Technique for storing and automated processing of weather station data in cloud platforms

S Sokolov\textsuperscript{1}, S Vlaev\textsuperscript{2}, M Chalashkanov\textsuperscript{3}

\textsuperscript{1}University of Telecommunications and Post, Department of Information Technology, Sofia, Bulgaria, strahil.sokolov@gmail.com
\textsuperscript{2}University of Telecommunications and Post, Department of Information Technology, Sofia, Bulgaria, svlaev@mail.bg
\textsuperscript{3}University of Telecommunications and Post, Department of Information Technology, Sofia, Bulgaria, mike83@abv.bg

E-mail: strahil.sokolov@gmail.com

Abstract. In this paper we present an approach for design of cloud-based systems for storing and processing of sensor data from weather stations. In recent years there are abundant IoT devices that are capable of fairly precise data acquisition of local atmospheric parameters and reliable transmission over wireless network. The used dataset is acquired from a modern weather station model from a superior European manufacturer. 17 parameters are used for training neural network based LSTM model. In our research we have also compared several models and approaches for modeling this type of data. We propose a design of container based system for automated processing of the received data, visualization and cloud storage technique. We are comparing two types of open-source container technologies for implementing data processing. We have evaluated the performance and provide the set of criteria for choosing a container platform. Storing the data in the cloud is an important part and we provide our consideration for implementing a particular open source platform that could reduce costs fulfil the requirements for weather data storage. The designed system is expected to be cloud-provider agnostic. The proposed features make the system especially suitable for purposes such as local weather forecast; disaster prediction: hurricane, severe storms, floods, tornadoes, etc.; severe droughts with fire dangers; prediction of landslides. Depending on the availability of additional sensors the future development of the system may be extended with parameters of the soil. Training the model on a freely available data set is in our future plans for system development.

1. Introduction

There is a growing demand for IoT and Cloud based applications worldwide and it is considered that one area is bound to exist with the other. Therefore there is a need for modern approaches and platform which are able to provide a combination of automated data acquisition, secure cloud storage, algorithms for data processing and improved availability. Our aim is to enhance the current practices in the field of IoT Cloud.

There exist various modern data acquisition systems and wireless devices for weather data, which are used in meteorology and smart agriculture. Several designs are available, as the researchers in [1] have shown that automatic weather stations seem to be more efficient and user-friendly. An automatic weather station is a device that is capable of measuring and recording weather parameters using

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd
sensors and does not rely on human intervention. The measured data can be stored in the embedded data logger or can be transmitted to a centralized data storage facility through a supported type of communication channel. In case that the measurements are stored in a local data logger, the records must be manually copied to a computer at a later time for further processing. This is not an acceptable scenario, especially when the weather station is located at a remote location and usually operates in unattended mode. Therefore, the availability of communication system is an essential building block in an automated weather station. The modern technologies \[2\], \[3\] allow that such stations be built be built at a different location and maintained on the remote location with minimum effort to transmit data at regular intervals. The work showed that multiple weather parameters can be captured in real-time, at pre-programmed intervals \[1\]. This type of measurements are important to detect extreme events such as floods, hurricanes, tornadoes, as well as to issue timely warning messages to the public \[4\].

The authors in \[5\] show that the Mobile Cloud Computing is an emerging technology that involves a type of infrastructure in which data storage and processing are done outside the smartphone. IoT (Internet of Things) has been around as a concept technology and attracts ever-growing interest in the telecommunications domain, especially when IoT relates to wireless telecommunications. As the main purpose of IoT the authors describe the interaction and cooperation between physical entities and abstract objects that are being sent over the wireless protocols to act and serve as a combined entity \[6\], \[7\]. This enables the rapid growth and demand of resources in both Cloud Computing and Internet of Things. The authors have provided a survey most frequently used IoT and Cloud Computing technologies with a focus on the security issues of both technologies and claim that Cloud Computing technology improves the IoT operations. In addition, there are security challenges of the integration of IoT and Cloud Computing which need to be taken into consideration and avoided if possible.

Cloud computing is the abstract underlying platform \[5\] which provides computing resources, storage, services, and middleware that enable us to server applications to the public over the Internet. The IoT cloud computing is defined as combination of cloud computing technology with mobile devices and sensors. It is a multidisciplinary area and can also be referred to as mobile cloud computing. There exist two main perspective to the mobile cloud: a) based on infrastructure, and b) ad-hoc based. The concept that is based on infrastructure relies on the fact that the hardware infrastructure remains static and also provides services to the mobile users and potentially to other services. In our current research, this type of more static cloud is covered.

Several challenges have been listed in \[5\], \[8\] regarding cloud computing: One of them is Heterogeneity and is related to the wide variability of devices, operating systems, platforms, and services available and possibly used for new or improved applications.

Another challenge which is pointed often by authors in the field is performance. There are IoT Cloud applications that introduce specific performance and QoS requirements at several levels (with aspect to communication, computation, and storage) \[9\].

Further challenge is often reliability: adopting the IoT Cloud for mission-critical applications, reliability concerns typically arise e.g., in the context of smart mobility, users are often on the move and the moving device communication and connection is often intermittent or unreliable.

Volumes of data or the so called Big Data is another major concern within the IoT Cloud computing field. With the ever growing number of IoT devices, the demands for storage, access, and processing of the huge amount of data they produce also rises. \[10\]

Last but not least, the authors have mentioned monitoring of the IoT Cloud as a major challenge. It represents essential activity in Cloud environments and serves as basis for billing, capacity planning, for managing resources, SLAs, performance and security audit, troubleshooting.

Processing of the weather-station data in the cloud requires efficient algorithms and approaches which are able to extract relevant features and make precise classification of the inputs in a computationally efficient way. With the recent focus of deep learning methods, especially long short-term memory (LSTM) neural networks, there have been a number of interesting applications especially in the field of network traffic prediction \[11\]. The true value of deep learning methods \[12\] in traffic event forecasting is similar to the feed which the sensor data is providing from weather station. There is a need, according to the authors that these methods need to be further exploited in
terms of the depth of the model architecture, the spatial scale of the prediction area, and the predictive power of spatial-temporal data.

The authors have proposed a deep stacked bidirectional and unidirectional LSTM (SBULSTM) neural network architecture, which considers both forward and backward dependencies in time series data, to predict network-wide traffic speed [13]. A bidirectional LSTM (BDLSM) layer is utilized to capture spatial features and bidirectional temporal dependencies from historical data. The authors propose a model which is able to handle missing values in input data by using a masking mechanism. Their model is scalable model and is applied on complex networks.

The rest of the paper is organized as follows: In chapter 2 we describe the used IoT Cloud architecture and its components; Chapter 3 provides details on the used approach for weather station data processing; Chapter 4 describes our approach for enhancing the performance of the system by using the selected components in comparison to traditional techniques; chapter 5 provides details on the used devices for weather data acquisition, experimental results and discussion.

2. Description of the proposed IoT Cloud architecture

The proposed architecture consists of input layer (the weather stations), the data is then transmitted over the Internet to our private cloud and the distribution layer (load balancer) which propagates the requests to the backend layer which is composed of Docker containers. The applications which run in Docker container have a small footprint and are relatively resilient. The bottom layer is the persistence layer and it is based on CEPH storage. It provides unmatched object storage, high availability and low cost of operation compared to many of the products which we have had experience with so far. More details will be provided below.

Below are some of the features which we are aiming to achieve via our design by incorporating elements of Cloud Computing technology for IoT and overcome some of the challenges that have been presented in the previous chapter:

- **Storage over Internet**: One of the features of the proposed architecture is that we achieve Storage provisioning through Software Defined Storage (SDS) mechanism [14]. The core technology which enables the persistent storage for the weather-station data and for the data processing algorithms storage is CEPH [14], [15], [16]. CEPH is an open-source software storage platform that implements object storage on a single distributed computer cluster, and provides simultaneously interfaces for object-, block- and file-level storage. Its main goal is completely distributed operation without a single point of failure, scalable to exabyte level, and freely available. CEPH replicates data and makes it fault-tolerant, [16] using off-the shelf hardware and requires no specific hardware support. The resulting storage subsystem is by design self-healing and self-managing with additional minimization of administration time and other costs.

- **Service over Internet**: The usage of Docker containers to provide backend services for data processing, presentation services for the processing results, our goal was to achieve high availability and continuous integration and delivery of the services. Docker containers being lightweight and faster than virtual machines are able to provide this type of flexibility [15], [17]. The applications that perform the weather-station data processing are Python based; they run in Docker containers and use the persistence storage layer to read the input values and store the predictions.

- **Applications over internet**: Our load balancing layer is based on HAProxy [15]. HAProxy is free software and open source. It is the tool which provides high availability load balancer service in our architecture for TCP and HTTP-based applications. It is able to distribute requests across multiple backend servers using high performance algorithms and is written in C. This makes this software famous for being fast and efficient.

- **Computationally capable**: The designed subsystem for weather station data processing is able to handle massive amounts of data fed by the weather stations. It is based on scalable architecture [15]. For automation of the environment while handling greater workload, we have used Red Hat® Ansible®. It is a famous IT tool for automating repetitive tasks that
transforms inefficient tasks of software release cycles into predictable, scalable, and simplified processes [18]. It is used in our design to provide automation for configuration management, application deployment, cloud provisioning and service orchestration.

![Image](image_url)

**Figure 1.** IoT Cloud architecture for weather station data processing.

### 3. Weather station data processing in the cloud

The proposed architecture on figure 1 relies on the LSTM scalable model [11] to process the input data and provide forecasting for the coming intervals of time. Each value or feature (such as temperature, solar radiation, wind speed) discrete time series from the used sensors can be represented as a vector with $n$ historical steps:

$$X_f = [X_{T-n}, X_{T-n+1}, X_{T-n+2}, ..., X_{T-2}, X_{T-1}],$$  

where $f$ is the $f$-th feature, $T$ is the value of the feature at time $T$. The prediction of the values at the time $T$ will be represented by the following matrix:

$$X_T^p = \begin{bmatrix} X_1^p & X_2^p & \cdots & X_f^p \end{bmatrix},$$  

where $P$ is the number of measurements which are provides by the weather station. Thus, the current snapshot of the data at time $T$ will be designated as (3):

$$X_T^p = \begin{bmatrix} X_1^p \\ X_2^p \\ \vdots \\ X_f^p \end{bmatrix} = \begin{bmatrix} X_{T-n}X_{T-n+1}\cdots X_{T-2}X_{T-1} \\ X_{T-n}X_{T-n+1}\cdots X_{T-2}X_{T-1}^2 \\ \vdots \\ X_{T-n}X_{T-n+1}\cdots X_{T-2}X_{T-1}^p \end{bmatrix}.$$  

RNN is a class of powerful deep neural network using its internal memory with loops to deal with sequence data. The architecture of RNNs, which also is the basic structure of LSTMs, is illustrated below. For a hidden layer in RNN, it receives an input vector $X_T^p$, and generates the output vector $Y_T$. RNNs exhibit the superior capability of adapting themselves to predict nonlinear time series problems [11]. Some of the RNNs are bound to reach the vanishing during the Backpropagation (BP) coefficient learning, and thus, become incapable of learning from long time lags, or keeping track of long-term dependencies. It is necessary to consider the above problems when using RNNs and use any of the
advanced techniques to deal with them, such as the Long short term memory LSTM architecture and Gated Recurrent Unit (GRU) architecture. According to [11] LSTMs have been shown to work better on sequence-based tasks with long-term dependencies. GRU represents a simplified LSTM architecture being fairly new – is used in the context of machine translation. Out of the many existing variants of LSTM, we have chosen to use the proposed algorithm [19] for processing of the weather station data, i.e. standard LSTM architecture.

4. Performance considerations while using cloud infrastructures
The standard approach to building a software delivery process involves supporting a large number of virtual machines, with their own environments and versions. They must be constantly available, which means that they constantly occupy resources, even when they do not perform any tasks, but simply idle. With the proposed dynamic infrastructure, everything is exactly the opposite. Containers are used instead of regular virtual machines. The advantage is that many containers can run on a single virtual machine without much interference. The containers needed to perform a particular processing task – that can be scheduled – are created when they are required, rather than working constantly, borrowing resources. All versions and variables in the environment can be used dynamically, and after the task is completed, the created resources are destroyed. New ones are created as needed.

5. Experimental results and discussion
In our experiments we used data from a weather station ATMOS 41 [3]. It has 12 weather sensors in a single device for measuring atmospheric conditions. It fulfills one of the prerequisites for being able to provide continuous deployment in severe conditions and climates, such as deserts or mountains. Two of the 17 features used in building our experimental model are precipitation and solar radiation. The device manufacturer claims high accuracy on these measurements, i.e. 0.017 mm resolution means it is said to accurately measure even the smallest of rainfalls and even heavy dew events that other rain gauges might omit. In addition, the device’s anemometer is also claimed to be very accurate even at low wind speeds due to the lack of moving parts. The attached accelerometer also contributes to the station’s level and accuracy.

To measure the performance gain while using of the proposed IoT Cloud system components in comparison to the traditional infrastructure, we have prepared an experiment of multiple deployments of the architecture on our private cloud infrastructure. The speedup which we were able to achieve of container-based deployment vs. virtual machine-based is more than 2 times measured at up to 10 runs of the automated infrastructure deployment.

![Figure 2. Speedup coefficient with 10 runs of the deployed architecture.](image-url)
The calculated Speedup coefficient \( S = \frac{T_{\text{hypervisor}}}{T_{\text{containers}}} \) is measured as the ratio between the time needed to deploy a number of instances (Tasks) with the runtime components required to operate this environment on a hypervisor and a container platform. The advantages of the container platform can be inferred from figure 2.

6. Conclusion

A design was proposed for a container based IoT Cloud capable system that delivers automated processing of the received data, visualization and cloud storage. We are comparing two types of open-source container technologies for implementing data processing. Performance evaluation has been calculated and we provide the set of criteria for choosing a container platform. Storing the data in the cloud is an important part and we provide our consideration for implementing a particular open source platform that could reduce costs and fulfill the requirements for weather data storage. The designed system is expected to be cloud-provider agnostic. The proposed features make the system especially suitable for purposes such as local weather forecast; disaster prediction: hurricane, severe storms, floods, tornadoes, etc.; severe droughts with fire dangers; prediction of landslides. Depending on the availability of additional sensors the future development of the system may be extended with parameters of the soil. Training the model on a freely available data set is planned for the next steps of the system development.

References

[1] Weerasinghe, R., Aroos M., Pannila A., Jayananda M. and Sonnadara D. 2011 Construction of an automated weather station for ground level weather measurements, *IESL Annual Transactions*, pp 450-455

[2] Tenzin S., Siyang S., Pobkrit T. and Kerdcharoen T. 2017 Low cost weather station for climate-smart agriculture, *9th International Conference on Knowledge and Smart Technology (KST)*, Chonburi, pp 172-177, doi: 10.1109/KST.2017.7886085

[3] ATMOS 41 Weather Station, Available at: https://www.metergroup.com/environment/products/

[4] Peek, S and Peek D. 2004 Cellular weather station and computer system using the public cellular data telephone system and internet for controlling irrigation and method of use, U.S. Patent US6343255B1, issued January 6, 2004

[5] Stergiou C., Psannis K., Kim B. and Gupta B. 2016 Secure integration of IoT and cloud computing, *Future Generation Computer Systems*, 78, Part 3, 2018, pp 964-975, ISSN 0167-739X, doi: https://doi.org/10.1016/j.future.2016.11.031

[6] Schuler T., Dunse T., Østby T. and Hagen, J. 2014 Meteorological conditions on an Arctic ice cap—8 years of automatic weather station data from Austfonna, Svalbard. *Int. J. Climatol.*, 34, pp 2047-2058, doi: https://doi.org/10.1002/joc.3821

[7] Royapoor M. and Roskilly T. 2015 Building model calibration using energy and environmental data, *Energy and Buildings* 94, pp 109-120, doi: https://doi.org/10.1016/j.enbuild.2015.02.050

[8] Vines, R. L., Krutz R. Dean, and Krutz R. L.. *Cloud security: A comprehensive guide to secure cloud computing*. Wiley Publishing, Inc., 2010.

[9] Singh A and Chatterjee K. 2017 Cloud security issues and challenges: A survey, *Journal of Network and Computer Applications*, 79, pp 88-115, doi: https://doi.org/10.1016/j.jnca.2016.11.027.

[10] Ivanova E, Iliev T and Stoyanov I. 2020 Modelling and simulation of big data networks, *7th International Conference on Energy Efficiency and Agricultural Engineering (EE&AE 2020)*, Ruse, Bulgaria, 2020, pp. 1-4

[11] Sundermeyer M, Schlüter R, Ney H 2012 LSTM Neural networks for language modeling. *interspeech*, doi: https://doi.org/10.1016/0165-6074(89)90269-X.

[12] Stoyanov I., Iliev T., Mihaylov G., Evtatieve B. and Sokolov S. 2018 Analysis of the cybersecurity threats in smart grid University of telecommunications and post, Sofia, Bulgaria, *IEEE 24th International Symposium for Design and Technology in Electronic Packaging (SIITME)*, Iasi, 2018, pp 90-93, doi: 10.1109/SIITME.2018.8599261.
[13] Cui Z, Ke R, Pu Z and Wang Y 2018 Deep bidirectional and unidirectional LSTM recurrent neural network for network-wide traffic speed prediction, *arXiv preprint arXiv:1801.02143*.

[14] Vlaev S, Sokolov S and Vukadinoff M 2018 Scaling and optimizing cloud storage, *International Journal “Information Theories and Applications*, 25, no 3, pp 280-289

[15] Sokolov S, Idiriz O, Vukadinoff M and Vlaev S 2020 Scaling and automation in cloud deployments of enterprise applications, *Journal of Engineering Science and Technology Review Special Issue on Telecommunications, Informatics, Energy and Management*, pp 103 - 106, ISSN: 1791-2377, 2020

[16] Weil S 2007 Ceph: reliable, scalable, and high-performance distributed storage, *PhD diss.*, University of California, Santa Cruz

[17] Merkel D 2014 Docker: lightweight Linux containers for consistent development and deployment, *Linux journal*, no. 239: 2

[18] Sokolov S, Iliev T and Stoyanov I 2019 Analysis of cybersecurity threats in cloud applications using deep learning techniques, *42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, Opatija, Croatia, 2019, pp 441-446, doi: 10.23919/MIPRO.2019.8756755.

[19] https://github.com/zhiyongc/Stacked_Bidirectional_Unidirectional_LSTM [last accessed September 2020]

**Acknowledgments**

This work has been sponsored by Grant NID-1/17.04.2020 - “Implementation and adoption of cloud services for the teaching process of the University of Telecommunications and Post”. 