

РАЗРАБОТКА АЛГОРИТМОВ СБОРА ДАННЫХ В РАМКАХ КОНЦЕПЦИИ ИНТЕРНЕТА ВЕЩЕЙ

Козлов А.С.¹, Дудник С.В.², Култазин Н.М.³, Ангапов В.Д.⁴, Гринер В.⁵

Email: Kozlov693@scientifictext.ru

¹Козлов Александр Сергеевич - старший системный администратор,
филиал

Корпорация "Алайн Текнолоджи Ресерч энд Девелопмент, Инк";

²Дудник Сергей Викторович - ведущий эксперт,
департамент инфраструктурных решений,
Сбербанк,
г. Москва;

³Култазин Нурлан Муратович - инженер инфраструктуры,
Astana International Exchange,
г. Нур-Султан, Республика Казахстан;

⁴Ангапов Василий Данилович - старший системный архитектор,
Digital IQ, г. Улан-Удэ;

⁵Гринер Вадим - главный инженер по качеству,
Red Hat, г. Модиин-Маккабим-Реут, Израиль

Аннотация: рассмотрены ограничения, связанные с развитием концепции Интернета вещей, инфраструктура которого базируется на наземных информационных сетях. Проведена систематизация типовых задач, характерных для построения аппаратно-программной базы спутникового Интернета вещей. Поставлена задача построения комплексной методологии работы со спутниковыми распределенными информационными системами в рамках концепции спутникового Интернета вещей на базе алгоритмов восстановления дискретных данных и технологии пространственно-временного компрессионного зондирования. Предложена модель, которая базируется на целевых функциях уровня сжатия данных, подлежащих сбору и точности сбора данных. Построена универсальная схема работы алгоритма восстановления дискретных данных для спутникового Интернета вещей, которая включает в себя такие функциональные элементы, как кластеризация входных данных, получение выборки и восстановление данных, которые не были получены.

Ключевые слова: спутниковый Интернет вещей, распределенная информационная система, беспроводная сенсорная сеть, алгоритм восстановления дискретных данных, пространственно-временное компрессионное зондирование, пропускная способность, выборка данных.

DEVELOPMENT OF DATA COLLECTION ALGORITHMS WITHIN THE FRAMEWORK OF THE INTERNET OF THINGS' CONCEPT

Kozlov A.S.¹, Dudnik S.V.², Kultazin N.M.³, Angapov V.D.⁴, Griner V.⁵

¹Kozlov Aleksandr Sergeevich - Sr. System Administrator,
BRANCH

"ALIGN TECHNOLOGY RESEARCH AND DEVELOPMENT INCORPORATED", EMEA RUSSIAN REGION;

²Dudnik Sergei Victorovich - Leading Expert,
DEPARTMENT OF INFRASTRUCTURE SOLUTIONS,
SBERBANK,
MOSCOW;

³Kultazin Nurlan Muratovich - Infrastructure Engineer,
ASTANA INTERNATIONAL EXCHANGE,
NUR-SULTAN, REPUBLIC OF KAZAKHSTAN;

⁴Angapov Vasily Danilovich - Senior systems Architect,
DIGITAL IQ, ULAN – UDE;

⁵Griner Vadim - Senior Quality Engineer,
RED HAT, MODIIN MACCABIM REUT, ISRAEL

Abstract: the limitations associated with the development of the concept of the Internet of things, the infrastructure of which is based on terrestrial information networks, are considered. The systematization of typical tasks characteristic of development the hardware and software platform of the satellite-based Internet of things is carried out. The task is to build an integrated methodology for working with distributed satellite information systems within the framework of the satellite-based Internet of things concept based on sampling-reconstruction algorithms and spatio-temporal compressive sensing technology. The model is based on the objective functions of the level of data compression to be collected and the accuracy of data collection. A

universal scheme of the discrete data recovery algorithm for the satellite-based Internet of things has been built, which includes such functional elements as clustering of input data, data sampling and data recovery.

Keywords: *satellite-based Internet of things, distributed information system, wireless sensor network, on sampling-reconstruction data algorithm, spatio-temporal compressive sensing, bandwidth, data sampling.*

УДК 004.75

Введение

Концепция Интернета вещей (IoT, Internet of Things) подразумевает объединение электронных устройств, которые могут быть рассмотрены в качестве аппаратно-программных платформ, характеризующихся такими параметрами как вычислительная мощность и ширина полосы передачи данных, в единую сеть. Тем не менее, в условиях масштабирования запросов, которые могут исходить от крупных корпораций и правительственных структур можно отследить ограничения, характерные для данной концепции, связанные с тем, что наземные информационные сети (НИС) неравномерно покрывают пространство планеты. Узлы НИС в значительной степени сосредоточены в городах и мегаполисах, но при этом данная инфраструктура отсутствует на территории лесов, пустынь, океанов, а также регионов Арктики и Антарктики. Поэтому для проектов связанных с добычей полезных ископаемых, экологическим мониторингом, координированием военных операций и экстремальным туризмом используется спутниковая передача данных. В рамках данной работы предлагается рассмотреть развитие концепции IoT до уровня спутникового Интернета вещей (S-IoT, Sattelite-based Internet of Things) [1-5]. Таким образом, **актуальность** исследования определяется как расширением базы спутниковых информационных сетей, так и экспоненциальным ростом запросов в рамках крупных проектов, что наблюдается в данной области последние два десятилетия [6, 7].

Анализ последних исследований и публикаций в области S-IoT показал перспективность развития алгоритмов сбора данных [8], которые базируются на подходах, которые используются в беспроводных сенсорных сетях (WSN, wireless sensor network) наземной инфраструктуры распределенных информационных систем (РИС). В частности, были рассмотрены следующие группы алгоритмов:

- алгоритмы, базирующиеся на модельном представлении (model-based algorithms) [9, 10];
- алгоритмы сжатого распознавания (compressed-sensing based algorithms) [11, 12];
- алгоритмы с доставкой по запросу (query-driven algorithms) [13, 14].

Проведенный анализ показал, что алгоритмы, базирующиеся на модельном представлении и алгоритмы сжатого распознавания характеризуются высокими требованиями к вычислительной мощности узлов РИС, а алгоритмы с доставкой по запросу, с другой стороны, эффективно работают только в случае их разработки для конкретных типов запросов, в то время как по отношению к более глобальной задаче сбора данных не могут быть рассмотрены как оптимальное решение. Соответственно, необходимость построения новой группы алгоритмов, ориентированную на работу в рамках инфраструктуры S-IoT было выделено как **нерешенная часть общей проблемы**.

Целью работы, таким образом, стала разработка метода на базе алгоритмов восстановления дискретных данных (SR, sampling-reconstruction), а также технологии пространственно-временного компрессионного зондирования (ST-CS, spatio-temporal compressive sensing) [15 - 21], адаптированных под спутниковые РИС в рамках концепции S-IoT.

1. Систематизация задач построения спутникового Интернета вещей

Главная особенность функционирования инфраструктуры IoT связана с крайне высокой себестоимостью масштабирования аппаратно-программного комплекса. На рис. 1 представлена обобщенная схема функционирования спутниковых РИС. С одной стороны мы видим рост объема собираемых данных, который связан с увеличением количества отдельных терминалов, а также узлов беспроводных сенсорных сетей и транспортных сетей, как то IoV (Internet of Vehicles), которые непрерывно расширяются. Одновременно растет количество и сложность программных приложений, в основу которых положена обработка собираемых данных. Таким образом, масштаб запроса исходящего от уровня приложений и уровня сбора данных в значительной степени превосходит пропускную способность, которой характеризуется сетевой уровень.

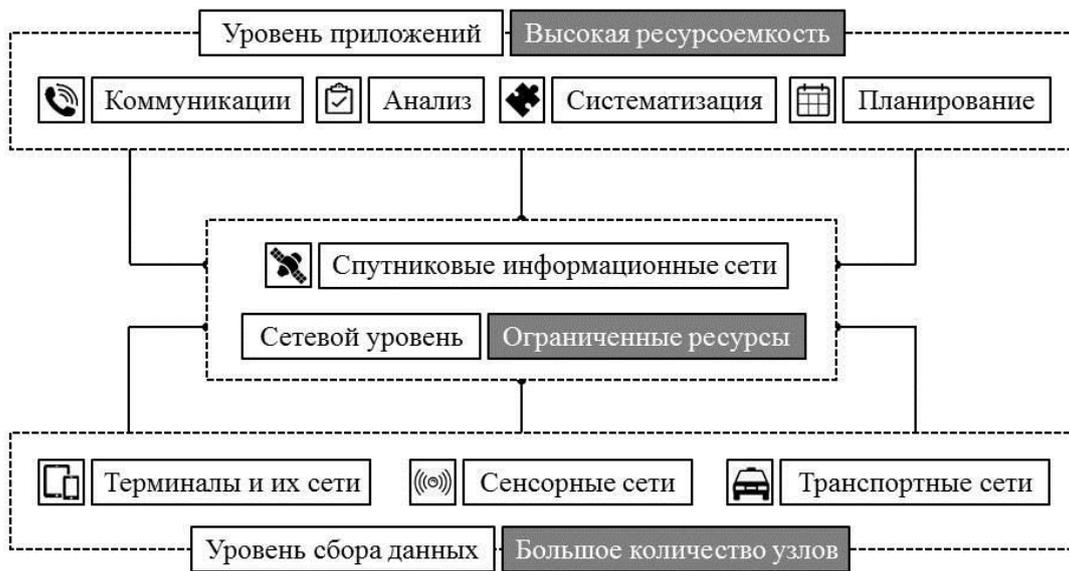


Рис. 1. Обобщенная схема функционирования спутниковых РИС

Базовым подходом для решения данной проблемы является применение наземных центров обработки данных (ЦОД) на которые через наземные станции приема и передачи данных передаются информационные блоки для дальнейшего анализа и систематизации (рис. 2).

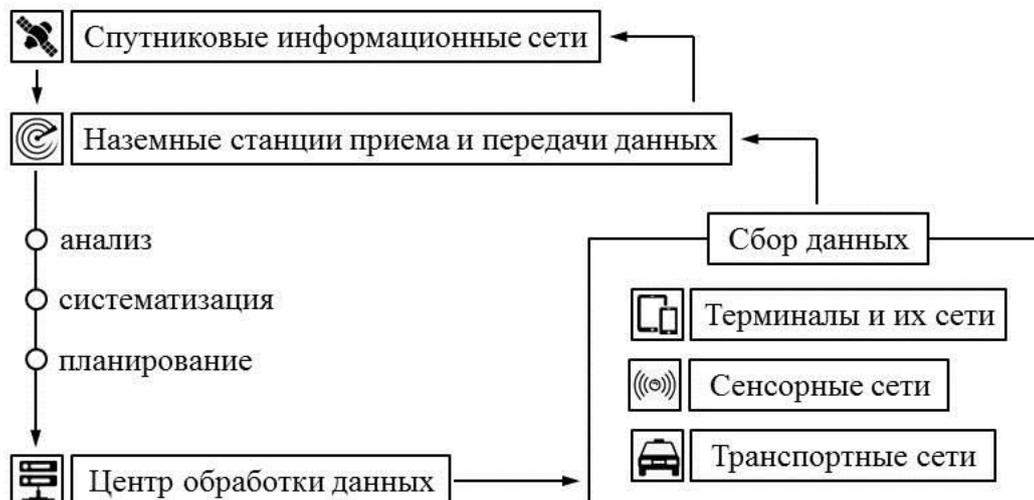


Рис. 2. Базовый алгоритм использования наземных станций в рамках S-IoT

Рассмотрим схему работы наземных станций в рамках S-IoT на уровне построения математической модели. Пусть множество N наземных узлов распределено в области $x \times y$, где $x \in [0; X]$ и $y \in [0; Y]$. Каждый узел в каждый момент времени $t_i = (t_0 + i \cdot \Delta t)$, где t_0 — начальный момент, а Δt , соответственно, минимальный период сбора данных, характеризуется парой координат $\{x_n(t_i); y_n(t_i)\}$, получаемых через систему позиционирования, причем $n \in [0; N]$, $x_n(t_i) \in [0; X]$, $y_n(t_i) \in [0; Y]$, а также массивом данных, который передается этим узлом информации $C_n(t_i)$.

SR-методология подразумевает ограничение информационного объема данных, который подлежит сбору, путем внедрения процедуры выборочного сбора и дальнейшую реконструкцию отсутствующих блоков данных. Пусть количество узлов, с которых собираются данные в отдельный момент времени t_i составляет M_i , а общий массив восстановленных данных $C_n'(t_i)$. В таком случае уровень сжатия данных подлежащих сбору DCR_i (Data Collection Ratio) и точность сбора данных DCA_i (Data Collection Accuracy) могут быть рассчитаны следующим образом:

$$DCR_i = \frac{M_i}{N}, \text{ где } M_i(0; N), \quad (1)$$

$$DCA_i = \begin{cases} \left(1 - \frac{\sum_{n=1}^N \left(\frac{|C_n(t_i) - C_n'(t_i)|}{|C_n(t_i)|} \right) \right) \cdot 100\% & \text{для } C_n(t_i) \neq 0 \\ \text{не может быть определена} & \text{для } C_n(t_i) = 0 \end{cases} \quad (2)$$

При этом точность сбора данных связана с конкретными методами получения выборки и реконструкции.

2. Принципы построения спутникового Интернета вещей на базе алгоритмов восстановления дискретных данных

Проведенный анализ показывает, что работа SR-алгоритма может быть представлена на основе трех ключевых этапов (рис. 3):

- кластеризация входных данных;
- получение выборки;
- восстановление данных, которые не были получены.

На этапе кластеризации SR-алгоритм использует базы ретроспективных данных сенсорных сетей, накопленные за большой период времени для кластеризации входных данных от наземных узлов, которая должна характеризоваться высоким уровнем пространственной корреляции. Далее этапы получения выборки и восстановления данных реконструкции выполняются отдельно для каждого кластера.

На этапе выборки SR-алгоритм определяет каждый наземный узел сенсорной сети, который должен передать данные, формирует выборку и, наконец, на ее основе агрегирует соответствующий информационный массив через систему приема-передачи спутниковой РИС.

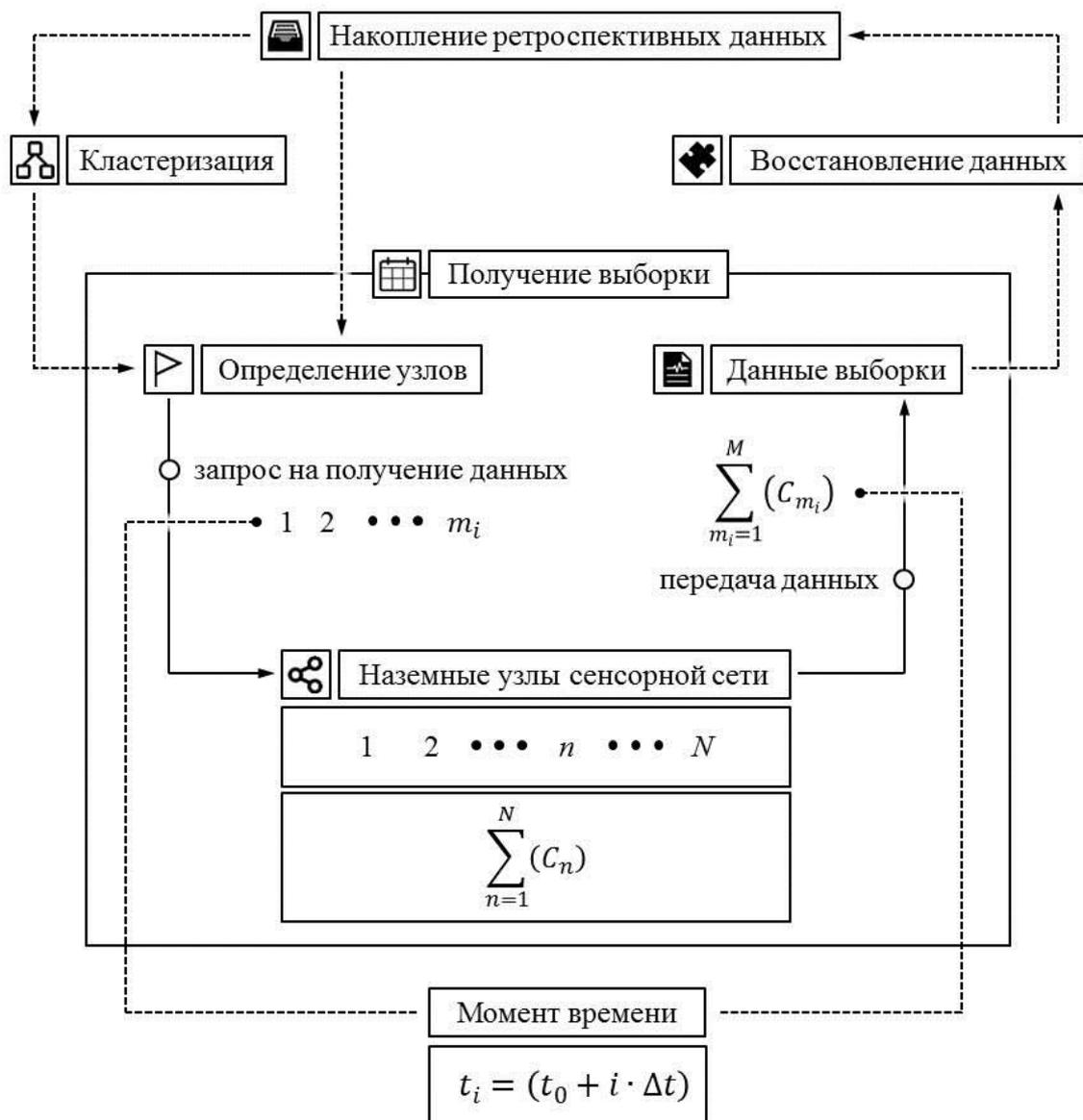


Рис. 3. Универсальная схема работы алгоритма восстановления дискретных данных для S-IoT

На последнем этапе SR-алгоритм восстанавливает недостающие блоки данных от наземных узлов сенсорной сети, используя временную и пространственную корреляцию между элементами полученного информационного массива. Восстановленные данные передаются в базу данных и используются на последующих этапах кластеризации, получения выборки и восстановления данных в рамках функционирования SR-алгоритма.

Выводы

В результате проведенного исследования были рассмотрены и систематизированы ограничения связанные с развитием концепции IoT. Проведена систематизация типовых задач характерных для построения аппаратно-программной базы S-IoT и поставлена задача построения комплексной методологии работы со спутниковыми распределенными информационными системами на базе алгоритмов восстановления дискретных данных и технологии пространственно-временного компрессионного зондирования. Предложена модель, которая базируется на целевых функциях уровня сжатия данных подлежащих сбору и точности сбора данных, и построена универсальная схема работы алгоритма восстановления дискретных данных для S-IoT, которая включает в себя такие функциональные элементы как кластеризация входных данных, получение ограниченной выборки данных и восстановление данных от полного набора наземных узлов сенсорной сети.

1. *Aiyetoro & Owolawi*, 2019. Spectrum Management Schemes for Internet of Remote Things (IoRT) Devices in 5G Networks via GEO Satellite. *Future Internet*. 11 (12). 257. doi: 10.3390/fi11120257.
2. *Wei J. & Cao S.*, 2019. Application of Edge Intelligent Computing in Satellite Internet of Things. 2019 IEEE International Conference on Smart Internet of Things (SmartIoT). doi: 10.1109/smariot.2019.00022.
3. *Kak A., Guven E., Ergin U.E., Akyildiz I.F.*, 2018. Performance evaluation of SDN-based Internet of Space Things. In: 2018 IEEE Globecom Workshops (GC Wkshps., Pp. 1–6. IEEE Press, Piscataway.
4. *Aiyetoro & Owolawi*, 2019. Spectrum Management Schemes for Internet of Remote Things (IoRT) Devices in 5G Networks via GEO Satellite. *Future Internet*. 11 (12). 257. doi: 10.3390/fi11120257.
5. *Bacco M. et al.*, 2019. IoT applications and services in space information networks. *IEEE Wirel. Commun.* 26 (2), 31–37.
6. *Soua R., Palattella M.R. & Engel T.*, 2018. IoT Application Protocols Optimisation for Future Integrated M2M-Satellite Networks. 2018 Global Information Infrastructure and Networking Symposium (GIIS). doi: 10.1109/giis.2018.8635784.
7. M2M and IoT via Satellite, 7th edn. [Электронный ресурс]. Режим доступа: <http://www.nsr.com/research-reports/satellitecommunications-1/m2m-and-iot-via-satellite-7th-edition/> (дата обращения: 28.02.2017).
8. *Cheng S., Cai Z., Li J.* Approximate sensory data collection: a survey, 2017. *Sensors* 17 (3). 564.
9. *Gedik B., Liu L., Yu P.S.* ASAP: an adaptive sampling approach to data collection in sensor networks. *IEEE Trans. Parallel Distrib. Syst.* 18 (12), 2007. 1766–1783.
10. *Kim R., Choi J. & Lim H.*, 2015. Channel-aware repetitive data collection in wireless sensor networks. *International Journal of Communication Systems*. 30 (7). doi: 10.1002/dac.2999.
11. *Nguyen M.T., Teague K.A.* Compressive sensing based random walk routing in wireless sensor networks, 2017. *Ad Hoc Netw.* 54, 99–110.
12. *Abdelaal M. & Theel O.*, 2013. An efficient and adaptive data compression technique for energy conservation in wireless sensor networks, 2013 IEEE Conference on Wireless Sensor (ICWISE). doi: 10.1109/icwise.2013.6728793.
13. *Silberstein A., Braynard R., Ellis C., Munagala K., Yang J.*, 2006. A sampling-based approach to optimizing top-k queries in sensor networks. In: 22nd International Conference on Data Engineering (ICDE 2006). P. 68. IEEE Computer Society, Washington DC.
14. *Guo L., Beyah R., Li Y.*, 2011. SMITE: a stochastic compressive data collection protocol for mobile wireless sensor networks. In: 2011 Proceedings IEEE INFOCOM. Pp. 1611–1619. IEEE Press, Piscataway.
15. Compressive Sensing Based Data Collection in Wireless Sensor Networks, 2016. Internet of Things (IoT) and Engineering Applications. doi: 10.23977/iotea.2016.11005.
16. *Meyer M., Desbrun M., Schröder P., Barr A.H.* Discrete differential-geometry operators for triangulated 2-manifolds. In: Hege, H.C., Polthier, K. (eds.) *Visualization and Mathematics III*. Pp. 35–60. Springer, Heidelberg, 2003. https://doi.org/10.1007/978-3-662-05105-4_2.
17. *Xu X., Ansari R. & Khokhar A.*, 2015. Spatio-Temporal Hierarchical Data Aggregation Using Compressive Sensing (ST-HDACS). 2015 International Conference on Distributed Computing in Sensor Systems. doi: 10.1109/dcoss.2015.15.
18. *Kong L., Xia M., Liu X.Y., Wu M.Y., Liu X.*, 2013. Data loss and reconstruction in sensor networks. In: 2013 Proceedings IEEE INFOCOM. Pp. 1654–1662. IEEE Press, Piscataway.
19. *Rallapalli S., Qiu L., Zhang Y., Chen Y.C.*, 2010. Exploiting temporal stability and low-rank structure for localization in mobile networks. In: *Proceedings of MobiCom 2010*. Pp. 161–172. ACM, New York.
20. *Zhou X. & Ling H.*, 2016. Zero-effort projection for sensory data reconstruction in wireless sensor networks. *International Journal of Distributed Sensor Networks*. 12 (8). 155014771665942. doi: 10.1177/1550147716659425.
21. *Fei C., Zhao B., Yu W. & Wu C.*, 2019. An Approximate Data Collection Algorithm in Space-Based Internet of Things. *Security, Privacy, and Anonymity in Computation, Communication, and Storage Lecture Notes in Computer Science*, 170–184. doi: 10.1007/978-3-030-24900-7_14.