Distribution and validation of meteorological data for the air traffic management systems

Ondrej Marik, Roman Marik

University of Pardubice, Faculty of Electrical Engineering and Informatics, Studentska 95, 532 10 Pardubice, Czech Republic
ondrej.marik@hotmail.com, roman.marik@hotmail.com

Abstract. The main aim of this article is to introduce you our research project which is focused on processing, validation and distribution of meteorological data for the air traffic management systems. We would like to suggest to reader the direction of our research project and present a new perspective on meteorological data processing and distribution for the industries with a high degree of dependency on weather. Usually research projects are focused on data collection in one central datacenter to do the data analysis and simulations but when we think globally, we think is better to have independent localities who talk to each other. And that is what we would like to research, and what we would like to present to reader. We are focusing especially on the air traffic industry where the need for current meteorological data is must. As a result, we would like to develop software solution which will be able to use decentralized and solid architecture for data exchange and analysis without high dependency on central datacenters.

Keywords: Meteorological; Data; Distribution; Validation; ATM; Air traffic;

1 Introduction

The main aim of this article is the introduction of research project with codename TEVI. The project is focused on realization of innovative approach for the processing and distribution of meteorological data for the air traffic management where the need for quality and accurate data information is a must.

Currently the project is still in research and prototyping phase so the presented data can change during time but we still think it is a good idea to present you what we know now, what we found out and how we would like to solve the problems. Also, we are going to publish all results and experiences after the project will be finished and certified so the reader can easily track progress and good or bad approaches we tried because we believe there is a huge space for improvement how to manage and use such a data in nowadays world.

For the rest of this article we will use model situation of the Czech Republic and our pilot project with the Army of the Czech Republic (ACR) where we would like to deploy our system for the distribution of data into air traffic management system

(ATMS) and other systems for support of air traffic management. We would like to deploy it on all active airbases ACR have and at two datacenters where ACR has supporting systems and services and training center.

2 Current situation and working procedures

In nowadays is international standard for the meteorological data publication and interchange in air traffic management (ATM) across the world METAR format. This format is defined and standardized by World Meteorological Organization (WMO) and International Civil Aviation (ICAO). Reader can find list and explanation of all codes in messages regarding to standard in [6] where are also some examples of reports and its structures. Detailed information about implementation of OPMET (Operational meteorological data) can be find in [4].
In a simple way, METAR can be defined as text format message which contain summary information about current meteorological conditions on any airbase or measuring location.
Nice example of METAR message is published in [2]:

```
METAR LBBG 041600Z 12012MPS 090V150 1400 R04/P1500N
R22/P1500U +SN BKN022 OVC050 M04/M07 Q1020 NOSIG
8849//91=
```

As you can see, message is represented by several information in one continuous string. First is indication of message type. It can be METAR or SPECI. Difference between these type of message is just about the time and situation when has been the message published. METAR is published regularly every hour in 365/24/7 mode. SPECI on the other hand is special type of METAR with completely same format (except keyword at the start) but it is published outside of the standard publishing window. This can happen because of technical issues, unexpected significant change of meteorological situation, wrong information in regular METAR message etc. After the message type keywords follows these data fields – **ICAO airport code, time and date, wind information, prevailing visibility, runway visual range, snow situation, cloud information, temperature and dew point, altimeter setting, trend forecast, runway condition, CAVOK information** (cloud and visibility ok).

2.1 Current architecture of data distribution

From international data distribution point of view looks current situation in Czech Republic very similar like in other countries. There is on central collecting point which is usually operated by one central institution which guarantee data collection and distribution for the whole country or region. This central point then distribute data into international interchange network or to its parent collecting point.
As you may notice, it is basically classical pyramid topology with all its cons and pros as you can see on Figure 1.
In our case is central collection and distribution point VGHMU Dobruska (military geographical and hydro meteorological office). This point is used as meteorological

data (METAR/SPECI) source for all military ATMS and locations. All process and data exchange of meteorological information in Czech Republic is defined in [5].

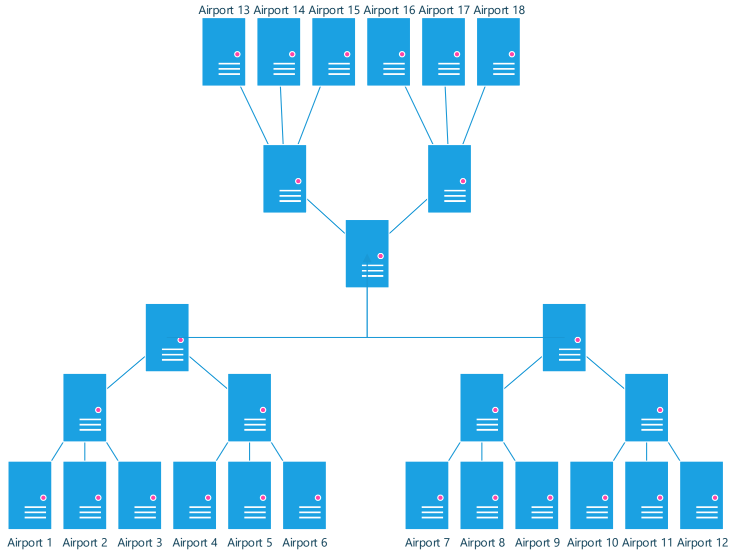

Figure 1 - Current architecture of collection and distribution

2.2 Disadvantages of the current state

Inflexibility and dependence

One of the biggest disadvantages of the pyramid topology is dependence on the central collecting points because each of it represents potential problems in case of errors but not only for the parent's points, but also for the points in a same region. It is because of that many real-life topologies or systems uses central points as exchange place also for the servers on a same hierarchy level so in case of unavailability of the central point, even neighbors' servers cannot exchange its' data.

It goes without saying that these points should use the high availability architecture and they should be redundant but in our opinion, this is not enough for example in a military environment where we can expect attack on the central datacenters during conflict.

Amount of data

Because in pyramid topology data are collected in a way up, the amount of data to keep is bigger and bigger as collection point is closer to top. That leads to situations when top points have all information from child points, but they cannot handle and store nothing detailed than hours information (METAR/SPECI).

For readers' imagination about data size, we can do a little calculation. During analysis and first tests of our project, we have found that one airbase of ACR generates about 64GB of RAW data. If we consider that ACR have four active airbases we get 256 GB

of RAW data per year. Then we, of course, needs data from civil airports in Czech Republic (let's count just the bigger ones – Prague, Brno, Karlovy Vary, Ostrava, Vodochody) we get 0.5 TB of RAW data and that's just from the Czech Republic (CZ). On the other hand, in [1] is written that in United States of America was in 2013 totally 15 513 airports. That means that in case we would like to store all RAW data we would have to have solution for storing and processing 992 TB of data per year.
So as reader can see this was also the reason for using just the METARs which represents just the fraction of RAW data (about 100B per message). Of course, the main question is if this is sufficient and accurate.

Data message transmissions
There are no general rules for the data collection and transmission between end localities and collection points so national organizations usually choose their own way how to handle it. Czech institutions for example usually uses FTP protocol to the central storage which basically represents batch processing than online data distribution so there can be smaller or bigger delay and in a same time different system can have different information.

Data quality
Another disadvantage of current solutions is absence of validation and data quality control. Data generation for the ATM is usually solved by two modes – automatic and manual.

Automatic mode
If the measuring system is in automatic mode, it automatically reads current values from the sensors then generates message and sends it for publishing. In this mode systems also usually enables modification of messages before its publishing.
Problem of this mode could be situation when something happens with sensors so it indicates wrong data. This mode is also usually unable to detect short term external influences. Point is, that in many cases is measure "garden" close to the runway or taxiway which is necessary for having accurate data on the edge of the runway for example. So hypothetically can occur situations when airplanes with started engine can directly influence the sensors like wind measure, temperature etc. That can lead to publication of wrong data and to unexpected situations.

Manual mode
In manual mode is message generated by worker of the meteorological service on airbase. System for data measuring just indicates current value and its evaluation is up to worker who fill it to the final message.
Disadvantage of this mode is high dependency on human which can lead to mistakes. There also potentially exists same problems with sensor error or short term external influence as in automatic mode.
As can be seen from the above, the current situation and working procedures are functional and proven, but certainly not ideal, and in our opinion, there is space for improvement.

3 Aim of project

The main aim of our project is development of solution which will deals with most of the disadvantages of current systems for distribution of meteorological data with focus on distribution to the ATMS and its quality.
When we were looking for similar research projects, we found out that many of them are focusing more on data collection and simulation like in [3], others for data analysis and their use for decision making [7], but we haven't found any that would be focused on data distribution. Lot of research projects also focuses on centralized architecture which we believe has more disadvantages than advantages so we decided to try to solve this problem kind different than others.
The proceeding text will be describing our pilot project with the ACR were we would like to apply all our ideas.

4 Planned system architecture

As it was written in chapter A, in Czech Republic is used architecture with central collection and distribution point operated by VGHMU. All the ACR airbases and its meteorological systems send data directly on this sever via FTP. That means there exists only one connection and in case of its failure none of the airbases has data from the others which is not quite ideal.
If consider that it is a military environment, it should be our main aim to develop decentralized system architecture with similar principles like telecommunications networks with great emphasis on security, resistant and flexibility of whole system.
On the other hand, we must also consider the backward compatibility of system because lot of applications and systems are proprietary and it its reconfiguration or modification will be possible only its future upgrade.
That leads us to design system which will works on full mesh topology and which will also have central analytical cluster. By full mesh topology is meant such a connection when every server of our system has connection to any part of our system across the decentralized military communication network. The reader should not understand topology in terms of directly connected physical data carriers. The topology is in fact deployed across the Czech Republic.

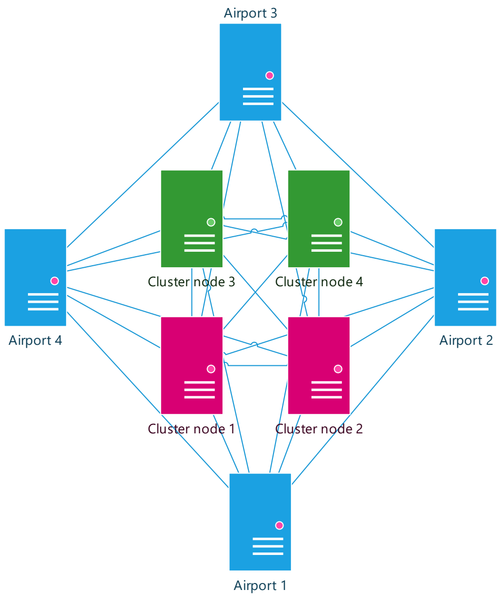

Figure 2 - Proposed architecture

On Figure 2 is shown current architecture of our pilot project. The reader can see that every airbase has its owns server pro data processing and distribution of meteorological information. At the same time, we have also central analytic cluster, which has three main roles.
The first role is replacement of the VGHMU collection and distribution server (from inside of our system). This provide us backward compatibility solution for old systems or system waiting for its upgrade. Another effect is that we can act as common system to the “outside” world without need of reconfiguration at national data collecting point.
Second role is planned advance real-time analytics over data from all airbases where would like to evaluate current situation, calculating short and long-term trends, historical analytics etc.
Third role is backup of active servers on every airbase. This is basically only protection for every airbase before hardware problems but it saves us part of the budget because we don’t necessarily need two active servers on every airbase.
The color difference of cluster nodes on the Figure 2 indicates its locations. As you can see, we spread them into two independent datacenters and unlike other projects (for example [3]) we are focusing on decentralized architecture with central support instead of building large central computational cluster.

5 Data processing and distribution

5.1 Data processing on airbase

On every airbase is deployed on independent server which can provide all functions of system even in case when the whole airbase is disconnected from network (in that case provides just local data). The procedure workflow of services on servers look like this:

5.2 Data mining from measuring system

System is directly connected into meteorological measuring system which sends current sensor values in predefined intervals via TCP/IP – defined by manufacturer of measuring device so it is basically in real-time mode.

5.3 Data storing

Once system receive any data, it validates its source and saves it into the SQL server. Before final commit into the database system is passing standard data procedure like writing RAW data, cleaning, sorting, basic checking on invalid characters etc.

5.4 Validation and evaluation

Once are RAW data stored system starts with its validation. This step is crucial in the whole project, because it is this feature that differentiates our system from others.
In common we can say that data validation is examine these areas:

- Comparing with previous values
- Validation against common values
- Validation based on calculated long-term and short-term trend
- Validation against historical and statistics data

This process is passed for every data value of received data and based on its result, every value gets indication of its state – ok, error, waring, unexpected change etc. These indications are then passed with data during it publications so end user now if value is save or not.
The main idea is of this process is not deleting of data which looks suspicious or which fails on validation but passing that information to the end user of that value no matter if he accesses it via web interface of our system or if get it in ATMS.
It can happen that for example unexpected change of wind speed or direction is not error of sensor or error because of external influences but it can be simply just gust of wind. In case of ATMS is then up to worker how he will deal with that information.

This part of system is one of the biggest challenge right now because we trying to find such a method, which eliminates false alarms on minimum and this needs lot of testing, data observing, analytics etc.

5.5 Data publishing

Once data have been validated and tagged, they will be published. Unlike standard systems where data is distributed over a central collection point, we have decided to provide a complete public API for system and data access. In addition, the data will not be published in any periodic cycles (except for the reports like METAR), but it will be send via PUSH notifications to each of currently registered and listening client.
This solution has a huge advantage and gives the entire system a huge potential for the integration with ATM systems of all current vendors and practically with any system because we will provide full description of API and we will keep it open as much as possible.
This API also helps in decentralize topology because server can talk directly to each other over standardized interface and they don't need to keep any information (except its address) from other locations. It is also important to keep in mind that every data is published with validation indications.

5.6 Archive

At this point of research, we have no valid data about archive but our planned approach is to keep all data in a SQL at least for two years (depends on HW). Once the data reach the deadline, we are going to flush detailed data into files and saved them on place, where we can index it and access it very fast. But we will keep the summary data in the SQL for the maximum performance and common statistical task. We are going to use data in the files only in case when someone call API request on it with maximal granularity.
The archive also must be accessible via API so the archive storage cannot be offline storage.

6 TECHNICAL BACKGROUND

The system is currently based on Microsoft technologies – Windows Server, SQL Server, C#, ASP.NET MVC / CORE. It is mainly because we are counting on fact, that big environments need big analytics power so we build system from the start for compatibility with Microsoft Azure because it offers great ratio between cloud features, power and integration with local environment so we can deploy completely hybrid solutions.
In a future, we count we will be Linux compatible once Microsoft release its SQL on Linux.

7 Future potential

If everything will be fine, we think that our system has huge potential for the future use. Our philosophy is to build solution in the box which will be compatible with as many meteorological measure systems as possible and which will be able to communicate with whole world right after the initial configuration. If you imagine situation when many locations would be equipped with our system and all of them would be connected into the internet. That could totally change the way how we use meteorological information in air traffic now.

For example, in nowadays is common thing that airplanes have direct data communication to the ground or into the internet. If they would be able to communicate with big network of our systems, the crew on board can know about change of weather conditions miles before they reach the destinations and system can offer them the closest airport with good conditions for landing or the airplane can inform closest system about current condition in its flight altitude to accurate the data of ground systems.

8 Conclusion

This article introduced to the reader our project which is still in prototyping and research phase. We introduced our motivation, basic functions of system and our plans into the future development.

We would like to thanks to all of readers and you have any questions do not hesitate to contact us.

9 Reference

[1] Central Intelligence Agency, „Total Number of Airports by Country," 2011. [Online]. Available: https://www.cia.gov/library/publications/the-world-factbook/fields/2053.html. [Accesed 20 6 2017].

[2] Wikimedia Foundation, Inc., „Wikipedia - METAR," 22 4 2017. [Online]. Available: https://en.wikipedia.org/wiki/METAR. [Accesed 21 6 2017].

[3] C. S. Tascón a J. Gonzalo, „Advances in digital meteorological services (DMET) for ATM: Aviation meteorology in support of air traffic optimization and automation," *SIDs 2011 - Proceedings of the SESAR Innovation Days,* 2011.

[4] International Civil Aviation Organization, „Guidelines for the Implementation of OPMET Data Exchange using IWXXM in the EUR Region (EUR Doc 033)," 23 9 2016. [Online]. Available: https://www.icao.int/EURNAT/EUR%20and%20NAT%20Documents/EUR%20Documents/033%20%E2%80%93%20CONOPS%20for%20the%20Transition%20of%20OPMET%20Data%20Exchange%20using%20IWXXM/EUR

%20Doc%2033%20%20%20(EN)%20-%20Edition%202,%20Amd%200.pdf. [Přístup získán 20 6 2017].

[5] Civil aviation authority Czech Republic, „Letecký predpis L3: Meteorologie," 10 11 2016. [Online]. Available: https://lis.rlp.cz/predpisy/predpisy/dokumenty/L/L-3/data/print/L-3_cely.pdf. [Accesed 20 6 2016].

[6] World Meteorological Organization, „WMO Publication No.306: Manual on Codes, International Codes, Volume I.1, Annex II to the WMO Technical Regulations, Part A – Alphanumeric Codes," 2016. [Online]. Available: http://www.wmo.int/pages/prog/www/WMOCodes/WMO306_vI1/Publications/2016update/WMO306_vI1_en_2011UP2016.pdf. [Accesed 20 6 2017].

[7] C. Klüver a J. Klüver, „A Self-Enforcing Neural Network as Decision Support System for Air Traffic Control based on probabilistic Weather Forecasts," v *Neural Networks (IJCNN), 2017 International Joint Conference on* , Anchorage, 2017.