2021 XVII International Symposium Problems of Redundancy in Information and Control Systems (REDUNDANCY) | 978-1-6654-3308-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/REDUNDANCY52534.2021.9606477

Gateway Data Encoding, Packaging and Compression method for heterogeneous IoT-satellite network

1st Leonid Voskov
National Research University Higher School of Economics
Moscow, Russia
lvoskov@hse.ru

2nd Alexei Rolich
National Research University Higher School of Economics
Moscow, Russia
arolich@hse.ru

3rd Gleb Bakanov
National Research University Higher School of Economics
Moscow, Russia
gpbakanov@edu.hse.ru

4th Polina Podkopaeva
National Research University Higher School of Economics
Moscow, Russia
popodkopaeva@edu.hse.ru

***Abstract*—Reducing the cost of data transmission is actively pursued all over the world. A separate direction in this area is the study of possibilities to reduce the cost of transmission of messages via a satellite communication channel. However, studies of the possibility of reducing the cost of message transmission in heterogeneous networks using satellite communication channels in the context of an undeveloped terrestrial network infrastructure and a remote Internet of things have not yet been carried out. This paper reviews and analyzes protocols and technologies for transferring Internet of Things (IoT) data and presents an architecture for a hybrid IoT-satellite network, which includes a long range (LoRa) low power wide area network (LPWAN) terrestrial network for data collection and an Iridium satellite system for backhaul connectivity. Simulation modelling, together with a specialized experimental stand, allowed us to study the applicability of different methods of information presentation for the case of transmitting IoT data over low-speed satellite communication channels. We proposed a data encoding, compressing, and packaging scheme called GDEPC (Gateway Data Encoding, Packaging and Compressing). It is based on the combination of data format conversion at the connection points of a heterogeneous network, message compressing and packaging. GDEPC enabled the reduction of the number of utilized Short Burst Data (SBD) containers and the overall transmitted data size by almost fifteen times.**

***Index Terms*—Coding; IORT; Satellite communications; Compression; Packaging**

I. Introduction

The current trends in the field of the Internet of Things (IoT) are the creation of new hardware platforms, the development of specialized operating systems, the solution of problems in the field of cybersecurity, the development and optimization of methods and algorithms aimed at increasing the energy efficiency of IoT devices to increase their battery life.

In areas where there is no cellular communication at all, satellite communication systems SATCOM are used to collect and exchange data, such systems are called the Internet of Remote Things (IoRT) [1]. Integrated communication systems and heterogeneous networks [2-17] are intensively developing for data collection using SATCOM. These networks combine satellite and terrestrial networks. The generally recognized limitations in data collection systems using SATCOM are the high cost of transmitting messages from sensors and low bandwidth (capacity) of communication channels, which does not allow widespread implementation of SATCOM for connecting IoT devices

In previous articles [18-22], we noted the problems when deploying data collection networks in regions without any existing infrastructure, several areas of application of the Internet of Remote Things. In [21] we proposed a data encoding and packaging scheme called GDEP (Gateway Data Encoding and Packaging). It is based on the combination of data format conversion at the connection points of a heterogeneous network and message packaging. GDEP enabled the reduction of the number of utilized Short Burst Data (SBD) containers and the overall transmitted data size by almost five times.

Reducing the cost of data transmission is actively pursued all over the world. A separate direction in this area is the study of possibilities to reduce the cost of transmission of messages via a satellite communication channel. However, studies of the possibility of reducing the cost of message transmission in heterogeneous networks using satellite communication channels in the context of an undeveloped terrestrial network infrastructure and a remote Internet of things have not yet been carried out.

In addition to the obvious technical difficulties associated with placing sensors in an uncontrolled aggressive environment, the key problems in satellite IoRT are the high cost of sending messages and the occurrence of collisions and significant delays in the collection and transmission of data. As a result, many algorithmic problems are still being investigated, and the authors present new methods of access to channels, modulation methods, routing mechanisms, load balancing, algorithms and protocols of data transmission and methods of increasing the network throughput, as well as schemes for reducing the energy consumption of the network infrastructure and approaches to reducing the cost of message transmission. In addition, new architectures have appeared of heterogeneous Internet of Things networks using satellite communication channels, including those based on the technology of energy efficient long-range networks (LoRa) and satellite

communication channel Iridium.

Thus, all the works are exploring the possibilities of satellite communications for the IoT, formulating many open questions and solving some of them. A potential gap in the optimization of data transmission at the border of two parts of the network - terrestrial and satellite - can be identified.

The cost of transmitting a message in a heterogeneous network using a satellite communication channel remains the main requirement for data transmission on the Internet of remote things, since it directly affects the distribution and implementation of these technologies in real industrial sectors. One of the factors affecting the cost of transmitting a message in a heterogeneous network using a satellite communication channel is the amount of data transmitted or the ratio between overhead and payload. Optimization of this ratio can be carried out by changing the existing network algorithms and protocols. In a heterogeneous structure, new algorithms and approaches can be applied at the border of two networks, which is implemented in this work, and which is the main distinguishing feature in comparison with other existing ones.

One of the ways to reduce the cost of data transmission is to compress the transmitted data. Reducing the volume of transmitted data by 7.5-14 times makes it possible to use satellite communications in hard-to-reach areas at rates comparable to those of cellular operators and significantly expand the use of Internet of Things technologies.

Thus, the problem arises of reducing the cost of transmitting messages from IoT devices located in an undeveloped network infrastructure to a target data collection system through the development of models, methods and algorithms for serialization, packaging and compression of messages in heterogeneous LoRa-Iridium networks.

II. Literature review

The development of cellular networks (GSM), the introduction of 5G, NB-IoT technologies, expands the service area and leads to the growth of IoT-connected devices using SMS for data transmission. The cost of data transmission using SMS messages depends on the region, country, telecom operator, and tariff and may differ significantly.

In Europe [23], the tariff for one text SMS - 160 bytes costs 0.03-0.089 eurocents ($0.00020- $0.00061 per one byte). In Russia [24], the tariff for one SMS transmission is 3.8-8 rubles ($ 0.00037 - $ 0.00078) for one byte.

International satellite operators Iridium considered in the review [25], has a tariff plan for the transmission of short messages (Iridium SBD) set at $0.0015 per byte. On the territory of Russia, the satellite operator has a short message transmission (SBD) tariff plan defined in the range of $2.74 to $0.18 per kilobyte ($0.0028- $0.00018 per byte), the subscription fee for servicing the device is $6.40 (traffic 1-10KB), $ 22.86 (traffic is more than 50KB) per month.

Table 1 shows the tariffs for the transmission of messages from GSM and SBD Iridium telecom operators per one byte. Tables 2 and 3 show the range of the ratio of tariffs for GSM and SBD Iridium telecom operators per one byte.

TABLE I
Tariffs for transferring short messages from telecom operators per byte

Service provider	**SMS GSM**		**SBD Iridium**	
Tariff	Min	Max	Min	Max
International	$0,00020	$0,00061	$0,0015	$0,0028
Russian	$0,00037	$0,00078	$0,00018	$0,0028

TABLE II
Maximum ratio of GSM and SBD Iridium tariffs per one byte

Service provider	**SMS GSM**	**SBD Iridium**	**Ratio**
Tariff	Min	Max	SBD Iridium/SMS GSM
International	$0,00020	$0,00028	14,00
Russian	$0,00037	$0,00078	7,56

Analysis of Table 2 shows that the ratio of tariffs for the transmission of one byte of data using satellite SBD Iridium and cellular GSM communication channels is in a wide range from 0.23 to 14. In the worst case, this means that the transfer of one byte of information via satellite communication channels is more expensive in 14 times in Europe and 7.5 times more expensive in Russia than using SMS GSM channels of cellular operators. In areas where cellular communication is completely absent, SATCOM satellite communication systems are used to collect and exchange data. The generally recognized limitations in data collection systems using SATCOM are the high cost of transmitting messages from sensors and low bandwidth (capacity) of communication channels, which does not allow widespread implementation of SATCOM for connecting IoT devices.

When transmitting data over low-speed channels through the LEO satellite constellation, preliminary data compression is of great importance since it allows to reduce the amount of transmitted data and reduce the cost of satellite communication services.

III. Features of data from cyberphis. systems in IoRT

Devices of the Internet of remote things are considered, transmitting short messages from sensors in the LoRaWAN network [2] up to 500 bytes in size. In the absence of terrestrial communications, messages are transmitted over the Iridium satellite channel. Thus, data transmission is carried out in a heterogeneous LoRa-Iridum network [2] in small packets (Short Burst Data, SBD) up to 1960 bytes [3] with a low speed of 1200 bit / s on average [3], a delay of 5-20 seconds

TABLE III
Minimum ratio of GSM and SBD Iridium tariffs per one byte

Service provider	**SMS GSM**	**SBD Iridium**	**Ratio**
Tariff	Max	Min	SBD Iridium/SMS GSM
International	$0,00061	$0,0015	2,45
Russian	$0,00078	$0,0018	0,23

[4] and high cost of satellite communication services (from $1.4 / Kb) [5]. In this regard, the problem arises of the high cost of transmitting a data packet in the considered heterogeneous LoRa-Iridium network, which can be solved by using methods and algorithms for serializing and compressing data transmitted in small packets.

IV. Gateway Encoding and Packaging Method (GDEPC)

Data from cyber-physical systems can come in various formats (JSON, GeoJSON, CSV, XML, etc.). To reduce the amount of data, each structure is converted into a sequence of bytes [11], the data is serialized. One of the effective serialization tools [12] is Protocol Buffers (Google), which allows, for example, to reduce the amount of transmitted data in the JSON format by 4.6 times [13]. Protobuf messages are packaged in small packets limited by the maximum SBD container size for the Iridium 9602 modem (340 bytes).

After filling the container, the data can be reduced in volume using compression algorithms. After compression, the container is filled with Protobuf messages again and the contents are compressed. Steps 2 and 3 (fig. 1) are repeated until the container is completely full. The algorithm scheme is presented below.

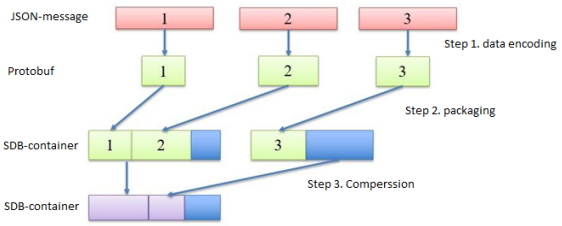

Fig. 1. GDEPC algorithm.

Messages coming from cyber-physical systems have a size of 30-2048 bytes, which does not allow using well-known archivers for compression (cmix, paq8hp12, nanozip, gzip, bzip2, and 7zip), since the size of meta-information when compressed is comparable to or exceeds the size of the original message. Let us formulate the problem taking into account the above initial conditions.

The real object is a queue of Protobuf messages up to 500 bytes in size, the property is the message size in bytes. Also, a real object is a container of 340 bytes in size (the maximum size of an outgoing message for the Iridium SBD 9602 modem), its property is its capacity (the amount of storage in bytes). The challenge is to reduce message volume and containerize.

To solve this problem, a comparative analysis of various compression algorithms was carried out for one hundred short messages (URL-sequences) in the range from 30 to 2048 bytes (see Experimental results). When using the well-known archivers cmix, paq8hp12, nanozip, gzip, bzip2 and 7zip, small data (30-45 bytes) are not compressed during archiving [14]. When archiving using the Huffman algorithm, data is compressed by an average of 1.3 - 1.5 times over the entire range (30-2048 bytes).

For short and medium messages, the highest compression ratio is obtained by the Huffman entropy algorithm, the volume of messages is reduced by up to 1.6 times. Huffman compression takes into account the frequency of individual characters in the coding dictionary, but does not take into account the frequency of occurrence of the aforementioned phrases. However, the redundancy of information in the messages under consideration is primarily associated with repeated from message-to-message field names (keys) and reserved words of the Protocol buffers format. Field names are quite common (due to the nature of the data (reports) received from the same sensors). This feature is called linguistic similarity and must be taken into account in the algorithm to achieve the best compression ratio.

Therefore, it is proposed to expand the frequency vocabulary from the standard Huffman algorithm (it consists of all the different symbols of the initial data array) with those words (phrases) that are encountered quite often. It is assumed that the coding of such words with a binary code of the optimal length in accordance with the priority queue principle will make it possible to develop an algorithm for coding and decoding the messages under consideration with a higher compression ratio than when coding by the Huffman method.

To implement the above approach, it is necessary to develop strict criteria for phrases to enter the dictionary and their subsequent encoding. Mathematical foundations of Huffman coding [26] require that the frequency of occurrence of a character (in our case, a phrase) exceed the Poisson mathematical expectation, and the Poisson probability is the lowest. In this case, the selected phrases will be the most dependent, and the resulting dictionary will be optimal.

The Poisson probability is calculated as follows:

$$\lambda = \frac{P(a_1) \cdot P(a_2) \cdot ...P(a_n)}{nsymbolsofcombination} \cdot m, \quad (1)$$

where a_1-a_n elements of symbol combination (phrase), $P(a_1)$ - $P(a_n)$ the probability of these characters appearing in the text, m the number of combinations in the text of n elements.

The probability of the occurrence of this phrase is calculated by the formula:

$$P(\lambda) = \frac{\lambda^k \cdot \exp(-\lambda)}{k}, \quad (2)$$

where k is the number of necessary combinations (phrases) in the text.

The creation of an extended dictionary (hereinafter referred to as the ED) begins with the alphabet of the original array of messages. It is proposed to create a dictionary for selecting many messages and pass it once. Otherwise, the compression of short messages, as was shown on the example of archivers, does not make sense (the compression ratio is less than or equal to one). Then a frequency dictionary is created (hereinafter referred to as an FD) for all field names and reserved words. Next, the weight of each remaining word from the

FD is calculated: the frequency of occurrence of each word is multiplied by its length. Checking the above criteria for adding words (phrases) from the FD dictionary to the ED dictionary goes from the words with the highest weight to the words with the lowest weight. Two necessary conditions are checked: the frequency of occurrence and the Poisson probability, but it is also necessary to evaluate how the length of the encoded sequence changes when the dictionary is expanded with a new phrase. This check is necessary due to the fact that adding a new character to the dictionary (in accordance with its weight) changes the dictionary character codes, therefore the tree constructed in accordance with the priority queue principle and, accordingly, the encoded sequence change. Therefore, when creating an extended dictionary, it is proposed to save the result of the i-th and i-1th iterations, comparing the encoding result before and after adding a new phrase to the dictionary. The dictionary is expanded as long as the addition of new words reduces the volume of the compressed message relative to the previous step. Thus, in a finite number of iterations, an optimal dictionary with a compression ratio will be found that takes into account the linguistic similarities of messages and is guaranteed (due to necessary and sufficient conditions) to compress no worse than the classical Huffman algorithm.

V. Experimental results

An experiment was conducted to test the archiving stage of the proposed method. As input data, 100 short messages up to 2 GB were used, consisting of URL sequences (for example, the sequence https://wikimapia.org/ 22 bytes in size). The above-mentioned well-known, most effective archivers participated in the experiment (table 5,6). To compile an extended dictionary, the approach described in the GDEPC section was used. Thus, frequently occurring combinations, such as https://, were encoded with the smallest number of characters in accordance with the principle of building a prefix tree for elements in the priority queue. The results are presented in the tables below. The algorithm "Huffman with extended dictionary" is described in IV. GATEWAY ENCODING AND PACKAGING METHOD (GDEPC).

TABLE IV
Analysis of compression algorithms (data of a small volume of 30-45 bytes)

Arch. program/compr. algorithm	**Compression ratio** $K = \frac{N_{in}}{N_{out}}$
cmix	0,86
paq8hp12	0,62
nanozip	0,34
gzip	0,56
bzip2	0,51
7zip	0,24
Huffman	1,33
Huffman with extended dictionary	2,13

The compression ratio is the ratio of the input data volume to the output data. Thus, the proposed Huffman algorithm with an extended dictionary made it possible to reduce the volume by 2-3.6 times in the entire range of 30-2048 bytes of data transmitted in a heterogeneous LoRa-Iridium network with a common linguistic feature.

TABLE V
Analysis of compression algorithms (data of a small volume of 45-2048 bytes)

Arch. program/compr. algorithm	**Compression ratio** $K = \frac{N_{in}}{N_{out}}$
cmix	2,4
paq8hp12	2,4
nanozip	1,9
gzip	1,8
bzip2	1,9
7zip	1,7
Huffman	1,6
Huffman with extended dictionary	3,6

Based on the conducted studies, it can be concluded that the GDEPC method allows to reduce the amount of data with a common linguistic feature in the range of 30-2048 bytes by 10-17 times (by 4.6 times due to serialization and by 2.13-3.6 times due to archiving by the proposed method). The data volume reduction (compression ratio) depends on the amount of data being transmitted.

VI. Conclusion

As a result of this work, methods of coding and compression of information were identified that allow solving the problem of collecting information from remote sensors located in the absence of traditional communication channels, and a practical check of the results obtained was obtained. The article uses simulation modeling to investigate the applicability of the Huffman method for compressing information in the case of IoT data transmission over low-speed satellite communication channels.

We see the following main contributions from this article:

1) A new method (GDEPC) was developed for encoding and compressing information at the OSI representation level of a heterogeneous IoT network, which, by combining several methods of encoding and packaging, increases the efficiency of data transmission by 13 times.
2) The proposed GDEPC method was tested on a simulation model and experimental equipment.

The GDEPC method proposed in the article allows the use of the IoT technology stack in remote regions, integrating it with the SBD satellite short message service. The GDEPC method made it possible to reduce the volume and number of SBD messages when transmitting data over low-speed satellite communication channels, which made it possible to reduce the cost of transmitting one data packet at the cost of an SMS message. Reducing the cost of data transmission leads to an increase in the economic efficiency of SATCOM for organizing data transmission networks in remote areas without telecommunications infrastructure.

VII. Future works

The Internet of Remote Things (IoRT) is a recently introduced paradigm describing monitoring and control networks in hard-to-reach locations. These networks usually have very limited bandwidth and tend to be heterogeneous, which paves the way for new models, techniques and algorithms to optimize data transmission, energy efficiency and traffic balancing. We have identified several challenges that need to be addressed within the IoRT framework that can benefit from the results of this paper.

We believe that the GDEPC method can be further improved and generalized for the satellite IoT scenario with any combination of raw data transmission technologies, encoding formats and container sizes. To form dictionaries for compression, neural networks and machine learning technologies can be used, which will allow the GDEPC method to be used for any heterogeneous LoRa-Iridium network. This will require additional modeling and experimental research.

The developed GDEPC method can be implemented based on hardware, embodying the concept of fog computing directly on the LoRa-Iridium network gateway.

References

[1] M. De Sanctis, E. Cianca, I. Bisio, G. Araniti, R. Prasad, Satellite Communications Supporting Internet of Remote Things. IEEE Internet Things J. 2015, vol. 3, pp. 113–123

[2] Sigfoxs CTO on Where Satellite Fits in an IoT-Only Network. Available online: https://spacenews.com/sigfoxs-cto-on-where-satellite-fits-in-an-iot-only-network/ (accessed on 27 December 2018)

[3] Z. Qu, G. Zhang, H. Cao, J. Xie, LEO satellite constellation for Internet of Thing. IEEE Access 2017, vol. 5, pp.1839118401.

[4] A. Augustin, J. Yi, T. Clausen, W.M. Townsley, A Study of LoRa: Long Range & Low Power Networks for the Internet of Things. Sensors 2016, vol.16, pp.1466.

[5] H. Huang, S. Guo, W. Liang, K. Wang, Online Green Data Gathering from Geo-Distributed IoT Networks via LEO Satellites. In Proceedings of the IEEE International Conference on Communications (ICC), Kansas City, MO, USA, 2024 May 2018; pp. 16.

[6] T. Ferrer, S. Cspedes, A. Becerra, Evaluation of MAC protocols for IoT satellite systems. In Proceedings of the IV School of Systems and Networks (SSN 2018), Valdivia, Chile, 2931 October 2018; pp. 5760. [Google Scholar]

[7] M. Gineste, T. Deleu, M. Cohen, N. Chuberre, V. Saravanan, V. Franscolla, M.Mueck, E.C. Strinati, E. Dutkiewicz, Narrowband IoT Service Provision to 5G User Equipment via a Satellite Component. In Proceedings of the 2017 IEEE Globecom Workshops (GC Wkshps), Singapore, 48 December 2017; pp. 14.

[8] B. Evans, O. Onireti, T. Spathopoulos, M.A. Imran, The role of satellites in 5G. In Proceedings of the 2015 23rd European Signal Processing Conference (EUSIPCO), Nice, France, 31 August4 September 2015; pp. 27562760.

[9] Z. Liu, J. Li, Y. Wang, X. Li, S. Chen, HGL: A hybrid global-local load balancing routing scheme for the Internet of Things through satellite networks. Int. J. Distrib. Sens. Netwo. 2017, 13

[10] Y. Kawamoto, H. Nishiyama, Z.M. Fadlullah, N. Kato, Effective Data Collection Via Satellite-Routed Sensor System (SRSS) to Realize Global-Scaled Internet of Things. IEEE Sens. J. 2013, 13, 36453654

[11] Y. Qian, L. Ma, L.; Liang, X. Symmetry Chirp Spread Spectrum Modulation Used in LEO Satellite Internet of Things. IEEE Commun. Lett. 2018, vol. 22, pp. 22302233.

[12] A. Muthukrishnan, J. Charles Rajesh Kumar, D. Vinod Kumar, M. Kanagaraj, Internet of image things-discrete wavelet transform and Gabor wavelet transform based image enhancement resolution technique for IoT satellite applications. Cogn. Syst. Res. 2019, vol.57, pp.4653.

[13] P. Li, G. Cui, W. Wang, Asynchronous Flipped Grant-Free SCMA for Satellite-Based Internet of Things Communication Networks. Appl. Sci. 2019, vol.9, pp.335.

[14] M. Bacco, M. Colucci, A. Gotta, Application protocols enabling internet of remote things via random access satellite channels. In Proceedings of the 2017 IEEE International Conference on Communications (ICC), Paris, France, 2125 May 2017; pp. 16.

[15] Y. Song, B. Song, Z. Zhang, Y. Chen, The Satellite Downlink Replanning Problem: A BP Neural Network and Hybrid Algorithm Approach for IoT Internet Connection. IEEE Access 2018, vol.6, pp.3979739806.

[16] O. Said, A. Tolba, Performance Evaluation of a Dual Coverage System for Internet of Things Environments. Mob. Inf. Syst. 2016, vol.2016, pp.120.

[17] I. F. Akyildiz, A. Kak, The Internet of Space Things/CubeSats: A ubiquitous cyber-physical system for the connected world. Comput. Netw. 2019, vol.150, pp.134149.

[18] S. Efremov, N. Pilipenko, L. Voskov, An Integrated Approach to Common Problems in the Internet of Things, Procedia Engineering, vol. 100, 2015, pp.1215-1223

[19] A. R. Fakhrazeev, A. Y. Rolich and L. S. Voskov, "Big telemetry data processing in the scope of modern Internet of Things," 2018 Moscow Workshop on Electronic and Networking Technologies (MWENT), 2018, pp. 1-4

[20] I. I. Lysogor, L. S. Voskov, A. Y. Rolich and S. G. Efremov, "Energy efficient method of data transmission in a heterogeneous network of the Internet of things for remote areas," 2019 International Siberian Conference on Control and Communications (SIBCON), 2019, pp. 1-6

[21] I. Lysogor, L. Voskov, A. Rolich, S. Efremov, Study of Data Transfer in a Heterogeneous LoRa-Satellite Network for the Internet of Remote Things. Sensors 2019, vol.19, pp.3384

[22] I. I. Lysogor, L. S. Voskov and S. G. Efremov, "Survey of data exchange formats for heterogeneous LPWAN-satellite IoT networks," 2018 Moscow Workshop on Electronic and Networking Technologies (MWENT), 2018, pp. 1-5

[23] Sending SMS in Europe. Available Online: https://www.allmysms.com/en/international-bulk-sms-gateway/Europe

[24] All MTS Tariffs. Available Online: https://moskva.mts.ru/personal/mobilnaya-svyaz/tarifi/vse-tarifi/supermts

[25] M. Prior-Jones, Satellite communications systems buyers guide. British Antarctic Survey

[26] Proof of Optimality of Huffman Codes.University of Toronto: CSC373, 2009