Technology of multidimensional contextual analysis using the Internet of Things, data with a large volume of Big Data as well as cloud and mobile solutions

Mariusz CZERWIŃSKI
Globedata Sp. z o.o., Toruń, Poland

mariusz.czerwinski@globedata.pl

Piotr MATUSZAK
Globedata Sp. z o.o., Toruń, Poland

piotr.matuszak@globedata.pl

Abstract

The presented project was implemented based on the following research areas: extraction of data of analytical significance from data sets registered by means of two-way BLE radio communication, a digital user profile built dynamically on the basis of data analysis from many sources, management and processing of big data in cloud resources, advanced analysis of profile data using classification, statistical and machine learning methods. As a result of the research and development work carried out, the technology of multidimensional contextual analysis using the Internet of Things, data with a large volume of Big Data as well as cloud and mobile solutions was created and implemented. This technology will allow in the near future to provide the market with new components and services related to the exchange of digital profiles of moving terminals, allowing for calculations and predictions helpful in the analysis of tourist data, matching offers for travelers and many other services, mainly from the SME sector. The use of systems using the BLE standard allows for the location of terminals in places normally inaccessible for classic GPS positioning, such as shop halls, shopping malls, public facilities, large hotels, historic churches, undergrounds, tunnels, etc. Moreover, it allows for significant energy savings and extending the operation of the device from the battery. The author's team applied approaches based on programmable FPGA gate systems, thanks to which it will be possible to optimize the functionality of future products in terms of market needs and significantly reduce their production costs.

Keywords: *Internet of Things, Big Data, mobile solutions, FPGA.*

1. Introduction

The implemented project concerned the development of a new technology of multidimensional contextual analysis using the Internet of Things, data with a large volume of Big Data as well as cloud and mobile solutions. Currently, many mobile devices communicate by exchanging information on various data transmission channels. An example may be the transfer of data to and from mobile devices via the cloud, i.e. distributed internet resources made available via GSM bands. Nevertheless, nowadays, the use of permanently switched on GPS positioning modules in terminals as well as radio data transmission causes the battery to quickly deplete, and thus to a reduced practical utility of these solutions. Therefore, a large number of users disable or block the positioning and data transmission by terminals when they absolutely do not need them. The exception to this rule is the Bluetooth Low Energy (BLE) transmission, which is characterized by a very low power consumption from version 4.0, a relatively long range (up to 100m in the field without obstacles), and a reasonable transmission speed of 1Mbps. Moreover, unlike GPS positioning, BLE technology does not require precise tracking of transmitters (in the case of GPS, these are satellites in orbit around the earth). The terminal can be in a pocket or bag, and if the materials are not HF attenuating (i.e. not manufactured exactly to attenuate them), the terminal will operate without problems. This effect is used today in the so-called BLE beacons, which position the user in objects where the GPS signal is not spread. An example may be shop halls, shopping malls, public utility facilities, large hotels, historic churches, undergrounds, tunnels, etc. Beacons are designed to allow the terminal (or more precisely, a mobile application) to determine the position, and secondly to give it a small portion information such as the ID of the beacon and a few - a dozen bytes to be used for additional marking of the place, e.g. determining whether it is a commercial, utility or historic place. Nevertheless, it turned out that these types of devices do not meet the many needs of users using mobile applications in tourism or while shopping. One-way communication does not allow for the full interaction of beacons and terminals that is currently expected.

The very limited data exchange makes it impossible, for example, to leave characteristic information about the user who was in the range of a given beacon at a certain moment. Hence, manufacturers of BLE devices are currently introducing the so-called smart beacons that can use two-way communication. This allows, for example, user authorization and access to limited or paid services, making micropayments or full use of promotional programs. The pace of the appearance of new BLE devices and mobile applications, as well as the popularity of smartphone or tablet terminals (and soon also smartwatches, smartbands and other devices from the Internet of Things category incorporated into e.g. clothes - wearable) causes a significantly combinatorial explosion of information circulating in highly dispersed networks . This causes two major problems. Firstly, it is difficult to control the standardization of data exchange processes, especially since there are many manufacturers of BLE solutions and terminals in the world. Secondly, there is a big problem with the analysis of such unstructured data belonging to the category of big data, i.e. large-volume data. For example, if the device records one terminal every

1 second for 10 hours, and the data exchange package is only 10kB, after a month we will get a data set of 10.8GB. A small network of 1000 BLE devices will already give 10.8TB, which already exceeds the analytical capabilities of many systems, including advanced commercial systems. Hence, the author noticed a dynamically growing need to provide the market with an appropriate environment that would implement advanced analytical processes based on large data sets. In order to save transmission and to optimize the amount of data transferred in relation to the capacity used, the author's team developed a methodology for automatically building and updating the electronic user profile (in bit form). The innovation of the profile is that it allows you to send the so-called BLE (intelligent beacon) to the BLE device. digital imprint, i.e. only parts of data that are contextually related to a given beacon (e.g. due to its location, role). Thanks to the author's solutions, it will be possible to massively analyze the traffic of users using mobile terminals, which will be used in many areas of life now and in the future, such as tourism, shopping, local information, promotional campaigns, short messages, monitoring of basic life parameters, etc.

The solution's operation can be compared to a basket analysis, sometimes called a receipt analysis. This approach is practically used in all stores with a sufficiently wide range of products. Each customer, when approaching the checkout and paying for the purchased goods, automatically transfers the data on what was in his basket, i.e. what products were purchased together. Thanks to the basket analysis, it is possible to answer many different, significant questions, such as: what products are bought together most often?, what is the distribution of the intensity of purchase of a given product on a daily, weekly, monthly, annual, seasonal basis?, is there a correlation between individual products?, does the sale of the products correlate with a specific time?, are there any autocorrelations in sales?, what is the statistical character of the distribution of sales of a given product over time? and many more. A practically correctly conducted basket analysis allows for such selection of the assortment and its distribution in a given store so that the value of sales is subject to specific parameterization, e.g. profit optimization, minimization of storage time, optimization of the supply chain, etc. Research of additional data, e.g. anonymous customer profiles, strengthens the effectiveness basket analysis. The area of this type of research in relation to medium and large-format stores is very developed and improved for many years. On the other hand, the author noted that this type of research is not carried out, or is carried out only in a very limited form, in relation to broadly understood tourism. The main limitation here is the lack of relevant data left by tourists who are usually not interested in appearing in the research. Moreover, the scope and type of data, especially unstructured data, is downright overwhelming and rarely any mechanism of their collection and processing can handle it. Appropriate marriages between distributed devices (agents) are required, e.g. in the form of the Internet of Things and mobile terminals, as well as efficient processing of large-volume (big data) sets in the cloud, preferably in learning networks [1].

Both of these factors effectively limit the ability to analyze the traffic of tens of millions of people. There were about 6.2 million Polish tourists traveling around the country in the first half of 2019 [2]. There were almost 10 million foreign tourists during this period [2]. This means over 90 thousand active authors of various analytical events daily. If each of them undertakes dozens of different activities daily related to stay and travel, as well as between them and the environment and interactions between them, a complex network of huge amounts of data is created, with almost no analytical structure, which cannot be analyzed without specific, advanced tools and methodology. On the other hand, the value of this type of data and analyzes is significantly high, as their use is of interest to many economic entities providing tourist and tourist services, as well as sellers of various types of goods (the tourist population is almost equal to the population of the entire country). Thanks to the results of the analyzes, they can better select their offers to meet the dynamically changing needs of travelers. Additionally, thanks to the Globedata system based on FPGA [3] [4] [5], they will be able to more precisely reach specific target groups with their offer. All this will translate not only into better service of tourist traffic in Poland and abroad, but will also result in optimization of related profits, generated by local economic entities, especially from the SME sector. Hence, the author noticed significant needs in the so far undeveloped analytical niche, and thus intends in the near future, as a result of the implementation of the project, to introduce advanced and innovative tools to the market to support it.

2. Research area

As part of the project, research and development work was carried out in the following research areas:

- Extraction of data of analytical significance from data sets registered by means of two-way BLE radio communication.
- A digital user profile built dynamically on the basis of data analysis from many sources.
- Management and processing of big data in cloud resources.
- Advanced analysis of profile data using classification, statistical and machine learning methods.

The research areas presented above were related to the implementation of the following technological issues:

- A methodology for collecting and basic processing of big data.
- Methodology of data processing by distributed mechanisms.
- Methodology of data extraction of analytical significance.
- Methodology of building bit markers used to carry profile data.
- A methodology for encoding the information carried through the digital user profile.
- A methodology for dynamically changing the digital user profile as a result of informational feedback loops.
- A methodology for distributing part or all of the fingerprints of user profiles on an IoT device network.
- Methodology for extracting contextual data from distributed digital fingerprints of user profiles.

- Methodology for processing analytical data from large volume in cloud resources along with the management of unstructured databases.
- A methodology for describing unstructured data with automatically generated metadata.
- Methodology for analyzing contextual data using statistical and predictive methods.
- Methodology of contextual data analysis using classification methods and machine learning.

3. Communication solutions

In the communication processes of terminals with BLE beacons, standardization was applied based on principles similar to modified protocols used especially in the field of the Internet of Things. The protocol used is distinguished by its lightness and simplicity, thanks to which it can be successfully used on small microcontrollers, with limited hardware resources, and also when we are dealing with low link bandwidth, which perfectly fits the IoT concept. Therefore, it is not surprising that the developed solution will also be used for connecting devices to clouds (eg Amazon Web Services or Microsoft Azure). Its implementation can be carried out on many platforms, such as Arduino, ESP8266 or Raspberry Pi, as well as on more and more popular specialized FPGA (Field Programmable Gate Arrays) systems. Systems of this type have been used by the author's team in the project, as well as in their own research equipment.

The data exchange protocol, like the ubiquitous HTTP standard, is at the top of the TCP/IP stack in the application layer, but the specificity of their operation is completely different. The data exchange protocol is based on the publish / subscribe model, which enables asynchronous communication between publishers and subscribers. A protocol client is any device connected to the network (e.g. FPGAs, microcontroller, computer, terminal) that has implemented the TCP/IP stack and data exchange protocol, i.e. any device that can communicate with it. Each individual protocol client can simultaneously act as subscriber and publisher.

Clients, publisher and subscriber, do not communicate directly with each other. Moreover, one client can forward a message to another client without knowing its IP address and its existence. This is possible because there is always an intermediate element between clients in the publish / subscribe model, the so-called message broker. It receives messages from publishing clients and then distributes them to the appropriate subscribing clients.

3.1. Publishing data

The question is, how does the broker know which subscribers to send a given message to? Well, the publisher, when sending a message, adds additional information to it, a text called topic, and the broker sends the message to those customers who subscribe to this topic. This process is shown in the diagram below:

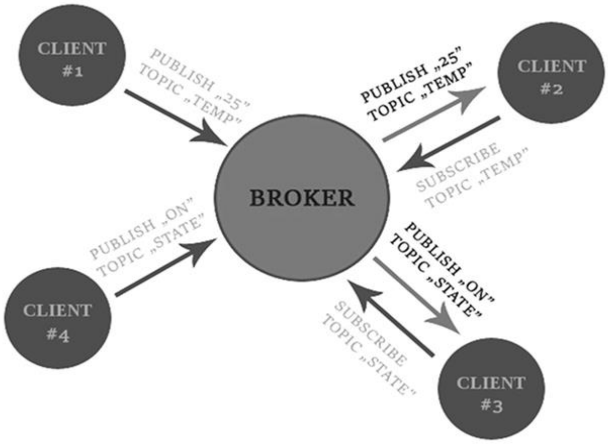

Figure 1. *Functional diagram of data publication and subscription processes in networks exchanging data related to digital fingerprints.*
Source: own materials.

Thanks to this mechanism, one node of the network can transmit information to many clients at a very low cost. It is enough for him to send a message to the broker and he will pass the information on to the other moving or stationary vertices of ad-hoc established network. It is thanks to this that the client can be implemented on small microcontrollers or cheap and simple FPGAs. Of course, in this case, the broker needs much more resources. After all, there may be many clients connected to it, it has to filter and send many messages, deal with client authentication and authorization, as well as message storage. Therefore, it is usually installed on more efficient platforms. Nevertheless, here too, specialized gate arrays come in handy, which can be strongly optimized for the tasks performed. This also lowers the manufacturing costs.

3.2.Client

The developed client libraries can be implemented in many languages, eg C#, C++, Python, Scala and many others. They can also be precompiled for platforms like Arduino, ESP8266, or FPGA. Libraries were created according to the POSIX standard. This applies, inter alia, to handle the TCP/IP stack (eg POSIX Sockets) as well as multithreading (POSIX Threads).

The client can work in two modes: synchronous and asynchronous. In synchronous mode, when executing one task, you must wait for it to complete before proceeding to the next task. The program works in a blocking manner, which very often introduces unnecessary delays. In the case of asynchronous mode, you can move to the next task without waiting for the previous one to finish. A client working in synchronous mode runs on one thread. When we send a message and want to be sure that the message has been delivered successfully, it is necessary to call the blocking function. When receiving a message, it is necessary to periodically call a function that blocks the program until a message arrives or the specified time has elapsed. A client working in asynchronous mode runs on multiple threads. Handshaking and connection hold are performed in the background, so you don't have to worry about it. The feedback function mechanism, the so-called callbacks. In this case, it is enough to define the necessary functions, register them, and then they will be automatically, asynchronously called when an event occurs, e.g. when a sent message is delivered.

4. Project results

The project implemented a new technology in both the test cloud resources and the test data collecting devices. Mobile terminals were also used as active data carriers, exchanging digital profiles in networks created in ad-hoc mode. The following briefly presents examples of the main components of a behavioral skeletal function implemented on the FPGAs of a broker implementing a device using the technology for collecting and transferring digital fingerprints and preprocessing using simplified deep learning mechanisms. The code has been optimized for the VHDL standard.

```
begin
mx: mux port map (a=>mux_in, s=>cycle, y=>mux_out);
Ai_mux: mux port map (a(0)=>tbl(A, idx(0)),
                a(1)=>tbl(A, idx(1)),
                a(2)=>tbl(A, idx(2)), s=>cycle, y=>Ai);
Bi_mux: mux port map (a(0)=>tbl(B, idx(0)),
                a(1)=>tbl(B, idx(1)),
                a(2)=>tbl(B, idx(2)), s=>cycle, y=>Bi);
Ci_mux: mux port map (a(0)=>tbl(C, idx(0)),
                a(1)=>tbl(C, idx(1)),
                a(2)=>tbl(C, idx(2)), s=>cycle, y=>Ci);
Di_mux: mux port map (a(0)=>tbl(D, idx(0)),
                a(1)=>tbl(D, idx(1)),
                a(2)=>tbl(D, idx(2)), s=>cycle, y=>Di);
mul1: FPP_MULT port map(Ai, p, fpu_clk, start, start, done(1), open, fequ_p(1));
mul2: FPP_MULT port map(Bi, mux_out, fpu_clk, start, start, done(2), open, fequ_p(2));
mul3: FPP_MULT port map(mux_out, mux_out, fpu_clk, start, start, done(3), open, fequ_p(3));
mul4: FPP_MULT port map(Ci, fequ_p(3), fpu_clk, start, done(3), done(4), open, fequ_p(4));
mul5: FPP_MULT port map(Di, PU2pPV2, fpu_clk, start, start, done(5), open, fequ_p(5));
done(6) <= done(1) and done(2);
sum1: FPP_ADD_SUB port map(fequ_p(1), fequ_p(2), fpu_clk, start, done(6), done(7), fequ_p(6));
done(8) <= done(4) and done(5);
```

sum2: FPP_ADD_SUB port map(fequ_p(4), fequ_p(5), fpu_clk, start, done(8), done(9), fequ_p(7));

div_mux <= div_opa when (do_div_opa_by_p = '1' and mux_s_buff = '0') else fequ_p(7);

div1: FPP_DIVIDE port map(div_mux, p, fpu_clk, start, done(9), done(10), open, fequ_p(8));

done(11) <= done(7) and done(10);

sum3: FPP_ADD_SUB port map(fequ_p(6), fequ_p(8), fpu_clk, start, done(11), done(12), fequ_p(9));

opa_div: if do_div_opa_by_p = '1' generate

div_done <= done(10);

p_cmp: fcmp port map(opa=>p, opb=>float_0, altb=>open, blta=>pLT0, unordered=>open, aeqb=>open, inf=>open, zero=>open);

end generate opa_div;

pipe_done <= '1' when do_div_opa_by_p = '0' else div_done;

pipe_result <= div_opa when do_div_opa_by_p = '0' else div_result when (div_done = '1' and pLT0 = '0') else float_0;

mux_s_buff <= '0' when state = RESET_WAIT else '1' when state = WAIT4DONE else mux_s_buff;

f_eq_buff <= float_0 when state = RESET_WAIT else f_eq_buff when state /= VALID_ON else pipe_result when mux_s_buff = '0' else alu_result;

```
process(rst, alu_done, cycle)
begin
  if rst = '1' then
    if rising_edge(alu_done) then
      case cycle is
        when "00" =>
          cycle <= "01";
        when "01" =>
          cycle <= "10";
        when "10" =>
          cycle <= "00";
        when others => NULL;
      end case;
    end if;
  else
    cycle <= (others=>'0');
  end if;
end process;
SYNC_PROC: process(clk, rst, out_valid_buff)
begin
  if rising_edge(clk) then
    if rst = '1' then
      state <= next_state;
      f_eq <= f_eq_buff;
```

```
          output_valid <= out_valid_buff;
        else
          state <= RESET_WAIT;
          f_eq <= (others=>'0');
          output_valid <= '0';
        end if;
      end if;
    end process;
    NEXT_STATE_DECODE: process(state, alu_done, enable, pipe_done)
    begin
      next_state <= state;
      case state is
        when VALID_ON =>
          if alu_done = '0' then
            next_state <= VALID_OFF;
          end if;
        when VALID_OFF =>
          next_state <= WAIT4DONE;
        when WAIT4DONE =>
          if (alu_done = '1' and enable = '1') then
            next_state <= VALID_ON;
          end if;
        when others =>
          if (pipe_done = '1' and enable = '1') then
            next_state <= VALID_ON;
          end if;
      end case;
    end process;
    OUTPUT_DECODE: process(state, mux_s_buff, rst)
    begin
      case state is
        when VALID_OFF =>
          start <= '1';
          out_valid_buff <= '1';
        when VALID_ON =>
          if mux_s_buff = '0' then
            start <= '1';
          else
            start <= '0';
          end if;
          out_valid_buff <= '0';
        when WAIT4DONE =>
          start <= '1';
          out_valid_buff <= '1';
        when others =>
```

```
            if rst = '0' then
               start <= '0';
            else
               start <= '1';
            end if;
            out_valid_buff <= '0';
      end case;
   end process;
end Behavioral;
```

As a result of the project, the concept of a device for generating promotional codes in terrain games (aka city games) supported by mobile applications for smartphones was also developed. It is a device for generating a promotional code based on the amount of money spent by the smartphone user in the place of use of the device, e.g. bar, restaurant, cafe. When the user of the promotional application makes the payment at the point of use of the device, the seller uses the numeric keypad to enter the value of the paid amount, which appears on the display above the keyboard. The device can also be directly connected to the cash register system. The processor of the device converts the entered amount into a promotional code, which is displayed on the display on the opposite side of the case. The application user reads the code and brings his smartphone closer to the NFC reader with the application running to start the procedure of entering the promotional code. The smartphone user can also use only NFC transmission since deployment of additional display is related to the compatibility with older phones not supporting NFC technology. The promo code allows you to earn loyalty points in terrain games supported by mobile applications. The promotional code is only valid for the specified time required to enter it. Only the application launched by touching the NFC reader allows the code to be entered, thus avoiding its use by a third party.

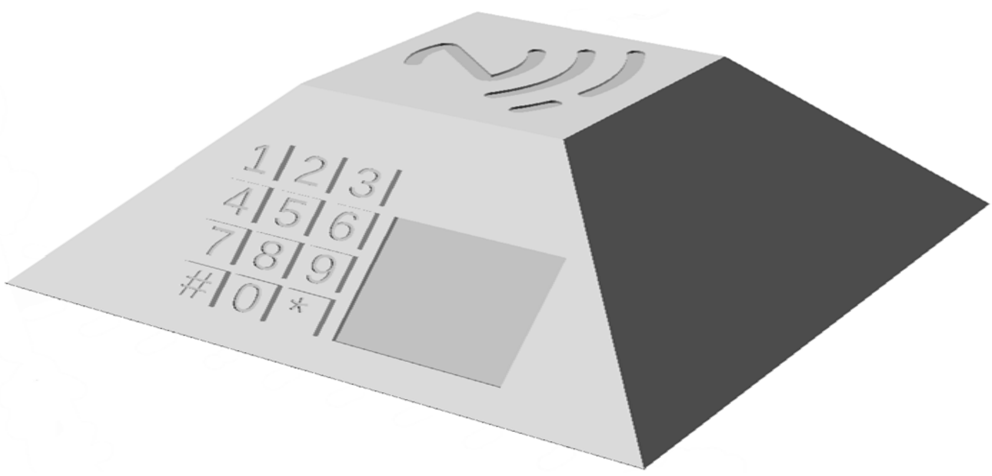

Figure 2. *An exemplary implementation of a device for generating promotional codes in field games supported by mobile applications for smartphones.*
Source: own materials.

Acknowledgements

The project was co-financed by the European Regional Development Fund under the Regional Operational Program of the Kujawsko-Pomorskie Voivodeship for 2014-2020. Project title: "Performing industrial research and development works leading to the creation of a new technology of multidimensional contextual analysis using the Internet of Things, data with a large volume of Big Data as well as cloud and mobile solutions".

(PL: Projekt współfinansowany ze środków Europejskiego Funduszu Rozwoju Regionalnego w ramach Regionalnego Programu Operacyjnego Województwa Kujawsko-Pomorskiego na lata 2014-2020. Tytuł projektu: „Wykonanie badań przemysłowych i prac rozwojowych prowadzących do stworzenia nowej technologii wielowymiarowej analizy kontekstowej z wykorzystaniem Internetu Rzeczy, danych o dużym wolumenie Big Data oraz rozwiązań chmurowych i mobilnych".)

References

[1] Omondi A.R., Rajapakse J.C., Bajger M. (2006), *FPGA Neurocomputers,* FPGA Implementations of Neural Networks, Springer, ISBN10-0-387-28485-0, p.1-34.
[2] *Estimated data from the Department of Tourism of the Ministry of Sport and Tourism regarding tourism for the first half of 2019*, https://www.msit.gov.pl/pl/turystyka/badania-rynku-turystycz/8537,PODSTAWOWE-DANE-STATYSTYCZNE-Z-TURYSTYKI-ZA-I-POLROCZE-2019-I-ICH-ZMIANA-W-ODNIE.html, access 24.09.2020.
[3] Girau B. (2006), *FPNA: Concepts and properties,* FPGA Implementations of Neural Networks, Springer, ISBN10-0-387-28485-0, p.63-97.
[4] Girau B. (2006), *FPNA: Applications and implementations,* FPGA Implementations of Neural Networks, Springer, ISBN10-0-387-28485-0, p.104-134.
[5] Ide A.N., Saito J.H. (2006), *FPGA Implementations of Neocognitrons*, FPGA Implementations of Neural Networks, Springer, ISBN10-0-387-28485-0, p.197-206.