewska-Mozrzymas E.: Electrical properties of discontinuous copper films. International Journal of Electronics 70, 1991, 499–504, [http://doi.org/10.1080/00207219108921300].
[3] Cemin F., Lundin D.: Low electrical resistivity in thin and ultrathin copper layers grown by high power impulse magnetron sputtering. Journal of Vacuum Science & Technology A 34(5), 2016, 051506-1–051506-7 [http://doi.org/10.1116/1.4959555].
[4] Chakarvarki S. K.: Track-etch membranes enabled nano-/microtechnology: A review. Radiation Measurements 44(9–10), 2009, 1085–1092, [http://doi.org/10.1016/j.radmeas.2009.10.028].
[5] Chebakova K. A., Dzidziguri E. L. et al.: Open AccessArticle X-ray Fluorescence Spectroscopy Features of Micro- and Nanoscale Copper and Nickel Particle Compositions, Nanomaterials 11(9), 2021, 2388, [http://doi.org/10.3390/nano11092388].
[6] Dingle R. B.: The electrical conducticity of thin wires. Proceeding of the Royal Society A Mathematical, Physical and Engineering Sciences, 1950, [http://doi.org/10.1098/rspa.1950.0077].
[7] Fedotov A., Mazanik A., Svito I., Saad A., Fedotova V., Czarnacka K., Kołtunowicz T. K.: Mechanism of Carrier Transport in Cux(SiO2)1-x Nanocomposites Manufactured by Ion-Beam Sputtering with Ar Ions, Acta Physica Polonica A 128, 2015, [http://doi.org/10.12693/APhysPolA.128.883].
[8] Giroire B., Ali Ahmad M., Aubert G., Teule-Gay L., Michau D., Watkins J. J., Aymonier C., Poulon-Quintin A.: A comparative study of copper thin films deposited using magnetron sputtering and supercritical fluid deposition techniques, Thin Solid Films 643, 2017, 53–59, [http://doi.org/10.1016/j.tsf.2017.09.002].
[9] Grimmet G.: Percolation, 2nd ed. Springer-Verlag, Berlin 1999, [http://doi.org/10.1007/978-3-662-03981-6].
[10] Hill R. M.: Electrical conduction in discontinuous metal films. Contemporary Physics 10, 1969, 221–240, [http://doi.org/10.1080/00107516908224594].
[11] Imantalab O., Fattah-alhosseini A., Keshavarz M. K., Mazaheri Y.: Electrochemical Behavior of Pure Copper in Phosphate Buffer Solutions: A Comparison Between Micro- and Nano-Grained Copper. Journal of Materials Engineering and Performance 25, 2016, 697–703, [http://doi.org/10.1007/s11665-015-1836-z].
[12] Kah-Toong Chan, Teck-Yong Tou, Bee-San Teo: Thickness dependence of the structural and electrical properties of copper films deposited by dc magnetron sputtering technique. Microelectronics Journal 37(7), 2006, 608–612, [http://doi.org/10.1016/j.mejo.2005.09.016].
[13] Kołtunowicz T. N., Żukowski P., Czarnacka K., Bondariev V., Boiko O., Scito I. A., Fedotov A. K.: Dielectric properties of nanocomposite (Cu)x(SiO2)(100−x) produced by ion-beam sputtering. Journal of Alloys and Compounds 652, 2015, 444–449, [http://doi.org/10.1016/j.jallcom.2015.08.240].
[14] Lacy F.: Developing a theoretical relationship between electrical resistivity, temperature, and film thickness for conductors. Nanoscale Research Letters 6, 2011, 1–14, [http://doi.org/10.1186/1556-276X-6-636].
[15] Lim J. W., Isshiki M.: Electrical resistivity of Cu films deposited by ion beam deposition: Effects of grain size, impurities, and morphological defect. Journal of Applied Physics 99, 2006, 094909-1–094909-7, [http://doi.org/10.1063/1.2194247].
[16] Lin Zhang, Xu Lu, Xinyu Zhang, Li Jin, Zhou Xu, Z.-Y. Cheng: All-organic dielectric nanocomposites using conducting polypyrrole nanoclips as filler. Composites Science and Technology 167, 2018, 285–293, [http://doi.org/10.1016/j.compscitech.2018.08.017].
[17] Liu H.-D., Zhao Y.-P., Ramanath G., Murarka S. P., Wang G.-C.: Thickness dependent electrical resistivity of ultrathin (<40 nm) Cu films. Thin Solid Films 384(1), 2011, 151–156, [http://doi.org/10.1016/S0040-6090(00)01818-6].
[18] Mech K., Kowalik R., Żabiński P.: Cu thin films deposited by DC magnetron sputtering for contact surfaces on electronic components. Archives of Metallurgy and Materials 56(4), 2011, 903–908, [http://doi.org/10.2478/v10172-011-0099-4].
[19] Mott N. F., Davies E. A.: Electronic process in non-crystalline materials. Claredon Press, Oxford 1979.
[20] Poklonskii N. A., Gorbachuk N. I.: Fundamentals of impedance Spectroscopy of composites. Belarusian State University, Minsk 2005.
[21] Svito I., Fedotov A. K., Kołtunowicz T. N., Żukowski P., Kalinin Y., Sitnikov A., Czarnacka K., Saad A.: Hopping of electron transport in granular Cux(SiO2)1–x nanocomposite films deposited by ion-beam sputtering. Journal of Alloys and Compounds 615, 2014, S371–S374, [http://doi.org/10.1016/j.jallcom.2014.01.136].
[22] Yang Yu.: Deposited mono-component Cu metallic glass: a molecular dynamics study Materials Today Communications 26, 2021, 102083-1–102083-5, [http://doi.org/10.1016/j.mtcomm.2021.102083].
[23] Yarimbiyik A. E., Schafft H. A., Allen R. A., Vaudin M. D., Zaghloul M. E.: Experimental and simulation studies of resistivity in nanoscale copper films, Microelectronics Reliability 42(2), 2009, 127–134, [http://doi.org/10.1016/j.microrel.2008.11.003].
[24] Zhigal'skii, G. P., Jones B. K.: The physical properties of thin metal films. Vol. 13. CRC Press, London 2003.
[25] Żukowski P., Kołtunowicz T. K., Partyka J., Węgierek P.: Dielectric properties and model of hopping conductivity of GaAs irradiated by H + ions, Vacuum 81(10), 2007, 1137–1140, [http://doi.org/10.1016/j.vacuum.2007.01.070].
[26] Żukowski P., Kołtunowicz T. N., Boiko O., Bondariev V., Czarnacka K., Fedotova J. A., Fedotov A. K., Svito I. A.: Impedance model of metal-dielectric nanocomposites produced by ion-beam sputtering in vacuum conditions and its experimental verification for thin films of (FeCoZr)x(PZT)(100−x), Vacuum 120, 2015, 37–43, [http://doi.org/10.1016/j.vacuum.2015.04.035].
[27] Żukowski P., Kołtunowicz T. N., Partyka J., Fedotova Yu. A., Larkin A. V.: Hopping conductivity of metal-dielectric nanocomposites produced by means of magnetron sputtering with the application of oxygen and argon ions. Vacuum 83(3), 2009, S280–S283, [http://doi.org/10.1016/j.vacuum.2009.01.082].

M.Sc. Eng. Aleksandra Wilczyńska
e-mail: aleksandra.wilczynska@pollub.edu.pl

Ph.D. student of Automation, Electronic and Electrical Engineering area at the Lublin University of Technology Doctoral School. She received Eng. Biomedical Engineering degrees at the Faculty of Mechanical Engineering and Master degrees of Mechatronics at the Faculty of Electrical Engineering and Computer Science.

http://orcid.org/0000-0002-5630-1078

M.Sc. Eng. Karolina Czarnacka
e-mail: karolina.czarnacka@up.lublin.pl

An employee as an assistant in the Department of Fundamentals of Technology at the University of Life Sciences in Lublin. Ph.D. student at the Lublin University of Technology in the discipline of Electrical Engineering.

http://orcid.org/0000-0003-1434-734X

Ph.D. Eng. Andrzej Kociubiński
e-mail: a.kociubinski@pollub.pl

He received the M.Sc. and Ph.D. degrees in Electronic Engineering from Warsaw University of Technology, Poland. In 2007 he joined the Lublin University of Technology where he was involved in research on semiconductor technology. His research interest is concentrated on semiconductor devices and technology. His recent activities are related to microsystems for biomedical applications.

http://orcid.org/0000-0002-0377-8243

D.Sc. Ph.D. Eng. Tomasz N. Kołtunowicz
e-mail: t.koltunowicz@pollub.pl

He received the M.Sc., Ph.D. and D.Sc. degrees in Electrical Engineering from Lublin University of Technology in 2004, 2011 and 2016 respectively. University Professor at the Department of Electrical Devices and High Voltage Technology, Faculty of Electrical Engineering and Computer Science, Lublin University of Technology, Lublin, Poland. Since 2018 Director of Lublin University of Technology Doctoral School, Lublin, Poland. Specialist in the field of electrical engineering, materials science (nanocomposites), physics and technology of condensed matter.

http://orcid.org/0000-0001-7480-4931

http://doi.org/10.35784/iapgos.2876 *received: 02.02.2022 | revised: 11.03.2022 | accepted: 27.03.2022 | available online: 31.03.2022*

NEURAL NETWORKS FROM KERAS IN SKIN LESION DIAGNOSTIC

Magdalena Michalska-Ciekańska
Lublin University of Technology, Department of Electronics and Information Technology, Lublin, Poland

Abstract. *Melanoma is currently one of the most dangerous skin diseases, in addition many others appear in the population. Scientists are developing techniques for early non-invasive skin lesions diagnosis from dermatoscopic images, for this purpose neural networks are increasingly used. Many tools are being developed to allow for faster implementation of the network, including the Keras package. The article presents selected methods of diagnosing skin diseases, including the process of classification, features selection, extracting the skin lesion from the whole image.The described methods have been implemented using deep neural networks available in the Keras package. The article draws attention to the effectiveness, specificity, accuracy of classification based on available data sets, attention was paid to tools that allow for more effective operation of algorithms.*

Keywords: dermatoscopic images, deep learning, melanoma, skin lesions, Keras

SIECI NEURONOWE Z KERAS W DIAGNOSTYCE ZMIAN SKÓRNYCH

Streszczenie. *Melanoma jest obecnie jedną z najbardziej niebezpiecznych chorób skóry, oprócz niej pojawia się w populacji wiele innych. Naukowcy rozwijają techniki wczesnego nieinwazyjnego diagnozowania zmian skórnych z obrazów dermatoskopowych, w tym celu coraz częściej wykorzystywane są sieci neuronowe. Powstaje wiele narzędzi powzalajcych na szybszą implementację sieci należy do niej pakiet Keras. W artykule przedstawiono wybrane metody diagnostyki chorób skóry, należy do nich proces klasyfikacji, selekcji cech, wyodrębnienia zmiany skórnej z całego obrazu. Opisane metody zostały zostały zaimplementowane za pomocą dostępnych w pakiecie Keras głębokich sieci neuronowych. W artykule zwrócono uwagę na skuteczność, specyficzność, dokładność klasyfikacji w oparciu o dostępne zestawy danych, zwrócono uwagę na narzędzi pozwalające na efektywniejsze działanie algorytmów.*

Słowa kluczowe: obrazy dermatoskopowe, uczenie głębokie, melanoma, zmiany skórne, Keras

Introduction

Skin diseases develop more and more often. The most dangerous of them is the rapidly developing malignat melanoma. In the process of diagnosing these diseases, doctors are helped by many other tools. Nowadays, diagnostic tools based on the use of artificial intelligence are being created [10, 12, 16, 36]. The fastest growing field is neural networks. They present in various ways extensive structures, they contain different counts of layers, weights, their operation is characterized by a different time needed to perform the whole process. Deep neural networks consist of many layers, each of which identifies more complex elements of the input image.

Neural networks allow for a number of actions. These include dermatoscopic images segmentation from different regions of the body, for this purpose very good are deep convolutional neural networks – DCNN [11, 19, 28]. The classification of skin lesions images is also important. A binary classification is distinguished, which applies to 2 classes, classification on malignat melanoma versus no-melanoma or begin nevi [6, 15, 22]. Multiclass classification gives the opportunity for greater interest in skin lesions diagnosis processes. The multi-stranded networks development has resulted in the development of its use. It is used for diagnosis based on network models that allow classification into 3, 5 and even 7 different disease [23, 24]. A more complex structure becomes more effective for data with greater diversity. Convolutional networks have allowed scientists to achieve great success in the features selection [1, 21]. In [34] based on the ResNet-50 based supervised deep learning networks gets extracted features from dermatoscopic images.

There are also many works that provide an overview of the available methods based on deep learning and machine learning [14, 26, 35]. The systems also combine multiple modalities to achieve the best possible effect [37]. Used frequently deep learning [8] approach based on CNN and recurrent neural network (RNN) [4]. Many works are based on deep learning [9, 36] and a Deep Residual Network (DRN), deep region based convolutional neural network (RCNN), fully convolutional neural network (FCN) and a specific convolutional neural network (CNN) in [4].

1. Segmentation and classification methods

Many studies have been conducted in melanoma classification from dermatoscopic images using deep learning and neural networks. To that date, there is no accurate data on the clinical use of artificial intelligence. There is a lack of a large data set to train and test algorithms [13]. Currently, the larger database is ISCI, it is a dataset for 7 different skin lesions. The correct diagnosis of skin cancer is not simple, many automated computer diagnostic systems based on deep neural network algorithms are being created.

Many scientists use the Keras library in their work with deep neural networks. It contains more than 20 models of neural networks, e.g. VGG16, VGG19 ResNet50, InceptionV3, MobileNet, InceptionResNetV2. Keras is implemented in Python. It allows you to define high-level blocks and train deep learning models, does not support low-level operations. It relies on a specialized and optimized tensor library (TensorFlow). Many convolutional neural network from Keras has been used for features extraction in many works [2, 3, 7]. They are characterized by different network complexity, number of layers, parameters and iteration time. They have a size from 88 to even 343 MB. Table 1 shows the models listed.

Table 1. Selected Keras networks models https://keras.io/api/applications/

Model	Size (MB)	Top-1 Accuracy	Top-5 Accuracy	Parameters	Depth	Time (ms) per inference step (CPU)	Time (ms) per inference step (GPU)
Xception	88	0.790	0.945	22,910,480	126	109.42	8.06
VGG16	528	0.713	0.901	138,357,544	23	69.50	4.16
VGG19	549	0.713	0.900	143,667,240	26	84.75	4.38
ResNet50	98	0.749	0.921	25,636,712	-	58.20	4.55
ResNet101	171	0.764	0.928	44,707,176	-	89.59	5.19
ResNet152	232	0.766	0.931	60,419,944	-	127.43	6.54
ResNet50V2	98	0.760	0.930	25,613,800	-	45.63	4.42
ResNet101V2	171	0.772	0.938	44,675,560	-	72.73	5.43
ResNet152V2	232	0.780	0.942	60,380,648	-	107.50	6.64
InceptionV3	92	0.779	0.937	23,851,784	159	42.25	6.86
InceptionResNetV2	215	0.803	0.953	55,873,736	572	130.19	10.02
MobileNet	16	0.704	0.895	4,253,864	88	22.60	3.44
MobileNetV2	14	0.713	0.901	3,538,984	88	25.90	3.83
DenseNet121	33	0.750	0.923	8,062,504	121	77.14	5.38
DenseNet169	57	0.762	0.932	14,307,880	169	96.40	6.28
DenseNet201	80	0.773	0.936	20,242,984	201	127.24	6.67
NASNetMobile	23	0.744	0.919	5,326,716	-	27.04	6.70
NASNetLarge	343	0.825	0.960	88,949,818	-	344.51	19.96
EfficientNetB0	29	-	-	5,330,571	-	46.00	4.91

In [30], several models of CNN neural network algorithms were used to determine their effectiveness in the diagnosis of several skin diseases and analysed their efficiency. Was used

Keras Sequential API and transfer learning model includes VGG11, RESNET50, DENSENET121, achieve highest accuracy of 90%. In [30], several models of CNN neural network algorithms were used to determine their effectiveness in several skin diseases diagnosis and analysed their efficiency. The solution is based on VGGNet and the transfer learning paradigm, a sensitivity value of 78.66%, which is much higher than the state of the current technic state in this dataset. Method achieves a sensitivity value of 78.66%.

2. Lesions classification process

The classification can be based on a set of data divided into 3, 6 or even 10 classes depending on how large the database is. Deep neural networks allow you to recognize even several skin diseases at the same time. More extensive classification systems allow you to assess more skin diseases causing the formation of various birthmarks. Figure 1 shows 8 pictures of selected skin diseases. The right number of these images allows you to evaluate as many as 8 different classes.

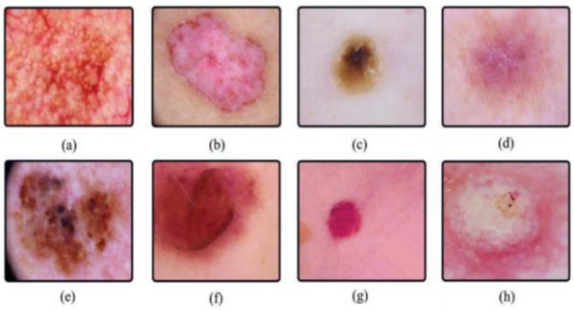


Fig. 1. Sample images of each class: (a) actinic keratoses, (b) basal cell carcinoma, (c) benign keratosis, (d) dermatofibroma, (e) melanoma, (f) melanocytic nervi, (g) vascular skin lesions, (h) squamous cell carcinoma [33]

Many works [6, 14, 15, 22] have been devoted to binary classification, which are based on extensive databases of dermatoscopic images confirmed by biopsy. Researchers use over 500 or even 1500 images. Figure 2 shows the steps of binary classification by the selected CNN network. The first step is to find databases that will contain as many cases as possible and good quality images without artifacts that prevent network diagnostics. The images are preprocessed, then the samples are cropped from the processed images. They are the ones who are classified. In earlier years, the most frequently chosen classification by scientists was binary. It concerns divisions into 2 classes, most often into malignat melanoma versus benign nevi (benign nonpigmented skin lesions, versus melanocytic nevi or atypical nevi) [14, 29].

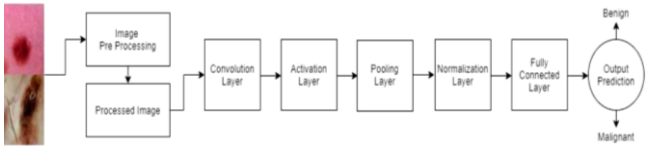


Fig. 2. CNN architecture from [29]used for binary classification

Batch normalization was used in the network architecture from [29]. It is reasonable to improve the performance, stability of the deep neural network, normalizes the input layer, removes parameters that reduce the performance of the model. Depending on the number of layers, the entered parameters, the classification process gives a different effect. In assessing the created model effectiveness, graphs of accuracy and loss values in relation to the number of epochs made by the network (figure 3) help. On the basis of the distribution of values for training and validation data, the algorithm or the type of data used is modified accordingly. It is important that the network is well trained, not retened, and that the data is well matched to the model.

In [33], neural networks from the Keras library were used to classify skin lesions: DenseNet-201, Inception-ResNet-V2 and Inception-V3. Figure 4 shows promising results and compares them. Inception-ResNet-V2 and Inception-V3 show similar results they have in terms of algorithm accuracy and losses. DenseNet-201 achieved the highest values of data validation efficiency, as well as the lowest loss values. Comparing several algorithms with each other allows you to find the best one, ensuring the highest efficiency.

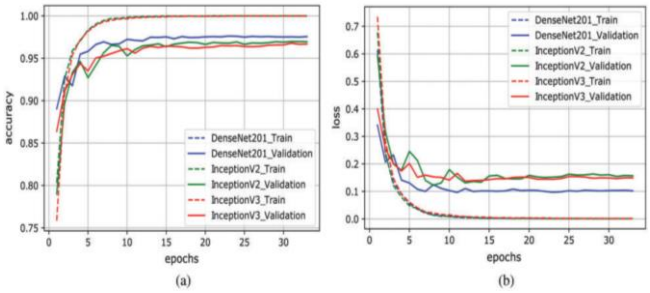


Fig. 3. Evolution of training and validation for (a) accuracy and (b) loss of the proposed model [33]

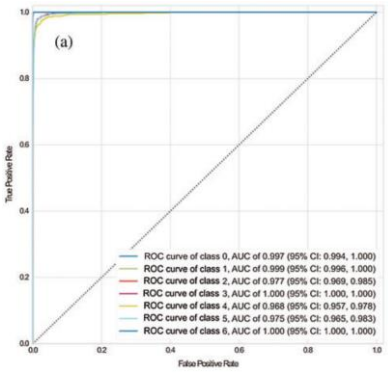


Fig. 4. ROC Curves with a confidence interval of 95% for 6 classes (0 'akiec' – actinic keratoses, 1 'bcc' – basal cell carcinoma, 2 'bkl' – benign keratosis, 3 'df' – dermatofibroma, 4 'mel' – Melanoma, 5 'nv' – melanocytic nervi, and 6 'vasc' – vascular skin lesions) [33]

The development of the ROC diagnostic curve gives a broader picture of the results obtained. The closer the curve to 1 in the TruePositive Rate, the higher the diagnostic capabilities of the algorithm. In [33] was obtained an average accuracy of 98% for all 6 classes with DenseNet-201. Classes 4 (melanoma disease) and 6 (Vascular skin lesions) were diagnosed with the highest success rate, reaching 1, and 0.999 for class 1 (basal cell carcinoma disease).

3. Features extraction

Many works are being created that allow for skin lesions diagnosis from dermatoscopic images. Researchers in the work on feature extration based on CNN architecture focus on geometry-based [20], color-based data augmentation [25] and ensemble of network architectures [17].

In [3] using binary classification, feature extraction was made. Two classes were considered, the first with images of melanoma and second colled „no melanoma" with unknow nad begin lesions. The results are presented in Figure 5. The authors were the first to combine RGB composition with the addition of texture information for relevant points. For this purpose, they generated key points with first-order statistical information, second-order statistical parameters. Then, the RGB components were analyzed to extract the features needed in the training set. A network with appropriately selected images was trained and classifications were made. The algorithm was based on finding points regardless of the location of the image.

In [21] a new function has been introduced – the difference between the maximum and minimum diameter of the Feret. It is assumed that it is the best suited ellipse to the shape of the skin lesion, using it melanoma was classified with an accuracy of up to 86.5%. The result of this step is a binary image of an segmented white skin lesion on a black skin background. To distinguish between malignant and benign melanoma, a neural network (BNN) back-propagation model was created. The results of the activities are presented in figure 6 Numerical values for various features AS1, AS2, B, C, D1, D2 have been quantified.

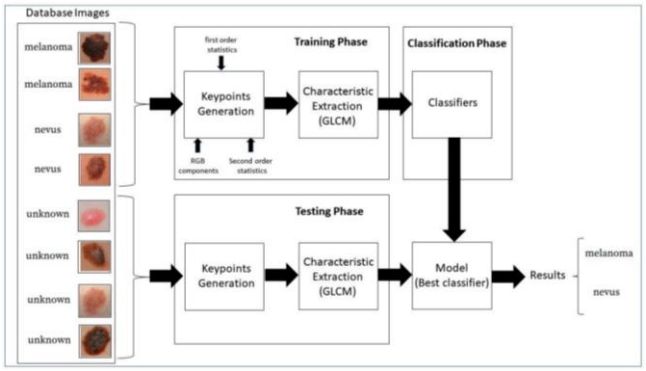


Fig. 5. Classification algorithm proposed in [3]

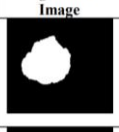
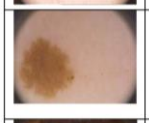
Original Image	Segmented Image	Features	Output
		*AS1*: 0.13 *AS2*: 0.13 *B*: 0.86 *C*: 2 *D1*: 4 *D2*: 0.96	Benign
		*AS1*: 0.16 *AS2*: 0.18 *B*: 0.86 *C*: 1 *D1*: 4.3 *D2*: 0.46	Benign
		*AS1*: 0.24 *AS2*: 0.18 *B*: 0.5 *C*: 4 *D1*: 8.6 *D2*: 1.5	Malignant

Fig. 6. Results from [21]

Scientists also try to modify the structure of the network in various ways. For feature extration in [18] was used Lesion Feature Network (LFN). The architecture of this network is shown in figure 7 Blue rectangles represent 12 convolution layers, the numbers on them indicate the size and number of kernels. The layers consist of 3×3 and 1×1 spots. 1×1 weave helps to integrate the functions created in 3×3. MP, AP, FC wool in the structure still other functions. Calculating softmax losses allows you to pay more attention to classes with fewer samples.

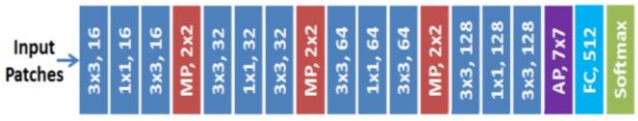


Fig. 7. Flowchart of Lesion Feature Network (LFN) [18]

4. Supporting functions and tools

Currently, a lot of work is aimed at tuning already pre-trained CNN networks with their known parameters. It is then used for well-prepared data, which is properly enlarged, important features are detailed. One should not forget about the sufficiently numerous classes, in an appropriate way divided into them images [5, 34]. Many face the problem of excessive do-fitting when working. One of the basic tools is a good preparation of input images, geometric transformations are made, images are scaled, colors are transformed. However, there are still a number of factors that affect the final result.

Beyond the structure of the neural network itself. An important element of the network is the optimizer, which determines the course of training. The loss function compares the results of network predictions with the target labels and calculates the loss value based on them. The Optimizer uses them when modifying network weights.

The process of teaching large networks can be time-consuming and require large computing resources. For this reason, it is worth choosing the right optimizer. The most frequently chosen are: momentum optimization, Nesterov algorithm (accelerated drop along the gradient), AdaGrad, RMSProp, Adam algorithm (adaptive moment estimation.

In [2], texture analysis was performed using the CNN model. Classification was made using appropriate labels in 16 classes from the Kylberg Texture dataset. In [2], texture analysis was performed using the CNN model. Classification was made using appropriate labels in 16 classes from the Kylberg Texture dataset. One of the most well-known optimizers is Adam [7], who learns very quickly, has no problems with fading learning indicators, leading to fluctuations in the function of loss.

In [32] the Max Pooling Layer was placed in the network structure. It causes a reduction along the weight and length, it reduces the matrix significantly, this can be seen in figure 8, which results in a smaller numerical representation of the parameters. With the help of the MAX function, the network is spatially reduced. In [16] a Max-Pool of 2×2 size was used, from 4 adjacent fields 1 – figure 8 was obtained.

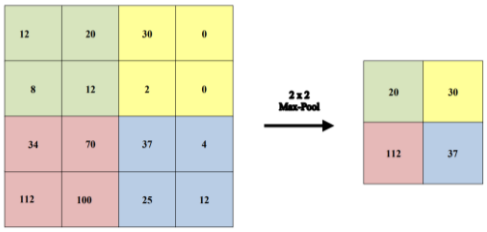


Fig. 8. Examples of how max pooling operates [14]

In the classification processes, other methods are also used to assess the effectiveness of neural networks and increase its effectiveness. Various comparisons of selected methods of melanoma classification are also made. Stolz's conventional method [27] was compared with the result of algorithms from [32]. The research was carried out on 700 samples. The STOLZ method achieved an accuracy of 76.6%, while the proposed one in [32] achieved 96.67%. Due to the emerging artifacts in the images and not very clear boundaries of changes, the STOLZ method has a lower efficiency in the extraction of skin lesion boundaries. Therefore, it incorrectly classifies the areas around the border.

5. Discussion and conclusions

Melanoma is now increasingly diagnosed skin disease. It is important to detect it quickly and fight the disease at an early stage. The development of non-invasive early diagnosis techniques is an issue of interest to many research teams. Neural networks are now becoming the fastest growing field of medical informatics. The use of methods based on various tools allows to increase the effectiveness of diagnosis of melanoma and other skin diseases using deep neural networks.

References

[1] Adegun A., Viriri S.: Deep learning-based system for automatic melanoma detection. IEEE Access 8, 2020, 7160–7172.
[2] Aggarwal A., Kumar M.: Image surface texture analysis and classification using deep learning, Multimedia Tools and Applications 80, 2021, 1289–1309.
[3] Almeida M., Santos I.: Classification models for skin tumor detection using texture analysis in medical images, J. Imaging 6(51), 2020, [http://doi.org/10.3390/jimaging6060051].

[4] Attia M., Hossny M., Nahavandi S., Yazdabadi A.: Skin melanoma segmentation using recurrent and convolutional neural networks. Biomedical Imaging (ISBI 2017), IEEE 14th International Symposium, 2017, 292–296.
[5] Barata C., Celebi M., Marques J.: A survey of feature extraction in dermoscopy image analysis of skin cancer. IEEE Journal of Biomedical and Health Informatics 23(3), 2019, 1096–1109.
[6] Brinker T. J., Hekler A., Enk A. H., Klode J., Hauschild A., Berking C.: Deep learning outperformed 136 of 157 dermatologists in a head-to-head der moscopic melanoma image classification task. Eur J Cancer 113, 2019, 47–54.
[7] Brownlee J.: Gentle introduction to the adam optimization algorithm for deep learning. Machine Learning Mastery Pty. Ltd., 2019, https://machinelearningmastery.com/adamoptimization-algorithm-for-deep-learning/ss
[8] Codella N., Cai J., Abedini M., Garnavi R., Halpern A., Smith J. R.: Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images. International Workshop on Machine Learning in Medical Imaging, 2015, 118–126.
[9] Codella N., Nguyen Q., Pankanti S., Gutman D., Helba B., Halpern A., Smith J.: Deep learning ensembles for melanoma recognition in dermoscopy images. arXiv preprint arXiv:1610.04662, 2016.
[10] Esteva A.: Dermatologist-level classification of skin cancer with deep neural networks. Nat. Res. 542(7639), 2017, 115–118.
[11] Ge Y., Li B., Zhao Y., Guan E., Yan W.: Melanoma seg-mentation and classification in clinical images using deep learning, ICMLC 2018: Proceedings of the 2018 10th International Conference on Machine Learning and Computing, 2018, 252–256.
[12] Ge Z., Demyanov S., Chakravorty R., Bowling A., Garnavi R.: Skin disease recognition using deep saliency features and multimodal learning of dermoscopy and clinical images. Descoteaux M., Maier-Hein L., Franz A., Jannin P., Collins D. L., Duchesne S. (eds.), Springer, Cham LNCS 10435, 250–258, 2017.
[13] Gupta A., Thakur S., Rana A.: Study of Melanoma Detection and Classification Techniques. 8th International Conference on Reliability, Infocom Technologies and Optimization, 2020 1345–1350, [http://doi.org/10.1109/ICRITO48877.2020.9197820].
[14] Haenssle H. A., Fink C., Schneiderbauer R., Toberer F., Buhl T., Blum A.: Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Ann Oncol 29, 2018, 1836–1342.
[15] Hekler A., Utikal J. S., Enk A. H., Solass W., Schmitt M., Klode J.: Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images. Eur J Cancer 118, 2019, 91–96.
[16] Hijazi S., Kumar R., Rowen C.: Using convolutional neural networks for image recognition. Cadence Design Systems Inc., San Jose 2015.
[17] Katapadi A. B.: Evolving strategies for the development and evaluation of a computerised melanoma image analysis system. Comput.Methods Biomech. Biomed. Eng., Imag Visual. 6, 2018, 465–472.
[18] Li Y., Shen L.: Skin Lesion Analysis towards Melanoma Detec-tion Using Deep Learning Network. arXiv.org > cs > arXiv:1703.00577, Computer Vision and Pattern Recognition 2017 (v2).
[19] Lopez A. R., Giro-i-Nieto X., Burdick J., Marques O.: Skin lesion classification from dermatoscopic images using deep learning techniques, [http://doi.org/JO.23J6/P.20l7.852-053].
[20] Lopez A. R.: Skin lesion classification from dermoscopic images using deep learning techniques. Proc. 13th IASTED Int. Conf. Biomed. Eng. 2017, 49–54.
[21] Majumder S., Ahsan Ullah M.: Feature extraction from der-moscopy images for an effective diagnosis of melanoma skin cancer. 10th International Conference on Electrical and Compu-ter Engineering Bangladesh, 2018, 185–188.
[22] Marchetti M. A., Codella N. C., Dusza S. W., Gutman D. A., Helba B., Kalloo A.: Results of the 2016 international skin imaging collaboration international symposium on biomedical imaging challenge: comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images. J Am Acad Dermatol 78, 2018, 270–277.
[23] Marchetti M. A., Liopyris K., Dusza S. W., Codella N. C. F., Gutman D. A., Helba B.: Computer algorithms show potential for improving dermatologists' accuracy to diagnose cutaneous melanoma: results of the international skin imaging collaboration 2017. J Am Acad Dermatol 82, 2020, 622–627.
[24] Maron R. C., Weichenthal M., Utikal J. S., Hekler A., Berking C., Hauschild A.: Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks. Eur J Cancer 119, 2019, 57–65.
[25] Mendoza C. S., Serrano C., Acha B.: Scale invariant descriptors inpattern analysis of melanocytic lesions. Proc. IEEE 16th Int. Conf. Image Process., 2009, 4193–4196.
[26] Murphree D. H., Puri P., Shamim H., Bezalel S. A., Drage L. A., Wang M., Pittelkow M. R., Carter R. E., Davis M., Bridges A., Mangold A., Yiannias J., Tollefson M., Lehman J., Meves A., Otley C., Sokumbi O., Hall M., Comfere N.: Deep learning for dermatologists: Part I. J Am Acad Dermatol, 1–9, 2020.
[27] Nachbar F., Stolz W., Merkle T., Cognetta A., Vogt T., Landthaler M.: The abcd rule of dermatoscopy. High prospective value in the diagnosis of doubtful melanocytic skin lesions. Journal of the American Academy of Dermatology 30(4), 1994, 551–559.
[28] Nida N., Irtaza A., Javed A., Yousaf M., Mahmood M.: Melanoma lesion detection and segmentation using deep region based convolutional neural network and fuzzy C-means clustering. International Journal of Medical Informatics 124, 2019, 37–48.
[29] Panja A., Jackson J. Ch., Quadir Md. A.: An Approach to Skin Cancer Detection Using Keras and Tensorflow. Journal of Physics: Conference Series 1911 012032, 2021, [http://doi.org/10.1088/1742-6596/1911/1/012032].
[30] Rahi M., Khan F., Mahtab M., Amanat Ullah A., Alam M. G., Alam M.: Detection Of Skin Cancer Using Deep Neural Networks, IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), 2019, 1–7, [http://doi.org/10.1109/CSDE48274.2019.9162400].
[31] Romero Lopez A., Xiro-i-Nieto X., Burdick J., Marques O.: Skin lesion classification from dermoscopic images using deep learning techniques. 13th IASTED International Conference on Biomedical Engineering (BioMed), 49–54, 2017, [http://doi.org/10.2316/P.2017.852-053]
[32] Sherif F., Mohamed W. A., Mohra A. S.: Skin lesion analysis toward melanoma detection using deep learning techniques. INTL Journal of Electronics and Telecommunications 65(4), 2019, 597–602.
[33] Villa-Pulgarin J., Ruales-Torres A., Arias-Garzón D., Bravo-Ortiz M., Arteaga-Arteaga H., Mora-Rubio A., Alzate-Grisales J., Mercado-Ruiz E., Hassaballah M., Orozco-Arias S., Cardona-Morales O., Tabares-Soto R.: Optimized Convolutional Neural Network Models for Skin Lesion Classification. Computers, Materials & Continua Tech Science Press, CMC 70(2), 2022, [http://doi.org/10.32604/cmc.2022.019529].
[34] Wang Y., Cai J., Louie D., Wang J., Lee T.: Incorporating clinical knowledge with constrained classifier chain into a multimodal deep network for melanoma detection. Computers in Biology and Medicine 137, 2021, 104812.
[35] Young A. T., Xiong M., Pfau J., Keiser M. J, Wei M.L.: Artificial intelligence in dermatology: A Primer. Journal of Investigative Dermatology 140, 2020, 1504–1512.
[36] Yu L., Chen H., Dou Q., Qin J., Heng P. A.: Automated melanoma recognition in dermoscopy images via very deep residual networks. IEEE Trans. Med. Imaging 36(4), 2017, 994–1004.
[37] Zhang J., Xie Y., Wu Q., Xia Y.: Skin lesion classification in dermoscopy images using synergic deep learning, Springer Nature Switzerland. LNCS 11071, 2018, 12–20.

M.Sc. Magdalena Michalska-Ciekańska
e-mail: magdalena.michalska@pollub.edu.pl

Ph.D. student at Department of Electronics and Information Technology, Lublin University of Technology. Recent graduate Warsaw University of Technology The Faculty Electronics and Information Technology. Her research interests include medical image processing, optoelectronics, spectrophotometry.

http://orcid.org/0000-0002-0874-3285

http://doi.org/10.35784/iapgos.2878 *received: 06.02.2022 | revised: 11.03.2022 | accepted: 28.03.2022 | available online: 31.03.2022*

DETERMINING THE WORKING LENGTH OF A ROOT CANAL USING INTRAORAL RADIOGRAPHY SEGMENTATION

Oleksii Perepelytsia, Tetyana Nosova
Kharkiv National University of Radio Electronics, Faculty of Biomedical Engineering, Kharkiv, Ukraine

Abstract. *Apexlocators are devices that determine the working length of the root canal. Measurement with the help of alternating currents of different frequencies in combination with the method of the ratio allows to achieve high accuracy in determining the location of the physiological tip of the root (over 80%). However, such devices cannot accurately visualize where a doctor's tool is. Using image processing algorithms, you can segment tooth root structures with high precision. Therefore, the combination of electronic determination of the root canal length (using apexlocator) with X-ray segmentation is relevant.*

Keywords: root canal, radiograph, endodontic treatment, apexlocator

OKREŚLANIE DŁUGOŚCI ROBOCZEJ KANAŁU KORZENIOWEGO ZA POMOCĄ SEGMENTACJI RADIOGRAFII WEWNĄTRZUSTNEJ

Streszczenie. *Apex locatory są urządzeniami, które określają długość roboczą kanału korzeniowego. Pomiar za pomocą prądów zmiennych o różnych częstotliwościach w połączeniu z metodą proporcji pozwala na osiągnięcie wysokiej dokładności w określaniu położenia fizjologicznego wierzchołka korzenia (ponad 80%). Urządzenia te nie są jednak w stanie dokładnie zobrazować miejsca, w którym znajduje się narzędzie lekarza. Wykorzystując algorytmy przetwarzania obrazu, można z dużą dokładnością segmentować struktury korzeni zębów. Dlatego też połączenie elektronicznego określania długości kanału korzeniowego (przy użyciu apex locatora) z segmentacją rentgenowską ma istotne znaczenie.*

Słowa kluczowe: kanał korzeniowy, radiogram, leczenie endodontyczne, apex locator

Introduction

The basis of successful endodontic treatment is the correct determination of the working length of the root canal (the distance between the external landmark on the crown of the tooth to the apical border). An apical constriction zone is recommended as a border for root canal treatment and filling. Intraoral radiograph allows you to obtain information about the direction of bending of the root canals, as well as to determine the working length. However, the radiograph is a two-dimensional total image and does not reproduce the entire anatomy of the apical part of the root therefore there are often layers and distortions of the image. When interpreting radiographs, there is a probability of error associated with the subjectivity of the evaluation result of the specialist. Thus, it is impractical to be guided exclusively by this method of determining the working length. The method of apexlocation is based on the difference of electrical resistance of tissues. The hard tissues of the tooth have a higher resistance than the mucous membrane of the mouth and periodontal tissue. Devices for electrometric determination of the working length of the root canal determine the impedance using alternating currents of different frequencies and apply the method of ratio. This measurement is stable and accurate even when working in wet channels and provides smooth visualization of all process of penetration of a top of the channel tool and high accuracy of definition of a place of physiological top of a root. Modern algorithms for electrometric determination of the working length of the root canal do not combine the data obtained from the radiograph. In this regard, it is important to develop new methods and means of displaying electrometric data on the radiograph to more accurately determine the location of the physiological apex of the root.

1. Electrometric method for determining the length of the root canal

There are 3 zones in the structure of the apex: the apex itself (radiological top of the root), large apical hole and apical constriction (the area of the apical part with the smallest diameter). The apical constriction zone is recommended as a limit for root canal treatment [8, 9]. The method of apexlocation is based on the difference in electrical resistance of tissues. The hard tissues of the tooth have a higher resistance than the mucous membrane of the mouth and periodontal tissue. The electrical circuit between the electrodes placed on the lip (lip electrode) and in the channel (electrode in the form of a file with a calibration stopper) remains closed until the periodontal tissue reaches the file. A sharp drop in resistance occurs in the area of apical narrowing, the circuit closes, this is what fixes the apexlocator [11].

Having used the mathematical method of determining the length of the canal (using tables), doctors began endodontic intervention. Knowing the range of tooth length is important factor in a successful operation.

Working length is the distance from the most protruding part of the tooth crown to the physiological narrowing – apical constriction. Apical constriction often has a complex configuration. The physiological tip is located at a distance of 1.0 mm from the anatomical – this is the final working length for the doctor. For an example, see the figure below.

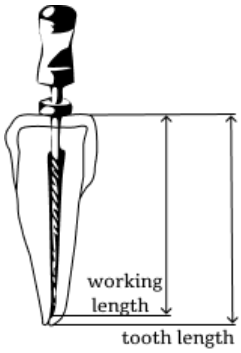


Fig. 1. Determination of working length

Devices for electrometric determination of the working length of the root canal determine the impedance using alternating currents of different frequencies. Using the ratio method, they allow you to find the total coefficient of resistance, which reflects the position of the file in the channel (Fig. 2).

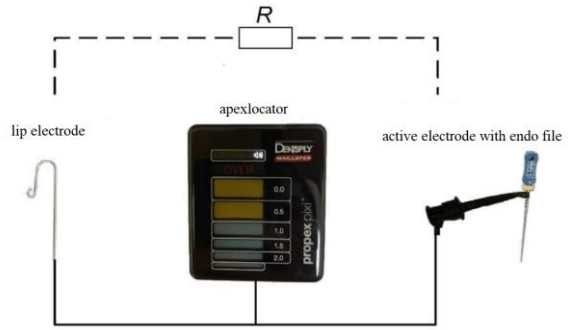


Fig. 2. Scheme of the apexlocator

The measurement scheme is shown in Fig. 2. This measurement indicates the presence of electrolytes in the pulp tissue, is stable and provides high measurement accuracy [10]. A significant disadvantage of the method is the requirement to work in relatively dry or partially dried channels.

The method of electrometric determination of the length of the root canal was performed in two stages. The first stage – with the help of the Propex Pixi device the previous working length was determined. In the process of processing the root canal is the expansion of the diameter and taper. The second stage – after the final treatment of the channel was performed another measurement by electrometric method. This value was defined as the final working length of the channel.

2. Method of radiological determination of root canal length

Monochrome photometric interpretation is used for dental radiographs (brightness of image pixels is represented by a gray scale with pointers from 0 to 255, where the brightness value of 0 corresponds to a pixel with black colour, the value of 255 – white).

The study used data obtained using a visiograph Planmeca ProSensor HD, which has a resolution of 1020 × 688 pixels. Matlab 2019 Image Processing Toolbox was chosen to develop software for segmentation and channel length measurement [5].

The threshold method was used to select the object of study (tooth root). The threshold method is a binarization method based on dividing an image into 2 parts based on threshold values. The value (T) is selected according to the task to be performed (from 0 to 255). All brightness values that are in the range of values higher than T are called object values, everything below is the background value. Next is the boundary layer – a curved line that separates the elements of the object and the background. A segment is selected along this line if it satisfies the low-pass noise filtering condition. Figures 3–5 show histograms.

This graphs of the distribution of image elements with different brightness, in which the horizontal axis shows the brightness from 0 to 255, and the vertical – the number of pixels with a specific value of the brightness of the corresponding images of teeth.

Physiologically sealed root canal corresponds to intervals with large brightness indicators. To select these areas, it is enough to select the value of T and determine all points that have $f(x, y) > T$, which belong to the object, and otherwise – belong to the background [14]. Then the original image (g) is defined by the following expression:

$$g(x,y) = \left\{ \frac{1, if\ f(x,y) > T}{0, if\ f(x,y) \le T} \right\}. \quad (1)$$

here *1* – the value of the object; *0* – background value.

The threshold T = 210 was chosen to binarize the image of the tooth root. Given the resolution, filtering with the removal of segments with a pixel count below 30,000 was used to exclude binarization artefacts [2, 13]. The result of binarization is shown in Fig. 6.

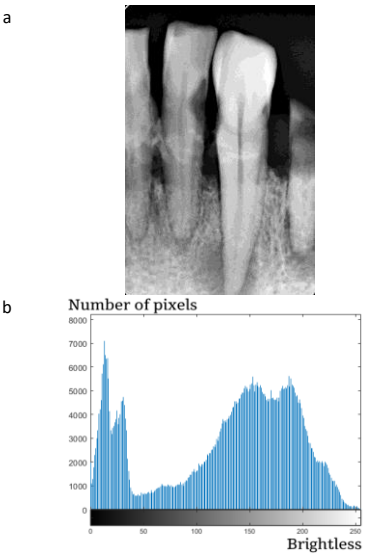


Fig. 3. Histogram analysis of the image: a — image of the tooth No 1, b — the corresponding histogram

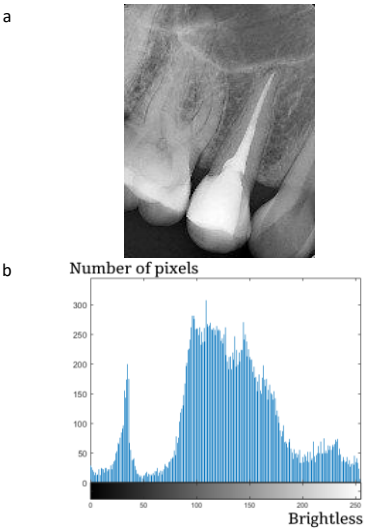


Fig. 4. Histogram analysis of the image: a — image of the tooth No 2, b — the corresponding histogram

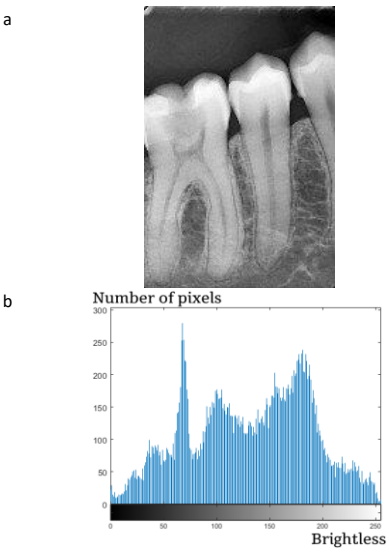


Fig. 5. Histogram analysis of the image: a — image of the tooth No 3, b — the corresponding histogram

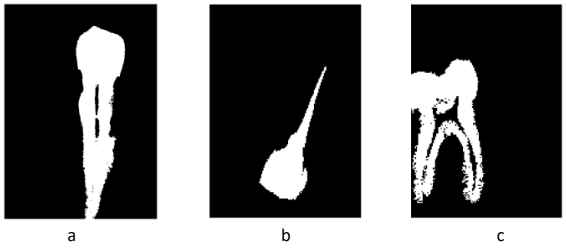


Fig. 6. The result of binarization of images with a threshold T = 210: a — segmented tooth No 1, b — segmented tooth No 2, c — segmented tooth No 3

At threshold segmentation it is necessary to consider the connection of components. If we consider two points of the image connected and there is a path between them, along which the characteristic function is constant, the points are connected. Marking objects on a discrete binary image is to select the point of the object from which the growth actually begins. The next step indicates the neighbouring points (except those already marked) and so on [3]. Upon completion of this recursive procedure, we obtain a closed loop (Fig. 7).

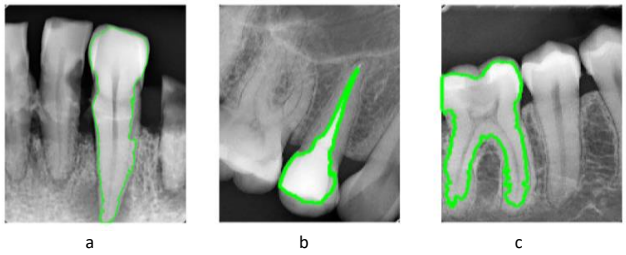


Fig. 7. The result of drawing segmented contours of the root of the teeth on the initial image: a — tooth No 1, b— tooth No 2, c — tooth No 3

3. Comparative characteristics of two methods

As a result of the work performed, the structures of the root canals of the tooth were segmented and their length was determined. Comparisons (Table 1) of the electronic determination of the working length with the radiological one showed that the electronic length and the radiological length determined by means of the program do not coincide. With lateral curvature of the canal, the X-ray may show a shorter working length than apexlocation devices [1, 15], and there is a possibility of incorrect segmentation of the tooth crown due to low brightness of the crown pixels and, consequently, the crown is not taken into account.

Table 1. Comparison of measurements performed using different methods of determining the length of the channel

Object	Mathematical method (average) according to J.I. Ingle, L.K. Buckland, J. Baumgartner [12]	Electrometric method	X-ray method with segmentation
tooth No1	22.4 mm	22.3 mm	21.64 mm
tooth No2	21.0 mm	21.5 mm	18.98 mm
tooth No3	20.9 mm	20.8 mm	20.31 mm

4. Conclusions

The advantage of measuring the length of the root canal with an apex locator is much greater accuracy (about 0.5 mm) compared to the method of radiography, but a combination of these two methods may be more reliable, which requires further statistical research. Particular attention should be paid to the peculiarities of methods of processing and segmentation of the obtained diagnostic images to ensure maximum quality of visualisation of the contours of the root canals. In segmentation, the main condition is the correct choice of the binary limit value. To do this, it is necessary to perform histogram analysis of the obtained images and in the process of post-processing to filter local artefacts using morphological operations.

References

[1] Avrunin O. G. et al.: Modern intelligent technologies of functional medical diagnostics. Press of the Kharkiv National University of Radioelectronics, Kharkiv 2018 [http://doi.org/10.30837/978-966-659-234-0].
[2] Avrunin O. G. et al.: Using a priori data for segmentation anatomical structures of the brain. Przegląd Elektrotechniczny 1(5), 2017, 102–105 [http://doi.org/10.15199/48.2017.05.20].
[3] Avrunin O. G.: Visualization of the upper respiratory tract according to computed tomography. Radio electronics and informatics 4, 2007, 119–122.
[4] http://www.dentsplysirona.com/content/dam/dentsply/pim/manufacturer/Endodontics/Motors__Apex_Locators/Apex_Locators/Propex_Pixi/PROPEX%20PIXI%20EUROP_DFU_1018_MASTER_DSE_EN.pdf (available: 10.01.2022)
[5] http://www.mathworks.com/products/image.html (available: 10.01.2022)
[6] Ingle J., Bakland L., Baumgartner J.: Endodontics. BC Decker, Hamilton, Ontario, Lewiston, NY 2008.
[7] Khomenko L. A., Bidenko N. V.: Practical endodontics. Tools, materials and methods. Kniga Plus, Moscow 2005.
[8] Kovetskaya E. E.: Comparative evaluation of the effectiveness of methods for determining the working length of the root canal. Modern dentistry 4, 2006, 11–13.
[9] Kovetskaya E. E.: Methods for determining the working length of the root canal. Modern dentistry 3, 2006, 35–39.
[10] Krainov S. V., Popova A. N., Firsova I. V.: Evaluation of the effectiveness of the electrometric method for determining the working length of the root canal using the example of the NovApex apex locator. Proceedings of the Conference Topical issues of modern dentistry 2010.
[11] Latysheva S. V., Abaimova O. I., Bondarik E. A.: Basic principles of endodontic tooth preparation. Dental journal 2, 2003, 2–6.
[12] Mounce R.: Determination of the true working length. Journal of Endodontics. 43(1), 2007, 17–19.
[13] Shamraeva E. O., Avrunin O. G.: Choice of a method for segmentation of bone structures on tomographic images. Bionics of intelligence: information, language, intelligence 65, 2006, 83–87.
[14] Shamraeva E. O., Avrunin O. G.: Construction of models of cranial implants based on radiographic data. Applied radio electronics 4(4), 2005, 441–443.
[15] Shchapov P. F., Avrunin O. G.: Obtaining information redundancy in measuring control systems and diagnostics of measuring objects. Ukrainian metrological journal 1, 2011, 47–50.
[16] Silva G., Oliveira L., Pithon M.: Automatic segmenting teeth in X-ray images: Trends, a novel data set, benchmarking and future perspectives. Expert Systems with Applications 107, 2018, 15–31 [http://doi.org/10.1016/j.eswa.2018.04.001].
[17] Solovyova A. M.: Features of conservative endodontic treatment for chronic periodontitis in teeth with incomplete root formation. Children's dentistry (Pediatric dentistry and dental profilaxis) 1–2, 2000, 79–83.

M.Sc. Oleksii Perepelytsia
e-mail: oleksii.perepelytsia@nure.ua

A participant of numerous exhibitions and scientific conferences. Received bachelor's and master's degrees with honours. Was a member of the Erasmus+ student exchange program with a stay at the Leibniz University in Hannover, Germany.

http://orcid.org/0000-0001-6850-2013

Prof. Tetyana Nosova
e-mail: oleksii.perepelytsia@nure.ua.

Associate Professor of the Department of Biomedical Engineering, Scientific Secretary of Specialized Scientific Council, Member of STC, Deputy Head of the Section 4 of STC, Doctor of Phylosophy. Scientific interests: development of methods of processing and analysis of biomedical signals.

http://orcid.org/0000-0003-4442-8001