

High-speed stream data collection and processing system of the Earth's ionospheric sounding

V V Kopytov, A O Shulgin, N G Demurchev, P V Kharechkin and
V V Naumenko

“Infocom-S”, LLC, 7, Suvorova Street, Stavropol, 355035, Russian Federation

E-mail: p.harechkin@infocom-s.ru

Abstract. The paper describes the structure of the Earth's ionospheric sounding system which allows to determine the geographic coordinates of ionospheric regions with intense small-scale structures of electron content by total electron content data analysis. The system is based on the high-speed stream data collection and processing technology that performs sensed data preprocessing and scheduled data processing, including the estimation of total electron content in the ionosphere by distributed scalable computer cluster using a stack of open-source services.

1. Introduction

Nowadays Earth sensing and industrial monitoring provide important information about physical and environmental conditions which is used in various fields of human activity. One of these sensing services is near-Earth space sensing based on the estimation of total electron content (TEC) in the ionosphere obtained from signals sent by global navigation satellite system GLONASS and global positioning system GPS [1]. Data on Earth's ionospheric TEC are of interest to many technical and scientific applications. Global maps of TEC distribution are used to evaluate current radio-wave propagation conditions. Besides, many research studies have analyzed the impact of major seismic events on the ionosphere.

It's a fact that intense phase and amplitude noise in received radio signals can appear in transionospheric radio channel and causes reduction in the quality of global navigation satellite systems (GNSS) [2]. The reason is forming different regions in the ionosphere with small-scale structures of electron content that have natural (polar and equatorial latitudes) or artificial origin. Therefore, this feature poses an important task of TEC sensing for the following GNSS quality prediction.

In this paper, we propose the structure of the Earth's ionospheric sounding system which allows to determine the geographic coordinates of ionospheric regions with intense small-scale structures of electron content by the analysis of ionospheric TEC time series data obtained from dual-frequency GNSS receivers.

2. Methods and algorithms

The paper [3] describes the technique and the algorithm that determine the geographic coordinates and assess the linear dimensions of ionospheric regions with intense small-scale disturbances based on the ongoing TEC, angle of elevation and azimuth received from GNSS.

The technique consists of the following steps:

- computing TEC time series data using the values of phase change of the radio signal carrier frequency;



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

- smoothing TEC time series data by simple moving average shifting forward 60 seconds period, which matches $n = 3000$ data points at discretization interval of 0.02 seconds;
- computing the time series data of small-scale structures of TEC;
- smoothing the time series data of small-scale structures of TEC by simple moving average shifting forward 0.1 seconds period, which matches $n = 5$ data points at discretization interval of 0.02 seconds;
- computing the average value of small-scale structures of TEC at time interval of 60 seconds, where the discretization interval of each time interval is 0.02 seconds;
- computing the mean square deviation of small-scale structures of ionospheric TEC for each time interval;
- comparing the value of mean square deviation with the threshold value that characterizes the presence of small-scale disturbances;
- determining the geographic coordinates of a subionospheric point.

Figure 1 shows the diagram of determining the geographic coordinates of ionospheric regions with intense small-scale structures.

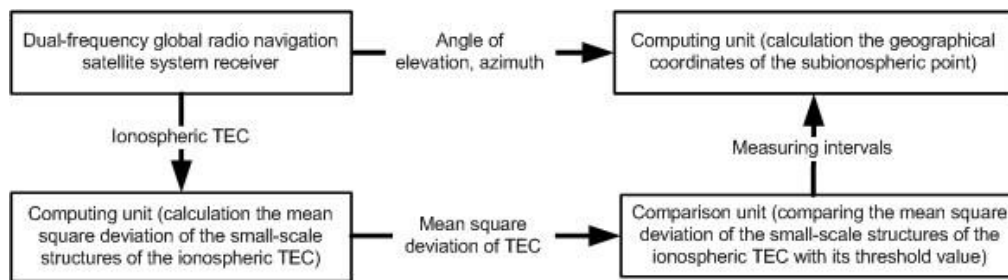


Figure 1. Diagram of determining the geographic coordinates of ionospheric regions with intense small-scale structures.

As presented in [4,5], the frequency of calculation the mean square deviation of TEC should be at least 50 Hz or higher to detect the small-scale disturbances in the ionosphere. Therefore, to build the information system that is able to sound the ionosphere, the development of special architectural and technological solutions is needed to perform high-speed stream data collection and data processing obtained from disturbed GNSS receivers.

3. The structure of the Earth's ionospheric sounding system

The proposed high-speed stream data collection and processing system combines the advantages of both approaches, the cloud computing and data preprocessing by interface modules, and allows to build an optimal solution in terms of performance, security, scalability and cost minimization.

The main tasks of the system for stream collection and processing of the Earth's ionospheric sounding data are as follows:

- obtaining data from a GNSS receiver to monitor the ionosphere;
- dividing the data into parallel streams and their preliminary processing;
- recording data streams into a time-series database;
- processing scheduled data to determine the geographic coordinates of a subionospheric point;
- visualizing the data processing results.

The architecture of suggested high-speed stream data collection and processing system consists of two basis components, a data processing center (DPC) and integrated monitoring devices (GNSS receivers).

A data processing center is a high-speed data collection and processing computer cluster that is part of high-performance computer clusters [6] and is used to accomplish the following tasks:

- use of various techniques to process sensor data streams stored in a time series database;

- scheduled data processing;
- interactive management of sensor data processing.

The architecture of a typical computer cluster is a group of joint operations servers sharing the same data warehouse. Depending on the task, the high-speed data collection and processing computer cluster shall also include software for collecting data from diverse measuring devices.

Therefore, the architecture for a high-speed data collection and processing computer cluster can be described as four interacting base subsystems:

- sensor data collection subsystem;
- sensor data processing subsystem;
- data storage subsystem;
- cluster management subsystem.

The sensor data collection subsystem is a bridge between sensor device terminals and data processing and storage components and provides a common interface for their connection to a computer cluster, i.e. an enterprise service bus, by means of device drivers or external monitoring systems. This study understands device drivers as software serving as a link between the enterprise service bus and a device (or an external monitoring system) and ensuring the compliance of the device's interface with that of the enterprise service bus.

The sensor data processing subsystem comprises tools allowing users to adopt various algorithms for sensor data stream processing, which are stored in a time series database, and for scheduled data processing.

A data warehouse should also meet the requirements to ensure horizontal scalability of DPCs depending on the number of connected sensors and the amount of sensor data requiring the application of the distributed data storage technology.

The cluster management subsystem provides mechanisms for interactively managing sensor data processing and, specifically, computing task scheduling.

Given this architecture, research was conducted on the following software classes to achieve the objectives defined:

- software designed to aggregate data from diverse devices;
- systems for managing distributed databases;
- software for unfolding and managing distributed data processing.

The choice of software is based on the analysis of state-of-the-art solutions in terms of their significance to achieve the stated objectives [7].

The proposed solution is based on the technology of distributed data storage and data processing, Apache Hadoop [8], that uses the resource manager Yarn as a computer cluster's resource management system. Below is the content and functionality of computer cluster program components ensuring high-speed data collection and processing in compliance with the architectural and technological solutions chosen in this study.

The following are the components of the diagram above:

- information sensors;
- intermediary data processing servers;
- Apache Kafka message broker;
- stream processing services;
- scheduled data processing services;
- distributed time-series database;
- Apache Zeppelin, a web-based notebook to work with Big Data;
- Apache Hue, a stream scheduling service for deferred computing;
- Grafana, a tool for metric analytics.

Figure 2 shows a data flow diagram of a computer cluster ensuring high-speed data collection and processing.

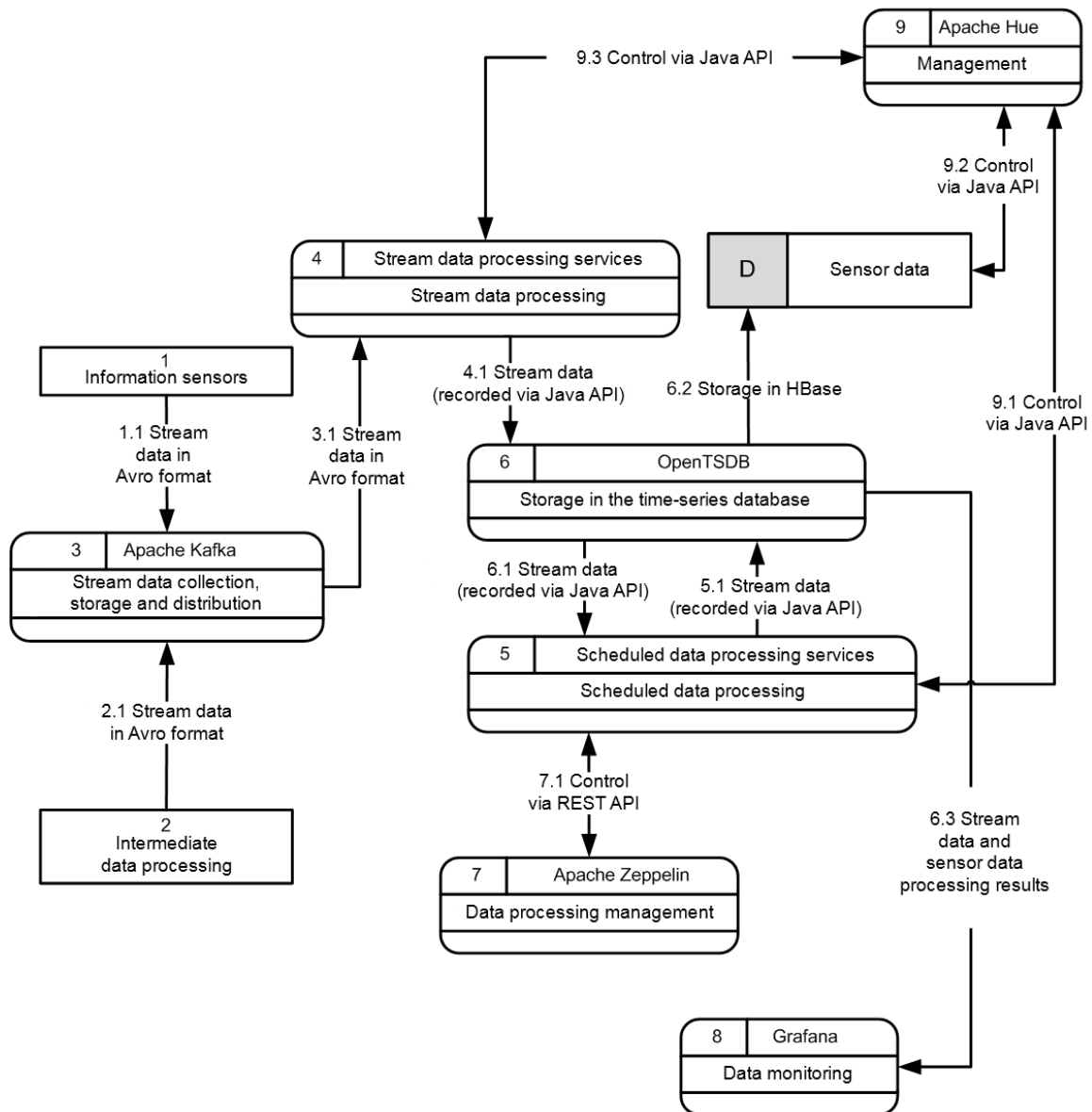


Figure 2. Data flow diagram of a computer cluster to ensure high-speed data collection and processing.

Sensor device drivers are connected to a single enterprise service bus based on the Apache Kafka message broker, shaping stream data in Avro format. Among Apache Kafka's functions are transmission, storage and distribution of data received.

Sensor data stream processing services and scheduled data processing services are the component parts of applications used in task analysis, projection or predictive modeling, to be discussed later.

Sensor data stream processing services obtain data directly from Apache Kafka, operate inside the Apache Spark platform and use the Spark Streaming library.

Scheduled data processing services interact with OpenTSDB, a scalable time-series database, operate inside the Apache Spark platform and use the Spark SQL library.

The interface for computer cluster management and for displaying sensor data and their processing results is based on Grafana, a tool using Varnish to have access to OpenTSDB.

The web-based notebook Apache Zeppelin supports remote program control for scheduled processing services in Apache Spark via Livy, a web service for Spark.

Apache Hue, a tool that interacts with the time-series database and Apache Spark, is used to schedule workflows for deferred computing and data analysis in Apache Hadoop.

The hardware and software system for high-speed stream collection and processing of the Earth's ionospheric sounding data comprises the following:

- three 1480Q1 Depo Storm computing servers;
- NovAtelGPSStation-6, a dual-frequency GNSS receiver;
- interface module based on Raspberry Pi 3 Model B.

The driver of a GNSS receiver is installed on the interface model with a view to read and pre-process monitoring logs and, then, to send the data to Apache Kafka, the enterprise service bus, at a speed of 0.8 Mbps. Fifty-six satellites collect data at a speed of 50 values per second for each data stream.

Data stream processing covered an area of 4,000 geographic coordinates. Scheduled processing aimed at calculating the geographic coordinates of TEC focused on an area of 180,000 geographic coordinates.

Grafana, a tool for metric analytics that is part of computer cluster program components (see figure 3), was used to visualize the data processing results and the satellites' current location.



Figure 3. Ionospheric condition monitoring system according to the satellites' current location and trajectory.

The interface modules preprocess stream data by computing the mean square deviation of small-scale structures of ionospheric TEC and the mean square deviation of wavefront phase fluctuations,

which are needed for the following analysis of near-Earth space parameters and identification the subionospheric point with small-scale disturbances.

4. Conclusion

The proposed architectural solutions for organizing a high-speed data collection and processing computer cluster fully support the basic functionality and the localization of program modules using the data processing and modeling methods and algorithms of near-Earth space sensing. What is special about the proposed architecture is the application of technological solutions ensuring high capacity and unlimited horizontal scaling. Further research leads to applying the developed architecture to other fields of activity that are in need of high-speed stream sensor data collection and processing (e.g., situation management systems).

Acknowledgments

This study has been carried out as part of the research project implementation on “Development of high-speed data processing facilities for data sensors in situation management systems” within the framework of Research and Development Federal Targeted Program 2014-2020 (unique identifier RFMEFI57916X0135) funded by the Ministry of Education and Science of the Russian Federation.

References

- [1] Sardón E and Zarraoa N 1997 Estimation of total electron content using GPS data: How stable are the differential satellite and receiver instrumental biases? *Radio Sci* **32** (5) 1899-910
- [2] Davies K 1969 *Ionospheric radio waves* (Blaisdell Pub Co) p 460
- [3] Pashintsev V P, Chipiga A F, Tsymbal V A and Peskov M V 2016 The complex definition of the field with small-scale ionospheric irregularities according gps-monitoring *Bulletin of Samara Scientific Center of the Russian Academy of Sciences* **18** (2) 941-5
- [4] Katkov K A, Pashintsev V P, Katkov E K, Gakhova N N, Gakhov R P and Titov A I 2017 Forecast accuracy of determining pseudo range in satellite navigation system through analysis of data from ionosphere monitoring *Journal of Fundamental and Applied Sciences* **9** (1S) doi:10.4314/jfas.v9i1s.744
- [5] Katkov K A, Pashintsev V P and Katkov E K 2016 Information ionosphere monitoring system *Bulletin of Samara Scientific Center of the Russian Academy of Sciences* **18** (2) 907-12
- [6] Furht B and Escalante A 2010 *Cloud Computing Fundamentals Handbook of Cloud Computing* (Springer) p 634
- [7] Díaz M, Martín C and Rubio B 2016 State-of-the-Art, Challenges, and Open Issues in the Integration of Internet of Things and Cloud Computing *Journal of Network and Computer Applications* **67** (C) 99-117
- [8] Kałużka J, Napieralska M, Romero O and Jovanovic P 2017 Data Locality in Hadoop *International Journal of Microelectronics and Computer Science* **8** (1) 16-20