

Enhanced dynamic data storage for enabling low cost space astronomy observations for capital constrained astronomy organizations

Abstract

Mechanisms that reduce the capital and operational costs are important for increased participation in astronomy. It is important that capital constrained organizations can engage in astronomy in cost effective manner. Approaches such as telescope conversion and using small satellites reduce the cost of astronomy observations. However, astronomy data observed by converted and small satellite telescopes require storage and processing by high performance computing infrastructure. High performance computing infrastructure acquisition is expensive for capital constrained astronomy organizations. The reduction in costs obtained by using converted and small satellite telescopes is not matched by a corresponding reduction in high performance computing. This paper addresses this challenge and proposes using a software defined space data storage system. The software defined space data storage system considers space telescopes as primary satellites and telecommunication and earth observation satellites as secondary satellites. The primary and secondary satellites are grouped in logical clusters. Secondary satellites are temporal data centers that store the astronomy data that cannot be held on primary satellites. The discussion in this paper presents algorithms that enable the identification of suitable secondary satellites and also influence the entry and exit of secondary satellite into dynamic clusters.

Keywords: dynamic data storage, space, astronomy, satellites, duration, capital constrained astronomy organisations

Volume I Issue 4 - 2017

Periola AA,¹ Ohize H²

¹Department of Electrical Engineering, University of Cape Town, South Africa

²Department of Electrical Engineering, Federal University of Technology Minna, Nigeria

Correspondence: Ayodele Abiola Periola, Department of Electrical Engineering, University of Cape Town, South Africa, Tel 0787369576, Email periola@crg.ee.uct.ac.za

Received: July 11, 2017 | **Published:** November 17, 2017

Introduction

Data storage and processing are important requirements in wireless systems that are used for telecommunications, earth observations and astronomy. Existing and future wireless systems require robust data processing. This has led to the emergence of new computing paradigms such as cloud computing.¹⁻³ Cloud computing makes use of both hardware and software components. The use of cloud computing has the capacity to enhance future radio astronomy by reducing computing costs. The computing cost is reduced because the astronomy organization no longer has to install expensive computing hardware i.e. the high performance computing infrastructure.⁴

The critical component in cloud computing that stores data is the data center. Currently, data centers are mostly on ground and have high operational costs due to the necessity of powering and cooling. These high costs have necessitated the design of novel mechanisms to reduce operational costs.⁵⁻⁷ The challenge of high operational cost arises due to the location of the data center. Currently, most data centers are land based. However, alternative locations such as the ocean⁸⁻⁹ have been identified.

Data centers also play an important role in astronomy where the analysis of exascale data is expected.¹⁰ It is important to design mechanisms that reduce the capital and operational costs of using data centers in astronomy. The data centers used in astronomy are terrestrially based like the conventional data centers being used in the cloud computing paradigm. Hence, data centers used in radio astronomy also have the drawbacks facing existing terrestrial data centre systems.

In addition, data center locations have also been identified to influence the latency and throughput associated with accessing data from the data center. This influences data access in astronomy because of the interaction of telescope arrays and the high performance computing. Astronomy telescopes can be either ground based or space based assets. These assets should be able to transmit data to the high performance computing infrastructure at high throughput and low latency.

Furthermore, data centers (high performance computing) used in astronomy have high capital costs. The availability of data centers (high performance computing) infrastructure is important for processing the data observed by telescopes. However, the high capital cost of owning data centers limits the ability of capital constrained organizations to conduct astronomy. Capital constrained organizations can make use of space telescopes designed using low cost small satellites.¹⁰ The use of converted telescopes has also been recognized to reduce the cost of telescope acquisition for capital constrained astronomy organizations.¹¹⁻¹²

However, the use of small satellites and converted telescopes by capital constrained astronomy organizations does not influence high performance computing costs. This affects the cost of computing for space astronomy observations but not terrestrial astronomy observations. Low cost high performance computing infrastructure can be used alongside telescopes realised from converted earth stations. However, this is not suitable for low cost small satellites used in space astronomy.

The use of low cost small satellites for space astronomy by capital

constrained astronomy organizations should be matched with low cost computing. It is also important that such low cost computing enhance the throughput and reduce the latency associated with processing and accessing processed astronomy data. In the case of space astronomy, a high performance computing infrastructure is most appropriate for this purpose. Hence, the use of space based high performance computing infrastructure is proposed in this paper. It is also important that the use of a space based high performance computing infrastructure does not increase the costs of conducting astronomy observations. A network architecture that meets these objectives for future space based astronomy applications is proposed in this paper.

This paper proposes novel data processing architecture for future astronomy observations. The proposed architectures uses data centers i.e. high performance computing infrastructures sited in space. It incorporates intelligence and enhances the data transfer quality of service in astronomy. This paper makes the following contributions:

- First, it proposes a novel fractionated data storage system for astronomy observations. The motivation for the design of a fractionated storage system is that the universe presents information in fractions contained in samples. The fractionated storage system is suitable for integration into existing and future space data center systems.
- Second, it proposes a space based dynamic storage network that is controlled from a ground control center. And enables the re-use of unutilized storage on-board in-orbiting satellites. This improves the ability of astronomy to benefit from serendipitous discoveries.

The proposed mechanisms aims to enable the design of a dynamic space based storage system with high data transfer rates. The rest of this paper is organized as follows. Section II defines the problem. Section III discusses the novel fractionate data storage and processing system. Section IV describes the dynamic space based storage system. Section V concludes the paper.

Problem definition

The scenario being considered comprises in-orbit satellites with on-board storage capacity. These satellites are inter-connected via high throughput inter-satellite links and can communicate with ground stations at different locations when necessary. In addition, the satellites can be controlled from a ground facility. Let S be the set of satellites.

$$S = \{s_1, s_2, \dots, s_k\} \quad (1)$$

The satellite $s_x; s_x \in S$ has instantaneous on-board power, utilised memory, non-utilized memory and storage required for holding observed data. Let $P(s_x, t_y); t = \{t_1, t_2, \dots, t_n\}; t_y \in t, \alpha(s_x, t_y); \beta(s_x, t_y), \gamma(s_x, t_y)$ and $I(s_x) \in \{0,1\}$ be the (1) instantaneous on-board power, (2) utilized memory, (3) non-utilized memory, (4) required storage and (5) status indicator of s_x at epoch t_y respectively. $I(s_x) = 0$ and $I(s_x) = 1$ indicates that s_x is not and is a space telescope respectively.

In a case given as $I(s_x) = 1, P(s_x, t_y) > 0, \gamma(s_x, t_y) > \beta(s_x, t_y)$, the space telescope s_x is not able to provide the required storage for the data arising from the observation at epoch t_y even though there is sufficient power. The space telescope s_x does not continue with observation when data cannot transfer to an out-of-view ground station. The discontinuing of observation at epoch t_y in this case

reduces the probability of serendipitous discovery. The probability of serendipitous discovery can be enhanced by using a satellite $s_q; s_q \in S$ for which $P(s_q, t_y) > 0, \gamma(s_q, t_y) < \beta(s_q, t_y), I(s_q) = 0$. However, s_x needs a mechanism to identify and use s_q at epoch t_y .

Another plausible case is one given as

$$I(s_x) = 1, P(s_x, t_y) > 0, \gamma(s_x, t_y) > \beta(s_x, t_y); \gamma(s_x, t_{y+1}) < \beta(s_x, t_{y+1}); t_{y+1} \in t.$$

In this case, the observation of s_x at epoch t_y can be stored but not at epoch t_{y+1} . In this case, a mechanism that enables s_x to use s_q for which $P(s_q, t_y) > 0, \gamma(s_q, t_{y+1}) < \beta(s_q, t_{y+1})$ is required.

This paper proposes the mechanisms to enable the usage of s_q to meet the storage demands of s_x when $P(s_x, t_y) > 0, \gamma(s_x, t_y) > \beta(s_x, t_y)$ or when

$$I(s_x) = 1, P(s_x, t_y) > 0, \gamma(s_x, t_y) > \beta(s_x, t_y); \gamma(s_x, t_{y+1}) < \beta(s_x, t_{y+1}); t_{y+1} \in t.$$

Novel fractionated data storage and processing mechanism

This section discusses the novel fractionated data storage and processing mechanism. The proposed storage and processing mechanism becomes necessary for s_x when:

$$(-\gamma(s_x, t_y) + \beta(s_x, t_y)) \leq \left(\sum_{m=1}^f U_m(t_y) \right), I(s_x) = 1 \quad (2)$$

$$(\gamma(s_x, t_y) - \beta(s_x, t_y)) \geq \left(\sum_{m=f+1}^h U_m(t_y) \right), I(s_x) = 1 \quad (3)$$

The relation in (3) show that s_x is a space telescope that can provide storage for f samples of observed data at epoch t_y .

However, s_x is not able to provide storage for the remaining $(h-f)$ samples of observed data at epoch t_y . The satellite s_q can be used to host $(h-f)$ samples of observed data at epoch t_y when:

$$(-\gamma(s_q, t_y) + \beta(s_q, t_y)) \leq \sum_{m=f+1}^h U_m(t_y), I(s_q) = 0; I(s_q) = 0 \quad (4)$$

The satellite s_q can be used to host data for up to y seconds when:

$$\left(\frac{1}{y+1} \sum_{\alpha=1}^{y+1} \frac{\alpha(s_q, t_\alpha)}{\alpha(s_q, t_\alpha) + \beta(s_q, t_\alpha)} \right) < \Phi(s_q) \quad (5)$$

Where $\Phi(s_q)$ is the storage utilisation threshold of s_q . The use of the proposed data fractionation paradigm enables satellites to acquire multi-modal functionality. A satellite incorporating the proposed software is used for astronomy observations and is considered a primary satellite. In the primary role, the satellite is used for astronomy observations. In the secondary role, the satellite is used as space based data center. The relations in (2) and (3) show that $I(s_x) = 1$ is a space telescope while the relations in (4) and (5) show that $I(s_q) = 0$ and is not a space telescope. Satellites such as s_x can be referred to as primary satellites and are used for space astronomy observations only. Satellites such as s_q are considered secondary satellites and are used for non-astronomy observations applications. Examples of such applications are telecommunications and earth observations.

Telecommunications and earth observation satellites can be used as secondary satellites for temporarily holding astronomy data. This occurs when satellite subscriber network activity is below an expected

threshold. Data fractionation arises when part of the samples observed by a space telescope i.e. primary satellite is stored on a suitable secondary satellite. This is demonstrated in relations (3) and (4). The relations in (2)-(5) concern network state at epoch t_y of satellites s_x and s_q . The proposed data storage and processing mechanism is implemented in software and uploaded to primary satellites i.e. s_x from a ground control facility during a communication window. The block diagram of the proposed mechanism incorporated on the primary satellite is shown in Figure 1. As seen in Figure 1, the primary satellite located in the low earth orbit comprises the astronomy payload, memory utilization module, data storage and communications system. The payload executes the observation requirements. The memory utilization module executes computations related to the expressions in (2), (3), (4) and (5). It also coordinates relations with the communication system when necessary.

In Figure 1, there is bi-directional communication between the memory utilization module and the data storage unit. The memory utilization module sends data to the storage unit in the forward direction and receives information on the utilized memory, un-utilized memory from the data storage unit. The information on the un-utilized and utilized memory is used to execute the relations in (2), (3), (4) and (5) alongside the required storage.

The functional flowchart of the memory utilization module in the proposed mechanism is shown in Figure 2. In Figure 2, it is assumed that the primary satellite is able to acquire information on secondary satellites that meet the requirement in (5). The relation in (5) is verified by the primary satellite using the information received from secondary satellites. The required information is received via the inter-satellite communications.

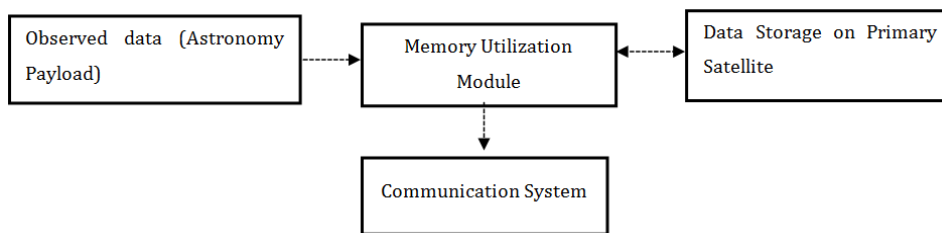


Figure 1 Block Diagram of primary satellite incorporating the proposed mechanism.

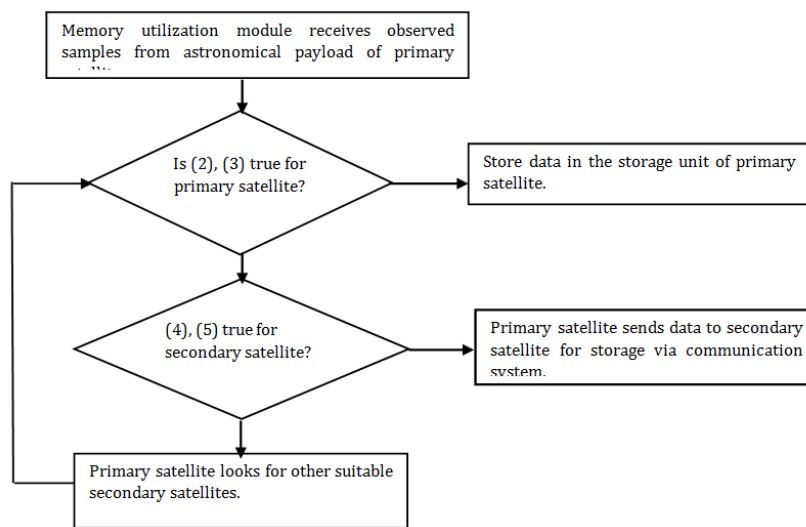


Figure 2 Functional flowchart of the memory utilization module.

Dynamic space based data storage system

The proposed fractionated data storage mechanism is incorporated in in-orbit satellites. In this paper, the mechanism that enables fractionated data storage is implemented in software and uploaded from a ground based station. In addition, the fractionated data storage mechanism is operational aboard in-orbit satellites. The primary and secondary satellites are located in the low earth orbit and can communicate with each other. The primary and secondary satellites can be described using different logical groups.

A logical group can comprise different numbers of primary and secondary satellites that are mutually visible to each other. Let $s_{x+1}; s_{x+1} \circ S$ and $s_{q+1}; s_{q+1} \circ S$ be a primary satellite and secondary satellite cluster respectively such that:

$$s_{x+1} = \{s_{x+1}^1, s_{x+1}^2, \dots, s_{x+1}^m\} \tag{6}$$

$$s_{q+1} = \{s_{q+1}^1, s_{q+1}^2, \dots, s_{q+1}^n\}, I(s_{q+1}) = 0 \tag{7}$$

The satellite $s_{q+1}^j, s_{q+1}^j \circ S$ becomes unsuitable as a secondary satellite and uses $s_q^w, s_q^w \circ S$ as the new secondary satellite such that:

$$s_{q+1} = \{s_{q+1}^1, s_{q+1}^2, \dots, s_{q+1}^{n-1}, s_q^w\} \tag{8}$$

The satellite $s_q^w, I(s_q^w) = 0$ is in the cluster w and can be reached by existing secondary satellites specified in (7). The secondary

satellite s_q^y is suitable for use as a secondary satellite to replace s_{q+1}^j for up to y seconds for data storage when:

$$\beta(s_{q+1}^j, t_\alpha) < \gamma(s_{q+1}^j, t_\alpha), I(s_{q+1}^j) = 0; t_\alpha \in t \quad (9)$$

$$\beta(s_q^y, t_\alpha) > \gamma(s_q^y, t_\alpha) \quad (10)$$

$$\gamma(s_{q+1}^j, t_\alpha) < \beta(s_q^y, t_\alpha) \quad (11)$$

The relation in (9) describes a scenario where secondary satellite has insufficient storage space to meet the storage requirements. This can arise when the secondary satellite is required to store and process increasing amounts of either communication packets or earth observation data.

The relation in (10) shows that s_q^y has sufficient memory stage since the un-utilized space exceeds the expected storage required to be provided by s_q^y at epoch t_α . The relation in (11) shows that the storage requirements expected of s_{q+1}^j at epoch t_α i.e. $\gamma(s_{q+1}^j, t_\alpha)$ is less than the non-utilized storage of s_q^y at epoch t_α .

The secondary satellite s_q^y can be used as a temporal data center for up to y seconds when:

$$\left(\frac{1}{|t_\alpha|}\right) \left(\sum_{a=1}^y \frac{\alpha(s_q^y, t_\alpha)}{\alpha(s_q^y, t_\alpha) + \beta(s_q^y, t_\alpha)}\right) < \Phi(s_q^y); t_\alpha = \{t_1, \dots, t_y\} \quad (12)$$

Where $\Phi(s_q^y)$ is the storage utilisation threshold of s_q^y . The flowchart showing the execution of functionalities aboard primary and secondary satellites in the proposed dynamic space based data

storage system is presented in Figure 3. The flowchart shows the execution of the following tasks:

Astronomy observation and storage analysis in primary satellites:

In this stage, the astronomy payload aboard small satellite space telescopes is engaged with astronomy observations. The astronomy payload can be engaged in observing different types of signals such as X-rays, Gamma rays or radio signals. The storage analysis in this stage examines the un-utilized memory and the amount of memory required to store the observed data on each satellite. This analysis is done for individual satellites that form the primary satellite cluster. Individual primary satellites are organised in logical clusters and identify suitable secondary satellites.

Secondary satellite search and secondary satellite cluster update procedure:

The computation at this stage enables the primary satellite to search for a suitable secondary satellite. The identified secondary satellite is added to the logical cluster of the secondary satellite. This stage also involves determining the expiration epoch of secondary satellite in the secondary satellite cluster. In a case where the expiration is not imminent, secondary satellites are retained in the existing secondary satellite cluster.

Data Retrieval and transmission to ground station: This stage occurs during the communication window of the space telescope. It enables the transmission of observed data to the ground station during the downlink of data from the space telescope.

Intersatellite link enabled data transmission: The intersatellite link enables the transfer of data stored in the secondary satellite to the ground station via the primary satellite downlink. This is executed by the primary satellite communication system.

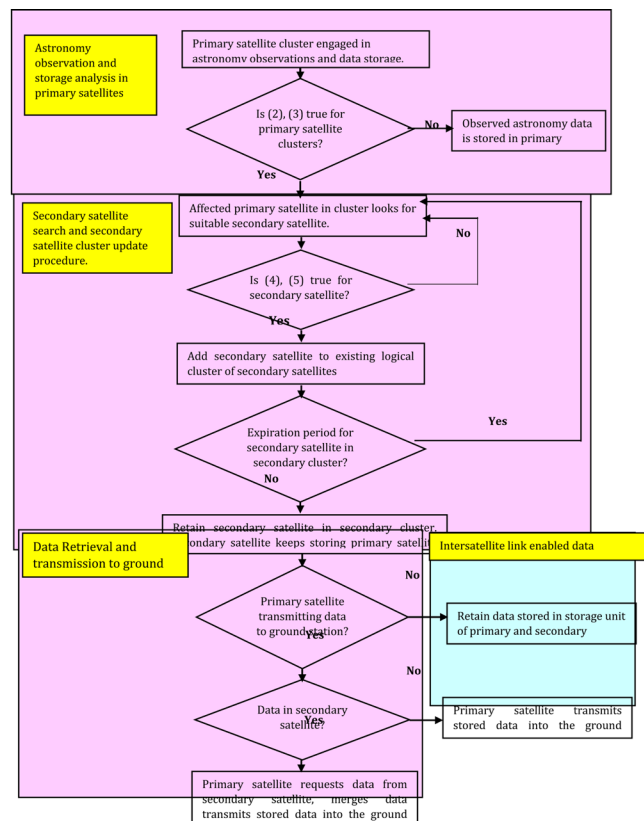


Figure 3 Flowchart showing relations between primary satellites, secondary satellites and related clusters in the space segment.

Conclusion

The discussion in this paper presents a novel architecture that enables capital constrained organizations to conduct astronomy observations with less concern for data storage costs. The presented architecture re-uses satellites with underutilized memory resources. The underutilized memory resources are used to provide storage for data observed by astronomy telescopes deployed in the low earth orbit. This is realized by dynamically using unused storage aboard satellites to store and process data. The proposed novel architecture demonstrates the potential of leveraging space assets for improved astronomy observations. In addition, the architecture separates storage requirements from space telescope observation capability. This separation enables a greater control over how data storage can be realized in space. In addition, the proposed architecture has the benefit of enabling the storage of increased amount of storage data than otherwise possible if storage was limited to the observing telescope. In addition, the paper presents a cross-over from cognitive radio concepts into astronomy systems. The motivation for primary satellites and secondary satellites has been derived from primary users and secondary users in cognitive radio networks.

Acknowledgements

The author acknowledges the efforts of reviewers and colleagues from the Department of Electrical Engineering at the University of Cape Town for the fruitful discussions held on the idea and concepts presented in the paper.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

References

1. Marotta MA, Faganello LR, Schimunek MAK, et al. Managing Mobile Cloud Computing Objective and Subjective Perspectives. *Computer Networks*. 2015;93(3):531–542.
2. YuW, Xie J, Li G. *Cloud Computing for Environmental Monitoring using Multi-Source Earth Observation Data*. International Conference on Agro-Geoinformatics, Tianjin, China, 2014. p. 1–4.
3. Cui C, He B, Yu C, et al. *AstroCloud: A Distributed cloud computing and application platforms for Astronomy' Astrophysics, Instrumentation and Methods for Astrophysics*. Cornell University Library, USA. 2017.
4. Berriman GB, Juve G, Deelman E, et al. *The Application of Cloud Computing in Astronomy: A Study of Cost and Performance' Instrumentation and Methods in Astrophysics*. Cornell University Library, USA. 2010.
5. Iazeolla G, Pieroni A. Energy Saving in Data Processing and Communication Systems. *The Scientific world Journal*. 2014;1–12.
6. Gao J. *Machine Learning Application for Data Center Optimization*. Google, 2014. p. 1–13.
7. Zhang S, Yang J, Shi Y, et al. *Dynamic Energy Storage Control for Reducing Electricity Cost in Data Centers*. Mathematical Problems in Engineering. 2015. p. 1–14.
8. Cutler BF, Whitaker NA, Fowers SG. *Artificial Reef Datacenter: United States Patent Application Publication*. UK, 2016. p. 1–15.
9. <http://natick.research.microsoft.com/>
10. Bentum MJ. *The search for exoplanets using ultra long wavelength radio astronomy*. IEEE Aerospace Conference, Big Sky, MT USA, USA, 2017. p. 1–17.
11. Mao Y, Zho Y, Huang T, et al. *DAG Constrained Scheduling Prototype for an Astronomy Exa-Sclae HPC Application*. IEEE International Conference on Smart City, Sydney, Australia, 2016. p. 631–638.
12. Periola AA, Falowo OE. *Intelligent Cognitive Radio Models for Enhancing Future Radio Astronomy Observations*. Advances in Astronomy, 2016. p. 1–16.

