Application Services in Space Information Networks

Anders Fongen
Norwegian Defence University College, Cyber Defence Academy (FHS/CIS)
Lillehammer, Norway
Email: anders@fongen.no

***Abstract*—As Low Earth Orbit satellites (LEO) are evolving from "radio mirrors" to "network entities" it is reasonable to foresee a development of satellite networks which not only forwards network traffic, but engage in middleware operations, discovery services and even collaborate applications. However, efforts already investigating "data centers in space" do so without taking into regard the special properties of orbiting spacecrafts and their relation to the demographic characteristics on the ground below them. In this position paper, an outline of a distributed system hosted in a Space Information Network (SIN) will be presented. Of special interest is the cyclic properties of the link topology and the predictability of the workload offered by its client on the surface below. Beside the general operating principle, a selection of discovery and application services will be used to demonstrate the feasibility of the principles, as well as to identify remaining problems and research topics.**

***Keywords*—LEO satellites; space information networks; AaaS in space**

I. Introduction

A number of Geosynchronous Equatorial Orbit (GEO), High Eliptical Orbit (HEO) and Low Earth Orbit (LEO) satellites together with high altitude aircrafts (balloons, drones) may form a distributed system sometimes called a Space Information Network (SIN). Although most often proposed for the provision of communication services only, some efforts also investigate the SIN as "cloud computing in space".

The advantages offered by a SIN should not only be global coverage, but also faster Round Trip Time (RTT). In theory, a satellite at 300 km altitude can offer an RTT of only 2ms, far better than any surface based path to a data center. Shorter RTT enables new applications for synchronous collaboration and remote control.

The main challenges with the AaaS (Application as a Service) principle in a SIN is the transfer of state during handover operations. For low-orbit spacecrafts, handover takes place at short intervals and must be subject to scalability analysis since a potentially high number of client sessions are kept in the spacecrafts.

The rest of the paper is organized as follows: In Section II, the distinct properties of a SIN, compared to other distributed systems are presented. In Section III, a small modeling effort is shown together with a plot considering population density during orbital period. Related research is briefly discussed in Section IV. A more technological perspective on a number of basic elements are discussed in Section V followed by a presentation of three example application services in Section VI. Finally, a summary and suggested future research problems are given in Section VII.

II. Properties of a SIN

A SIN has properties distinct from an ordinary distributed mobile system. Unlike a mobile system, where the users move around a more or less stationary infrastructure, a SIN system will consist of a mobile infrastructure and stationary (as seen from space) users. Besides, the mobility patterns of the SIN infrastructure are highly predictable, and cyclic. At any time in the future, the position of the spacecraft, the topology of the inter-satellite links and the relative position of the ground stations are known, as well as the demographic characteristics of the population at the surface below.

Demographic characteristics refer to population density, age distribution, language, etc, as well as other preferences learned through earlier pass or from other satellites. The characteristics can be predicted to estimate type and volume of service requests for which the satellites can prepare in advance.

The workload estimation may also be used to schedule delay tolerant communication and computational activities to periods of low activity. The population pattern of the earth leaves short, but frequent periods of no human activity below (except for expeditioners and unmanned sensors) where background activities like synchronization and replication can take place. Ground stations can be used for intermittent high-speed communication with the terrestrial network and should be located in locations with low population density. Also, inter-satellite communication for background activities can take place in the polar regions, where the longitude distance between the satellites is short.

III. Population Density Estimations

For the purpose of studying the varying population density on the ground below a satellite during its orbital period, a Python program was made which models the orbit and correlates the satellite position with demographic data available for download [1].

The downloaded data is easily accessed and incorporated into the program. Each grid element limited by longitude great circles and latitude small circles is represented with a value indicating the number of people per square km.

The modeled orbits are all assumed to be circular with 90 minutes orbit cycle o, and with inclination angles ϕ of 60, 70 and 80 degrees, respectively. From the length of the arc a, which is expressed as a function of time t:

$$a = (t/o) \bmod 1 * 2\pi \qquad (1)$$

 ISBN: 978-1-61208-893-8

The corresponding longitude lo and latitude la is calculated as a function of a and ϕ:

$$la = \arcsin(\sin(\phi) * \sin(a)) \quad (2)$$

$$lo = \arccos(\cos(a)/\cos(la)) \quad (3)$$

A graphical representation of the result is shown in Figure 1, with the horizontal axis showing time in minutes and the vertical axis the corresponding population density. Different fill patterns represent the different orbit inclination angles as indicated in the legends. As expected, the plot shows that there are short bursts of high population density below with idle periods in between. Please observe that there seems to be one high peak for every 90 minutes cycle, with only modest traffic in between.

The plot is far from accurate in its details, but indicates the opportunity for the satellite to spend its communication resources on maintenance and updating tasks between busy periods.

Since higher inclination angles means relatively more time over polar regions the results were expected to show bigger variations than what is indicated in the figure. But the most important question remains as: How can these idle periods be used for communication and computation tasks not related to client requests, and to prepare the spacecraft for the next populated area on its itinerary?

IV. Related Research

Most efforts on SIN have envisioned the structure as a provider of communication services only, where the communication endpoints are located on the earth's surface. Existing infrastructure like the Iridium system has been in operation for three decades and proves the feasibility of a LEO system for global telephone service. More recent initiatives like the Starlink infrastructure are currently being deployed on a very large scale for the provision of global Internet access [2].

Other projects have proposed "cloud computing" in space by deploying data centers in larger satellites with volume and energy resources sufficient for their operation. Other smaller satellites may carry units for "fog computing" in a distributed fashion in order to provide replicated services through communication paths with fewer hops [3] [4].

The resource management in satellites with great variations in its workload can be predicted, but machine learning approaches may also contribute to a management scheme which better adapts to unpredictable variations. Zhou et al. offer their results on a study based on neural networks in [5].

In order to improve the communication capacity of SIN units, lots of research has gone into the development of antennas for spatial multiplexing (space-division multiple access, SDMA), beamforming, non-orthogonal multiple access, optical communication links etc. [6] [7].

The proposals made in this position paper will not deal with technical details in the communication technology, but rather view the SIN as a distributed system which borrows its methods and solutions from the field of distributed computing. The author is not aware of other efforts to investigate this perspective to the extent necessary for a SIN offering application services.

V. Technological aspects of SIN services

Services offered by a SIN will have properties different from ordinary distributed systems. A technological discussion on how to support these properties follows in subsequent paragraphs.

A. State Transfer Between SIN units

For any location on earth, a LEO satellite will pass from one horizon to the other in a matter of minutes. Frequent handover to a trailing satellite every few minutes is necessary in order to provide a ground terminal with continuous service. For the provision of communication services only, this is a solved problem. For application services, however, a handover requires a state transfer of possibly complex and large data structures, e.g.:

- Web session objects, with open connections to files and databases
- Cached contents in Domain Name Service (DNS) proxies
- Shared document collections expected to be available and up to date
- Ongoing chat and media-rich conferences

Some of these data structures will be bound to a client session and their content subject to client actions. Others, like cached content from discovery protocols, are dependent on the collective actions of the client community. This kind of information can also be speculatively pre-fetched, based on expected client requests derived from the correlation of orbital parameters and demographic data.

State bound to client sessions cannot be pre-fetched, but must be transferred from leading satellites, i.e. the previous satellites (front and front-west, since the earth is rotating counter clockwise), and passed on to trailing satellites (back and back-east). Communication between satellites in the same orbital plane may effectively use optical links for high speed communication since their relative positions are fixed, while links to satellites in the neighbouring orbital plane would need adjustable beams as their relative position changes with the latitude position. The optical links will not compete with ground links for transmission capacity.

One can visualize the collection of client states being "fixed" in the sky above the client's position while the satellites move and continuously relay the state objects in the opposite direction. Some session objects will be passed to trailing satellites in the same orbital plane, others to the orbital plane to the east, depending on the hand-over operation (of which the sending satellite is assumed to be informed of).

The handling of shared storage (document collections and databases) which should be available for ground terminals takes the form of a replication problem where availability, consistency and ordering semantics must be taken into account. The timeliness of updates is also of interest, since the predictable patterns of user request allows some updates to

 ISBN: 978-1-61208-893-8

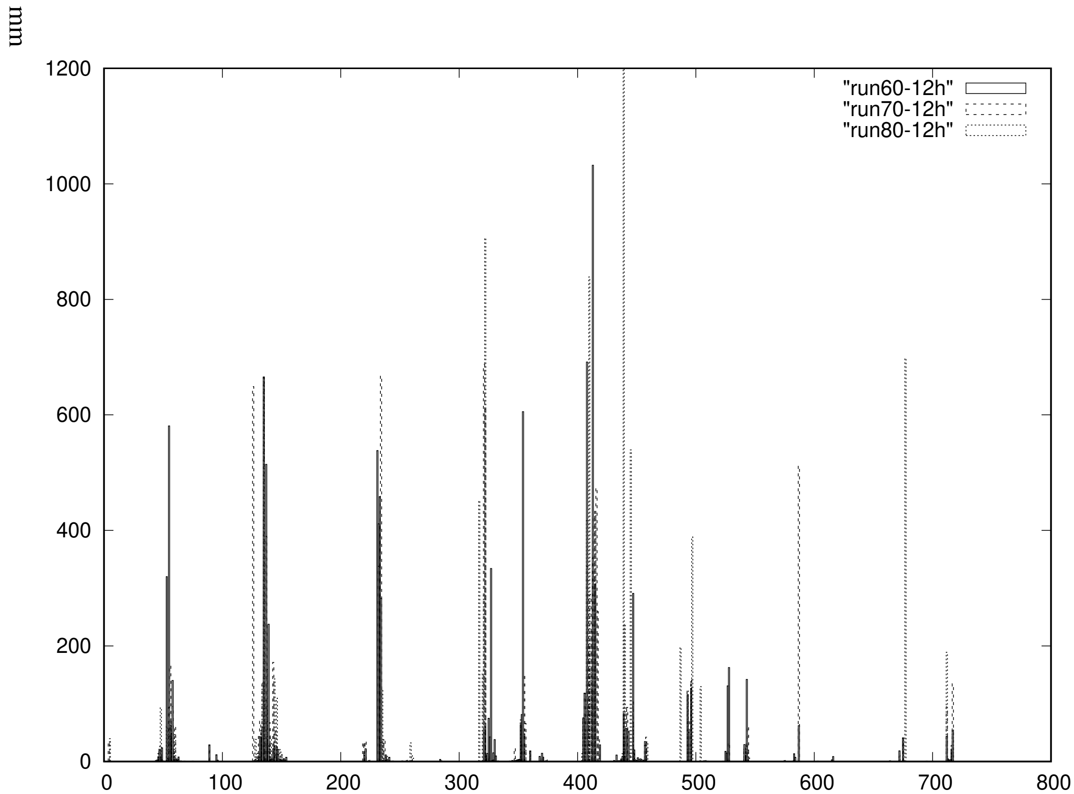

Figure 1. Population density below a satellite during 12 hour orbit

be postponed until more communication and computational resources are in supply.

Only objects with *serializable* properties (in the sense used in the Java programming language) may be passed on between spacecrafts. Objects representing local OS resources like sockets, files and mutexes will not. This limitation will impose the same system design restrictions as what are found in traditional mobile and distributed systems.

B. Delay Tolerant Services

A significant fraction of client requests is expected to be forwarded to other resources, possibly terrestrial. Due to response time requirements, the requests and the resulting reply need to be given high priority. Other requests can be regarded as delay tolerant, and the forwarding operation may be postponed to a later instant where resources have better availability, e.g. when the satellite is within range of a ground station. Requests falling into this category can be:

- E-mail operations, both inbox update and sending (subject to message priority)
- Certain type of sensor readings, e.g., environmental observations
- Software updates
- Transfer of large files
- BitTorrent applications
- Cache refreshes
- Speculative (proactive) replication

C. IP address structure

The satellites will expose their network endpoint for client connections, but it is not advisable to give each satellite a unique IP address. With separate Internet Protocol (IP) addresses, handover operations would require complicated arrangements for maintaining existing flows and connections. A more appealing approach is to give all satellites the same IP address for their service endpoint, since clients on the surface will never address more than one of them. An IP address that is preserved across handover will allow a transparent transfer of User Datagram Protocol (UDP) flows from one satellite to the next. Transmission Control Protocol (TCP) connections may also in theory be transferred, but would require a modified TCP protocol stack.

Satellites will need to invoke services in other satellites through their inter-satellite links. The interfaces used for this purpose must have unique IP addresses within a private/non-connected subnet, since these interfaces should not be reachable from endpoints elsewhere, e.g., in the Internet.

For the forwarding of IP datagrams between satellites, the inter-satellite link interfaces could be used for IP-based routing, but a more flexible solution would be to leave forwarding actions to the link layer, e.g., by employing Multiprotocol Label Switching (MPLS). The connection between a terrestrial client and a spacecraft should be handled likewise by a link-layer tunnel which also take care of keeping the handover operation transparent at the IP layer.

D. Speculative caching

As a satellite enters a populated region, the expected workload offered by terrestrial clients is estimated from a range of data sources: The demographic data of the region, the set of requests from earlier passes, the requests received from immediate leading spacecrafts etc. These data sources can be subject to statistical or machine-learned algorithms in order to determine the content of on-board data storage that will provide the best and fastest response to the requests from that region.

This content may be loaded from leading satellites or from ground stations. For this reason, the location of ground stations become a significant factor for the caching scheme. Even existing data objects may be loaded if the master copy is updated.

The cached data objects may be updated through client requests, and the replication of updates to other storage units must observe the decided consistency and ordering semantics.

VI. Outline of Example Applications

The following paragraphs will discuss the SIN's ability to support a small selection of simple applications, chosen in order also to identify architectural limitations. The selection includes both discovery service, streaming service and storage service.

A. Discovery Service - DNS

Domain Name Services (DNS) is of utmost importance for the daily transactions taking place on the Internet. The performance of the DNS service may well form a bottleneck in these transactions since most of them require the DNS service invocation to complete prior to the actual application service. The SIN-based DNS service will act as a DNS proxy and pass on requests through the communication service if they cannot be resolved by the local cache.

The chosen strategies should aim to maximize the cache-hit rate, which is an easily measured number. Due to the diversity of requests during the complete orbital period, a single body of cached DNS entries of reasonable size is not expected to produce an acceptable hit rate. Due to the predictability of requests during the orbital period, ground stations may receive request statistics from recently passed areas and upload a selection of cache entries based on previously received statistics. Ground stations need to be interconnected in order to obtain this, and cache entry selection may well be subject to machine-learning algorithms.

This approach exploits the great variations in request traffic during the orbital period, and the ground stations should be located in less populated areas in order not to compete with traffic from ground terminals.

B. Conversational Service - SIP

As was mentioned in the introduction, a LEO spacecraft can offer a shorter RTT than most terrestrial paths, but requires that clients are served by the same spacecraft or by a short path of spacecrafts. This poses an interesting research problem, since client handover should happen in a synchronous manner to keep the path short.

A Voice-over-IP (VoIP) service depends on the Service Description Protocol/Session Initiation Protocol (SDP/SIP) protocols for signalling, but not for media traffic, which is typically handled by the Realtime Transport Protocol (RTP) protocol. The latency requirements will not be applied to the signalling phase, which can progress and finish in a relaxed manner, but the media transfer phase should benefit from the low latency offered by the SIN to enable applications like music rehearsals, remote control with video feedback etc.

A SIP server can be implemented in each satellite and will locate and connect VoIP users even on other parts for the SIN, as well as those connected to terrestrial SIP servers. SIP servers do not normally facilitate media traffic, but leave that to realtime-centric protocols like RTP. The perceived quality of the VoIP service relies more on the sound quality than the connection latency, and the properties of the IP communication services of the SIN becomes the most important factor.

Multi-part VoIP sessions identify the need for multicast services in the SIN. IP multicast is not a candidate technology, since the SIN is not an IP network. The link layer tunnelling technology used for the SIN infrastructure must be able to establish and maintain a forest of multicast trees in the presence of frequent handovers and relaying to endpoints in the IP-based terrestrial network. Multi-part VoIP has been researched in RFC4353, RFC4579 and, most detailed, in RFC5850 [8]. The latter identifies mechanisms for multicast endpoints, based on IP multicast protocols. Arrangements for multicast through L2 tunnels must be developed and implemented in the VoIP User Agent for the end systems directly connected to the SIP.

C. Storage Service - Content Delivery Network

Since the vast majority of web transactions are protected by the `https` protocol, traditional web caches do not offer any reduction in latency or traffic volume. Content Delivery Networks (CDN), however, works fine over https and offers improved web performance. Can a CDN be deployed in a SIN?

A CDN combines DNS and Hypertext Transfer Protocol (HTTP) services to provide content from the replica located closest to the client. One replica may be contained in the satellite above, and the DNS provided by the SIN will refer the client to its IP address if the content is to be found there.

Contrary to terrestrial CDN services, the CDN provided by the SIN will not benefit from the storage of localized information, since the satellite traverses the entire earth's surface. Ground stations may assist the satellites in loading information for the surface area ahead in its path, information which may be discarded at a later instant. The content of any CDN instance is therefore subject to speculative/pro-active replication from ground stations based on demographic properties ahead, and possibly from leading satellites in the same orbital plane.

As earlier stated, the SIN satellites may share the same IP address. It simplifies the client configuration, and ongoing sessions (TCP connections etc.) with SIN services are not interrupted during a handover to another satellite. The CDN design is also simplified since the service will always be found on the same IP address. On the other hand, it is infeasible to give guarantees that the actual CDN instance under invocation contains the actual information object. This problem cannot be solved by the DNS service, and a form of request forwarding must be implemented in order to forward requests to supplementary CDN servers.

VII. Conclusion

The architecture principles of a service-oriented satellite network is the focus of this position paper. The contribution in the proposal is to take the cyclic properties of the orbit and a map of the population density into regard for the scheduling of traffic with ground stations and other satellites. A number of example services are discussed in more detail.

The conclusion from the initial analysis is that several established protocols will need modification in the User Agent software in the terrestrial clients, and that certain service qualities need to be chosen by the client user, e.g., the priority of mail messages. This is somewhat similar to the required modifications found in Delay Tolerant Systems.

Trust management has not been addressed in this article, which is an obvious requirements for the protection of data and separation of activities. Trust Management is being investigated and the results are subject for publishing in a future paper.

The satellite platform will need a Service Provider Interface in order for independent service providers to implement services in the satellites.

Other future research activities in this field of interest include machine-learning for workload prediction, optimization of replication activities, resource management and scalability analysis.

References

[1] "Gridded population of the world v.4.11," https://sedac.ciesin.columbia.edu/data/collection/gpw-v4/sets/browse, [Online; accessed 30-Aug-2021].

[2] "Starlink web site," https://www.starlink.com/, [Online; accessed 30-Aug-2021].

[3] S. Briatore, N. Garzaniti, and A. Golkar, "Towards the internet for space: Bringing cloud computing to space systems," in *36th International Communications Satellite Systems Conference (ICSSC 2018)*, 2018, pp. 1–5.

[4] S. Cao *et al.*, "Space-based cloud-fog computing architecture and its applications," in *2019 IEEE World Congress on Services (SERVICES)*, vol. 2642-939X, 2019, pp. 166–171.

[5] H. He, D. Zhou, M. Sheng, and J. Li, "Mission structure learning-based resource allocation in space information networks," in *ICC 2021 - IEEE International Conference on Communications*, 2021, pp. 1–6.

[6] X. Zhang, L. Zhu, T. Li, Y. Xia, and W. Zhuang, "Multiple-user transmission in space information networks: Architecture and key techniques," *IEEE Wireless Communications*, vol. 26, no. 2, pp. 17–23, 2019.

[7] Y. Su *et al.*, "Broadband leo satellite communications: Architectures and key technologies," *IEEE Wireless Communications*, vol. 26, no. 2, pp. 55–61, 2019.

[8] D. Petrie, J. Rosenberg, A. Johnston, R. Sparks, and R. Mahy, "A Call Control and Multi-Party Usage Framework for the Session Initiation Protocol (SIP)," https://rfc-editor.org/rfc/rfc5850.txt, May 2010, [Online; accessed 30-Aug-2021].