Grand Programming Challenges Hackathon @ ICM University od Warsaw - Map the Gap

Piotr Bała (ICM UW e-mail: bala@icm.edu.pl), 21.02.2017

Problem statement

The subject of the first edition of the hackathon was the problem of the best design of fiber optic network development in Poland. At present, 9.5 million people use the fixed line, which translates to 22.7% penetration - below the European average, which is 30.88 percent. The fiber line penetration in Poland is worse than European average in terms of the number of people using the Internet with high speed, considered to be above 100 Mbps. In Europe, such high-speed internet is used by 6.5% of clients, in Poland only 2.3%. Interestingly, in the European Union wins Latvia, where is as much as 37.8% links faster than 100 Mbps (http://www.telko.in/p-divers-declaration-for-preferences-for-the-first-course-1-1-popc). The most important problem is the provision of broadband internet access in less urbanized areas. Expansion of the fiber optic network to offer high bandwidth connections is one of the challenges faced by telecom operators and government agencies.

Investments in ICT infrastructure are difficult and costly. The return on investment is planned for decades (typically 20 years). Therefore, investment planning is extremely important and difficult at the same time. Moreover, not only telecommunications companies are involved, but also public resources, for example within the framework of the Digital Poland Operational Program. Completed calls include public funding of connection of at least 880,000 white spots and a number of other locations including schools.

Current methods of designing fiber optic network development are based on statistical data such as the number of potential customers in a unit area. These methods are supplemented by a number of additional conditions: they take into account the preferences of the recipients or their potential interest, mainly on the basis of statistical analyzes. There are also solutions for designing and optimizing the fiber optic network in small areas - including one or more neighboring municipalities (eg https://www.softelnet.pl/wp-content/uploads/2017/01/Concept-Styci-FTTH -dla-POPC1_QGIS_GNIFREE_1.1.pdf). There is, however, no method for locating and optimizing the development of a fiber network on a global basis, which would allow for a deeper optimization and thus connecting more customers and lowering investment costs.

For this reason, it is worthwhile to use a new approach to the problem and to globally optimize fiber links based on a list of addresses and other available information, such as a list of access points to which new fiber networks may be connected, or a list of streets and roads along which fiber optic investments should be conducted. Due to the unavailability of the list with the existing fiber optic network, a list of LTE1800 BTS's could be used as a realistic approximation. Such list for Poland is available at the Electronic Communications Authority's (list of 25.01.2017). This list can easily replace the list of points of access to the existing fiber optic network of one or more operators.

Analysis of data at the national level requires the involvement of large computing power with multiple processors. This, in turn, requires appropriate algorithms, applications, and large computational resources. Unfortunately, knowledge on this subject is limited and available only in a few institutions such as ICM. Even in supercomputer centers, most of the calculations are done using only single processors, and the capabilities of available multiprocessor computers can be used by very few people. Grand Programming Challenges Hackathon participants were given access to computational resources that allowed them to analyze large data and their effective use was one of the challenges to be solved.

Hackathon results

The first edition of Great Programming Challenges Hackathon attended 3 teams (two more teams were registered but did not participate). All teams presented the solution to the problem (connecting as many addresses as possible from the list of white spots). Solutions differed in approach to the problem and the result. The teams were solving the problem in two variants:

- first, in which point connections are made by geometrically shortest segments,
- second, in which the fiber optic sections must be routed only along the roads (based on OpenStreetMap).

The team: **Aleksandra Kardaś, Rafał Kowalczyk, Piotr Konorski, Piotr Witkiewicz** (all students of newly created master degree programme in computational engineering run by ICM University of Warsaw) presented the method of connecting all address points from the list of white spots (over 2.7 million locations) with access points (BTS) in the first variant (the shortest way) using about 650 thousands kilometers of optical fibers. The team presented a preliminary proposal to guide optical fiber along the roads defined in OpenStreet Map. One of the interesting effects was the density analysis of points from the list of white spots.

The software was implemented in Java using the PCJ library. For calculations, 60 nodes of the Cray XC40 prvided by ICM were used (2880 computational cores), and a one-time data analysis using all resources lasted 5-10 minutes (if using one processor it would require 250 hours!).

The team composed of: **Michał Żak, Kamil Żył**a (students of the 4th year of computer science programme at MIM University of Warsaw) performed a series of analyzes based on searching for the closest locations. The team presented a connection of about 90,000 locations using 1 million km of fiber optics. Locations without internet were connected to the nearest access point (BTS) by direct fiber connection. The result shows the importance of global link optimization, which, as shown by the work of other teams, allows for a significant reduction in the total fiber length.

Results – discussion

Hackaton has shown that the analysis of large data covering more than 2.7 million locations is possible using existing computational resources, development tools, and competencies held by students and staff at ICM UW.

Analysis and optimization of large connection networks can be carried out on a global scale (nationwide) with any resolution - in this case resolution of a single building.

Hackathon has shown potential hidden in publicly available data from various sources (MC, UKE, public domain). Unfortunately, data is made available in a variety of formats, making it difficult to reuse them, and one of the challenges was converting data into a common format that allowed them to be correlated and used. Hackathon has shown that these barriers can be overcome; however, efforts to unify data and reduce the number of data formats used by public administrations could be useful.

The teams had a wide range of development tools available. The most effective solutions were developed using tools developed in the ICM UW (PCJ - Parallel Computing in Java: http://pcj.icm.edu.pl) and using Java. The Java language enabled the use of available libraries for reading data in various formats, including OpenStreet Map data stored in XML. The use of traditional tools such as the MPI (C ++) library proved to be much more difficult and caused many problems with writing the correct code and launching it (these problems are well known and motivated to find other solutions for programming methods and languages).

Hackathon has confirmed the validity of the ICM studies in computational engineering which gather students with a various background who gain knowledge in computer simulation, data analysis or large computing systems.

Results – examples

Below we present sample graphs showing hackathon results. The charts were prepared by a team composed of **Aleksandra Kardas**, **Rafał Kowalczyk**, **Piotr Konorski** and **Piotr Witkiewicz**. These are very preliminary results that require refining, but they illustrate the software capabilities of the hackathon participants.

Charts can be published only with the information on the authors and information that the results were obtained during the Great Programming Challenges Hackathon at ICM University of Warsaw - Map the Gap (http://wwp.icm.edu.pl).

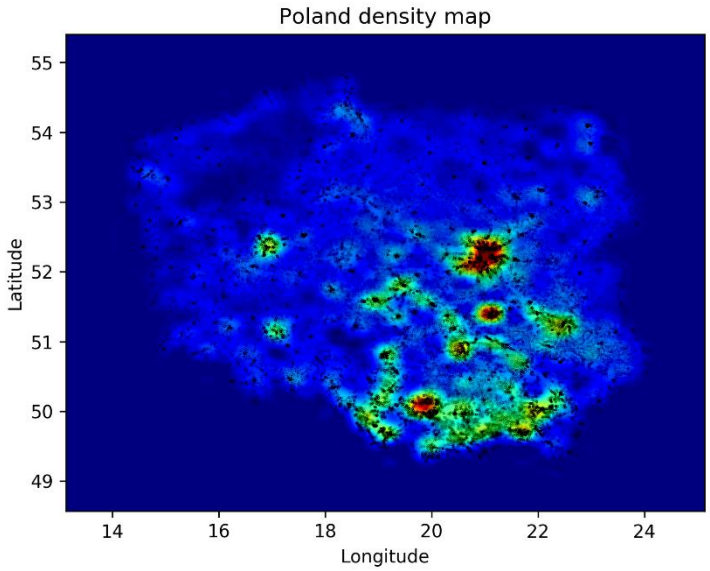

Fig. 1 Density of the location without broadband internet access in Poland (based on the list of white spots).
The chart shows areas of high density (yellow and red) in the vicinity of large agglomerations (eg. Warsaw, Kraków and southern Poland). In these areas it may turn out that public support is not necessary due to the large number of potential customers.

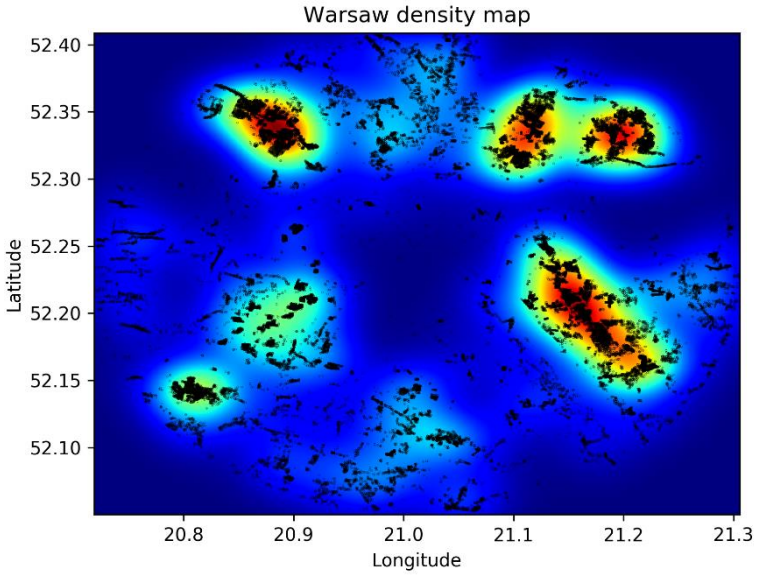

Fig. 2. Density of the location without broadband internet access in Warsaw and surroundings (based on the list of white spots).
Clearly visible areas densely populated with poor fiber infrastructure in suburbs (Wawer, Jabłonna).

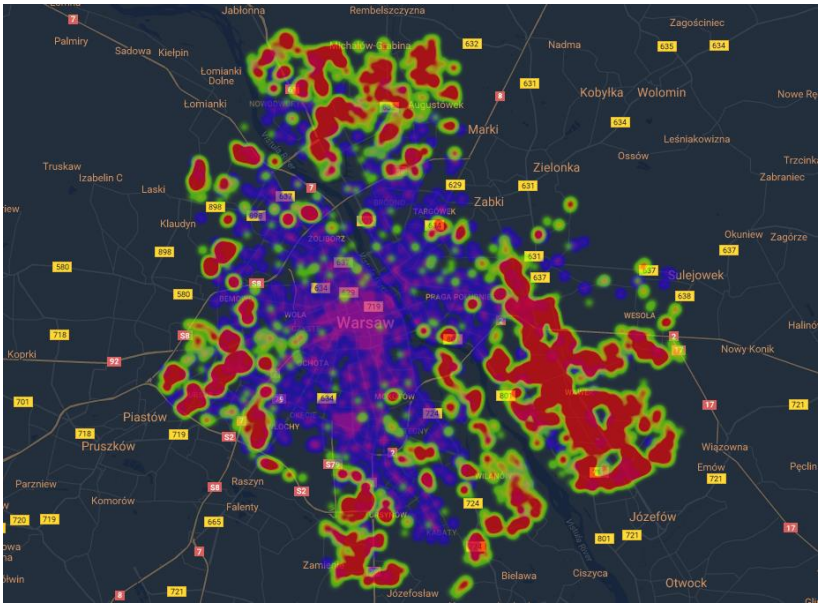

Fig. 3. Density of the location without broadband internet access in Warsaw and the surrounding area (based on the list of white spots, red means high density of location).

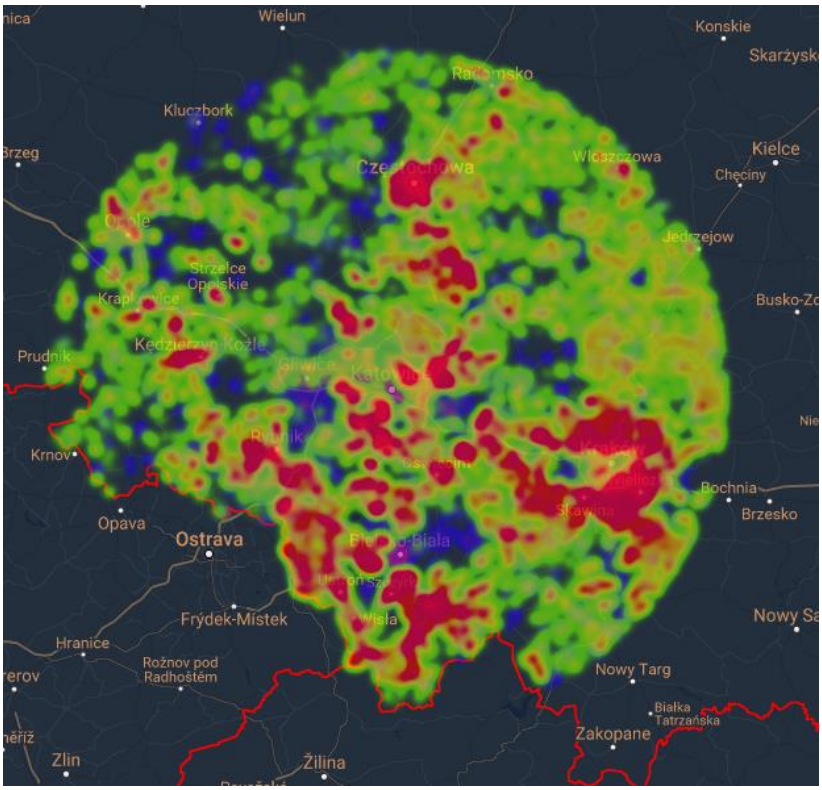

Fig. 4. Density of the location without broadband internet access in Cracow and surroundings (based on the list of white stains, red means high density of location).

Fig. 5. Example fiber connections in the shortest distance model (green lines). Points marked red indicate locations in the list of white spots, blue dots indicate the location of the BTS.
The area covered includes the Konotopa highway junction in Warsaw.

Fig. 6. Examples of fiber optic connections in a model where connections (green lines) are routed along roads. Points marked red indicate locations in the list of white spots, blue dots indicate the location of the BTS.
The area covered includes Włochy (Warsaw).

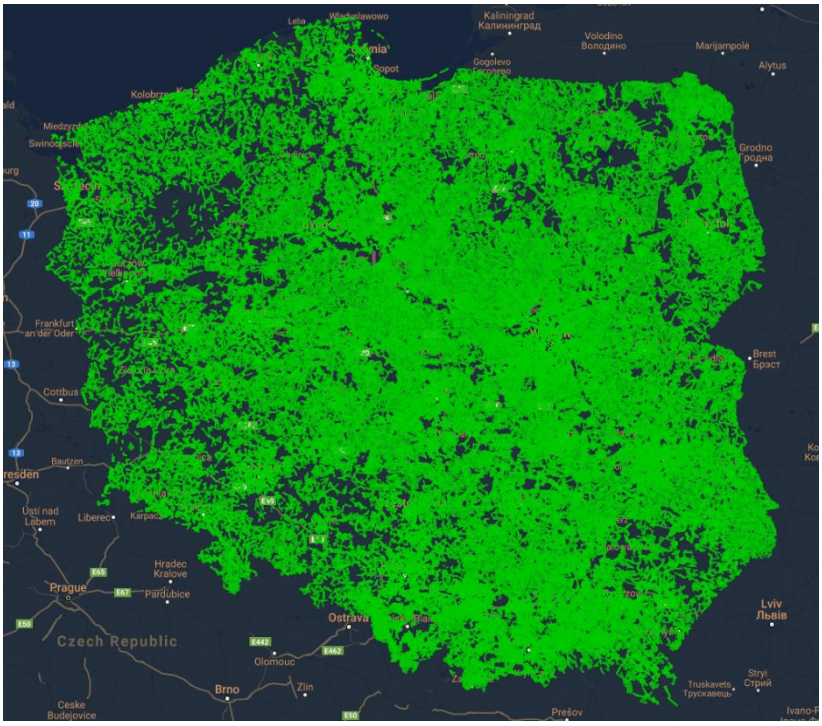

Fig. 7. Proposed path of optical fibers in the model of the shortest connections in the Polish scale. The total length of the connections required to connect all points from the list of white spots is 650,000 km.

With the further optimization performed after hackathon the length of the network was reduced to the 310,000 km (450,000 km in the model 2 – along the roads).