Technical Report:

A Commentary on the Nature of Mobility and Connectivity in Space and Implications on the Design of Space Network Protocols

Prepared As a part of
NASA Faculty Fellowship Program
August 2005 by:

Dr. Javed I. Khan
Media Communications and Networking Research Lab
Department of Computer Science
Kent State University, Kent OH 44242
javed@kent.edu

<u>NASA Colleagues</u>

Brenda L. Ellis, Larry McFarland, and Chuck Putt

Office of the CIO
Computational Environments Branch
NASA Glenn Research Center
21000 Brookpark Road, MS 142-1
Cleveland, Ohio 44135
Phone: (216) 433-5214

Technical Report 2005-08-01
Internetworking and Media Communications Research Laboratories
Department of Computer Science, Kent State University
http://medianet.kent.edu/technicalreports.html

A Commentary on the Nature of Mobility and Connectivity in Space and Implications on the Design of Space Network Protocols

Abstract: *What if the righteous place to implement "hop-and-forward" mechanism is in the network layer but not at application or transport layer? Is it possible that congestion is an earthly phenomena- byproduct of human freewill and hardly be an issue in cold lifeless space until it is deeply colonized? Can one conceive a space routing protocol which will require zero inter-router communication and still route packets with absolute precision? In this report we examine the technical nature of the notion of 'mobility' and 'connectivity' in the context of space networking and seek answers to the questions like above.*

1. Space Internet

NASA's current space communication technologies have evolved as mission specific systems from individual mission considerations. The success of the Internet on the earth has added a new vision towards developing an integrated communication infrastructure for space. The idea is to meet the demands of a new age in-space access, add technological vitality with mass participation in space activities and to offer cost effectiveness. While various wings of NASA and also the broader science and engineering communities are already converging to explore the idea of integrated space internet- the overall idea has got a vital boost in the recent US presidential articulation of the *Vision for Space Exploration*.

A key element of the new age space exploration is mass access. A new digitally savvy generation is emerging who will demand order or magnitude more access into the space systems and the exploration experience. This mass access view is radically different from the way NASA currently conducts its communication. Not to mention its current communication architecture is inadequate to support such access.

In a recent milestone work Bhasin & Hayden [1] have defined the vision of a space communication infrastructure in more architectural terms. It identifies an initial set of elements for Space Communication Infrastructure (SCI) with their possible space placement. The proposal envisions constellations of communication relay satellites, sensor web inter-spacecraft linkage elements, autonomous ground relay elements, and a host of terminal elements to offer 24/7 end-to-end connectivity to our solar system. It has identified various orbits around earth, moon and mars, orbits around Lagrange points between earth and sun, earth and moon, and earth and mars for the placement of SCI elements. It will provide a shared communication infrastructure for various manned and robotic communicating elements such as space crafts, aircrafts and constellation sensors deployed around the planet orbits and their surfaces, and other astronomical objects of future interest. The architecture also proposes an initial set of interconnectivity links identified for backbone, access and proximity links.

Now the focus is on building the network. The next step is the identification of the network software architecture. Like any evolutionary system, the goal is to use the proven Internet technology as much as possible and to evolve it in critical points and to deploy it in space Internet. The engineering risk requires as much as possible to use the standard, low-cost, off-the-shelve protocols. However, space offers several novel challenges which defy some basic assumptions of earth Internet. Current protocol suit faces fundamental qualitative challenge and quantitative performance issues if deployed in space in as-it-is form.

Current Internet communication- embodied in the TCP/IP suit is indeed a collection of over 200 protocols and standard applications. Below is just a sub-list of very widely used protocols which makes today's Internet system. Given the complexity of future networks, almost all of them probably will be considered in a space network as well.

Layer	Protocols
Applications:	Ping, Telnet, HTTP, SHHTP, NTP, GPRS, RTP, SMTP
File Transfer:	BitTorrent, NFS, FTP, MFTP, CFDP
Transport:	TCP, UDP
Security:	SSL, SSH, SCP, Mobile IP Security
Network:	IP, IPv6, MIP, OSPF, BGP4, SNMP, ICMP, NAT, DHCP, DNS

Also CCSDS has developed four protocols for a possible CCSDS network architecture [2] for space use SCP-FP, SCP-TP, SCPS-NP, SCPS-SP. If past is any indication, it is important to note that while the space communication system will be built (perhaps in next 10-15 years?) it is highly likely that several sweeping new Internet protocols will further emerge right during this infrastructure build-up period. (perhaps BitTorrent[1]?). The design decisions taken today will greatly affect this future.

It seems, over the last few years several of the above protocols have been tested onboard via number of projects such as UoSAT-12, U0SAT-12, AlSAT-1, BILSAT, CHIPSat, etc. in space though using very basic network scenarios [3]. However, yet a very challenging aspect in the design of space network is to obtain a cross-layer perspective. Network protocols are designed with layers perspective. A design decision on one protocol might perform well from its own layer perspective. But, it might be inefficient from the perspective of a different layer. Often various layers perform redundant operation- and a particular choice in lower layer can unwittingly affect conditions at

[1] A very interesting new file transfer protocol invented by Cohen and now spreading like wildfire over the Internet. It defies the 'law of congestion'- the speed of file transfer increases the more people wants to download a file! Consider millions of K-12 school kids want to download a video image from a mars event.

other/upper layers. It is very important to understand the integrated dynamics of a complex system.

The special features of space are special brand of mobility, intermittent/periodic connectivity, long delay, power awareness, high level of noise, CPU, and bandwidth constraints, and path asymmetry, etc. The design of the space Internet infrastructure is further complex because of the nature of space exploration program- which highly values fault tolerance, remote controllability, requirement for reservation and traffic schedule (space version of QoS), extremely high reliability, security, etc., in space communication. These requirements are expected to have significant impact on the design of network layers in space communication and may add new protocols, extension, and/or significant provisioning in existing network layers [4]. However, all are not bad news. Indeed straight importing of ground protocols may not be recommended for space environment because some of the issues- such as scale, congestion, and event prediction are much easier and may offer significant advantages and opportunities to space communication. Ground protocols often pay heavy price for aspects which may not be present in space. A cross-layer perspective is very important to identify such opportunities.

In this technical report I provide cross-layer perspective with respect to only two aspects of space communication- *mobility*, and *connectivity*. I only include views which probably received more isolated treatment in literature. I attempt a concise adequacy/inadequacy analysis from the perspective of all layers' joint operation. The re-engineering of the existing TCP/IP network to suit space will require many design choices to be made before completely knowing the future. This technical report may provide some insight to make possible ranges of design decisions from a holistic perspective of overall network systems engineering. The paper however, would not attempt to suggest any particular protocol, or solution within its limited scope.

2. Nature of Motion

Mobility is a central aspect in space communication. Almost all assets- satellites, space craft, ground stations, are in continuous relative motion with respect to each other over vast distances. This flavor of mobility however is quite distinct from the mobility that has been handled on earth's TCP/IP protocols.

On earth-network several protocols are used in combination to handle mobility. The core protocol is OSPF and Mobile IP. Also BGP, DHCP, NAT, DNS play interesting roles in handling current mobility scenarios. Also research level protocols such as IPMN, Freeze TCP, WTCP, Snoop etc [5] have been proposed features of which can potentially find use in space mobility. It is important to note that each mobility protocols have specific advantages and infrastructural assumptions. In their current state these are logically incomplete to handle space network- and also these incur performance penalties.

2.1. All Relative Motions are Not the Same

2.1.1. Issues

In space scenario the core network components are all in continuous relative motion. This includes the space proximity and inter-platform backbone networks. In contrast, on earth the core infrastructure components have no relative mobility. Thus the core network of the Internet is static. The relative motion is confined to some elements at the edge. I will call it edge-rider's mobility. In earth scenario (2005) this edge mobility has hop count just one- at the very last hop.

However, new systems are emerging where a network will move as a platform I will call it *platform-mobility*. *Personal Area Network* (PAN), Department of Defense Vessel Network, where an entire edge-network rather than just the edge-element is moving- are examples of this class of mobility. Any spacecraft onboard network including planetary surface network in the space will be in this class. Architecturally it means at least border router to be mobility-enabled.

A third form of mobility is where everything in a network is in motion relative to each other. I will call it *core-mobility*. Architecturally it means all router's and hosts to be mobility-enabled i.e. the protocols specific to all routing should be mobility-enabled. The planned core space network which will use relays onboard satellite constellation and ground stations will be an example of a network of this class.

2.1.2. Impact on Protocols

The oddity of TCP/IP networks in handling mobility originates from the very basic use of IP address. It bundles the *identity* (who is it) of a host and the *routing information* needed to get to this host in this one identifier. Mobile engineering over TCP/IP network is nothing but the attempt to decouple this dual use of IP address. DHCP has been designed to handle edge-mobility for client like hosts. Client systems can easily discard the identity. Thus when attached to a new location it can easily take on a new identify via new networks DHCP and get the updated routing information identifier (the new IP address). However, a server system must have a well known self identity. Thus, it cannot discard the old IP. Thus, it needs to incapacitate the routing information embedded in its old IP address. Thus, a server system requires Mobile IP or IPMN like protocols to obtain the routing. IP incapacitates the routing information found in the original IP by tunneling. IPMN maintains a chain of identity. It finds the original node based on old IP, but than as node moves it keeps on substituting the old IP address with the new one at the very end point.

The question of logical completeness of current protocols becomes little more interesting in the case of platform mobility. Lets' assume first that an area border router (it does not matter if those are BGP or OSPF) of a platform looses connectivity with a certain neighboring network, and then establishes contact with a new network. Logically, it can discard the routes from the previous network and exchange the new routing table with the new network. This system will work so long the handoff does not jump the hierarchical address allocation boundaries. But will they operate without trouble in the general case where address aggregation is in place (BGP's CIDR, of OSPF's subnetting)?

The same question will arise in core-mobility but with increased concern for route table explosion, thus route lookup speed, and hence in the effective capacity of the core, if just the old routing regime is packed and shipped on a set of moving platforms. Network wide address aggregation might have to be seriously blocked. If all mobile nodes within a network with core-mobility can be kept inside one network- perhaps the logical connectivity problems can be avoided- though it will restrict the size of it. One can conceive some interesting configurations- though it is not the intention of this report to discuss specific solutions.

2.2. Deterministic vs. Nondeterministic Motion

2.2.1. Issues

The motion in space is almost pin-pointedly predictable and is hardly stochastic. The relative motion among communication satellites, spacecrafts, ground stations, all are deterministic and precisely computable. In sharp contrast, the edge mobility found in the earth-network (or in a single-platform network) has higher degree of stochastic unpredictability. It is important to appreciate that elements of communication are in motion- but the system is based on a static mathematical model. While in earth-network the motions are not easy to model with any static formulation. Currently, communication events (including traffic) are pre-planned and prescheduled with at least few hours lead. In future with more spontaneous control systems- stochastic communication may increase in space communication. But, it may never be as severe as on the earth.

2.2.2. Impact on Protocols

Determinism is a very positive feature of space network. It will be pity if space network fails to take advantage of it. Stochastic behavior has costly implication in communication protocol design. The earth protocols are to be foundationally "reactive". All requires post-event communication to continuously "keep" up with the changes. OSPF incurs significant communication cost to exchange this post event information which is called route updates. Information has to be fast communicated between elements after each significant relative motion to ensure logical correctness of routing system. Also, often expensive stochastic guessing methods have been employed to enable some "pro-activity" and limit certain critical performance penalties (such as hand-off, jitter) in this highly unpredictable world.

In contrast, a predictable system is blessed. It can be "proactive", and conceptually can replace many communications simply by local computations. A classical OSPF or BGP routing 'system' is potentially in incorrect state for the time it takes the change of local state information to propagate system-wide. Space has very large end-to-end system wide delays. A 'reactive' routing system will therefore be at 'fault state' for significantly longer period of time- for the same fault. Pro-activity can remove this 'outage' time. A deterministic model based routing can reduce guesswork and reduce computation.

Deterministic routing protocols however, will require clock synchronization using protocols such as NTP. NTP seems to have performed well in space [3]. It will require onboard algorithms to compute route, and one time direct communication from a central point to obtain the model parameter, schedule update, etc.

Combining the two aspects, it seems that with platform-mobility and core-mobility the reactive protocols will face increasingly more complexity in space. But the deterministic properties of space mobility- if used by a new set of proactive protocols, can dramatically simplify handling of astromobility. Compared to the chaos on earth communication, the space communication scenario seems to be so much a-priory known, schedulable, and comfortably computable- who knows it might be possible to design routes with absolute precision.

3. Notion of Connectivity

3.1. Issues

A second aspect of space network is that the connectivity between various backbone and proximity network elements is intermittent in nature. All communication requires some form of connectivity. TCP/IP network have been designed with a specific assumption of *concurrent link* connectivity- where at least one pathway exists between the source and sink where all the links of the paths have connectivity simultaneously.

Concurrent Link-based Connectivity: *A type CLC connection exists from a point A to a point B passing through any intermediate point H if there are simultaneous type-CLC connections from A to H, and H to B.*

CLC connections are symmetric. Connection from A to B implies connection from B to A. Thus, routing computation is simple.

This core CLC assumption is situated in the IP route table calculation software of TCP/IP networks. Thus, combined effect of all the TCP/IP protocols is such that if a CLC connection does not exist then all other protocols assume the lack of any connectivity cease to support any communication.

However, more general notion of communication is possible with weaker notion of connectivity. A system with finite communication delay can communicate from point A to point B even if all the intermediate links are not up concurrently. One such form is Ripple Timed Connectivity.

Ripple Timed Connectivity: *A type RTC connection exists from point A to a point B passing through any intermediate point H, if there are type RTC connections from A to H, and H to B, where the later exists at time until* $t_{HB>}\ t_{AH}+d_{AH}$ *where* t_{HB} *and* t_{AH} *are respective the time instants when the connections existed and* d_{AH} *is the communication delay to arrive from A to H, and the data can be stored at H for the wait time gap* S_H *which is difference of time of the equality.*

All type-CLC connections are strictly a subclass of type-RTC connections. But the converse is not true. RTC connections are not necessarily symmetric. RTC-connection from A to B does not imply RTC connection between B to A. All type-CLC connections are strictly a subclass of type-RTC connections.

There are other interesting notions of connectivity, however can be considered out of scope for this paper. Guaranteeing earth like type CLC-connectivity is considered very expensive. The number of satellites might to too high to afford in short future. Though, that will be the eventual goal, but it will be years to get there. Thus space exploration has to work on the assumption of RTC-connectivity. RTC-connectivity implies four artifacts (a) significant per/link downtime, (b) occasional lack of bidirectional connectivity even when there is one way connectivity (c) inequality of bandwidth in two directions and (d) increased effective delay. There are also legacy operational reasons which makes current space network asymmetric in uplink and downlink capabilities or completely different routing (done manually) of the uplink and downlinks [3].

3.2. Impact on Protocols

Space links satisfies the RTC-connectivity but does not satisfy the conditions of CLC-connectivity. This is a logical gap in current TCP/IP system if ported on space. I will first discuss the impact in IP and then on transport.

At the very least two things have to be done to mend this IP level gap and keep packets moving. (these effects both UDP, and TCP). First, the data forwarding mechanism must include prolonged **data buffering mechanism** (data stops).

Data stops can be implemented either at application layer (such as proposed by several hop-and-forward file transfer mechanisms), at transport layer (using some combination of UDP) or by the generalizing of the current IP. There are advantages and disadvantages of each of the approaches. If implemented at application layer it would require transport layers to be present in space routers and be involved in the solution. Its advantage is that secondary storage management can be realized with relative ease at application layer.

However, it seems that an alternate is to modify network layer (or IP). IP is nevertheless the 'righteous' place for doing anything that has to do with the packet forwarding task. It indeed is very attractive performance point of view. It can be potentially handled by much simpler extensions of input/output buffer management of IP. Several schemes are possible but those are out of scope for this paper. In either approach it has to be minded, that buffering data inside network- whether at application layer or in IP layer, is potentially an exponential resource hogging proposition. In some cases is it possible that resource constraint might push the extended data buffering task back to the sending end-point. Only selected stops where type CRC-connections would exist at least between the CTC-connection hops might sometime do some buffering.

The second logical gap is that space routing mechanism must be aware of **packet traps**. Current routing algorithms can potentially run into serious unstable state where packets will get trapped in a maze of RTC-connections

What is the impact of the RTC-connectivity on the upper layers? Direct modification of IP does not guarantee efficient communication in the layers above. Lack of simultaneous bi-directional connectivity can be a logical as well as performance drag for protocols such as FTP, TCP, HTTP and all security enhancing protocols which depend on challenge/response. Several proposals such as PBF, MFTP, CFDP has been designed to operate mostly with one-directional link. There can be other cross-layer ripple impact on the functioning of other protocols- (including ping, Mobile IP, UDP-RTP).

In space mobility scenario the actual time for uni-directional connectivity might be a small fraction of the RTC-connectivity up time. However, the problem is when long delay is mixed with request/response based protocols- then it has similar impact as of uni-directional connectivity. On top of that, RTC connectivity increases communication delay by the cumulative gaps S_H at each hop. This can have quite adverse cross-layer effect. Thus, all upper layer entities which uses communication time estimate must prepare against this gap accumulation.

4. Conclusions

Space network has several more distinguishing aspects besides the two discussed. A frequently noted feature is that many of the hops and links have long delay compared to networks on earth. However, the longer delay by itself does not seem to be at odd with any fundamental assumption of TCP/IP's design (like the other two previous issues). However, it can degrade the performance in few cases. TCP 'tune' its many internal complex operations based on various estimates (RTT, queuing delay, etc.) on the observation of return trip delay (a random variable). Many of the estimates are based on the assumption of smooth variation of this random variable. TCP measurement system is expected to operate without serious problem if the variation is not random, but it can potentially become erratic and loose efficiency in a network when its different parts demonstrate significantly unequal delays at different times. Additionally, in a mixed delay network (moon-mars-terrestrial) remote hosts with relatively longer delay will receive unfairly lower share of bandwidth than the ones with relatively shorter delays. Because, its additive increase process marches at the pace of respective round trip time.

A second distinguishing aspect of space network is probably the scale. Though several additional countries (likely China, Korea, Japan, Brazil, India) besides USA, European Agency and Russia are expected to join in the exploration, but still it is never expected to be crowded like earth. How many IP (V6 presumably) elements will be deployed in space by the year 2050? More than 1000, 10,000, or 1,000,000? If I have to bet, I would bet in the middle. But even it is the highest one it seems various earthy components (protocols, hardware, memory capacity, etc.) are already way more "scalable". Radiation and environment hardiness requirement may keep the space components 10-100 times slower than their earthly versions. High number of potential concurrent users and the stochastic unpredictability in the traffic model are the two principal ingredients that contribute to the network congestion. Thus it is possible space network will have to worry surprisingly less about congestion if more computation is performed and advantages are taken from traffic planning can capacity accommodation.

Acknowledgement: This report has been prepared as a part of a summer fellowship completed by the author at NASA Glenn Research Center (NGRC). The author is indebted to many colleagues at the center and in space exploration team and particularly including Kul Bhasin, Brenda Ellis, Larry McFarland, Chuck Putt, T. Luu, R. Slywczak, K. Johnson, J. Follo, P. Keith and C. Tran for ideas and hours of stimulating technical discussions. The author would also like to thank Dr. David Kankam, and Tony Rusnak for the program support.

5. References

[1] Bhasin, K., and Hayden, J. L., Evolutionary Space Communications Architectures for Human/Robotic Exploration and Science Missions, Space Technology & Applications International Forum, (STAIF-2004), 2004, Albuquerque, NM.

[2] Robert Durst, Space Communications Protocol Standards Overview Protocol Standards Overview, April 1998, [http:// http://www.scps.org/scps/ Documents/ SCPSoverview.PDF

[3] Hogie., Keith, E. Criscuolo, R. Parise, Using Standard Internet Protocols and Applications in Space, Journal of Computer Networks, April 2005, 47 (2005) 603-650.

[4] Javed I. Khan and Raid Y. Zaghal, Interactive Transparent Networking—Modeling Examples of Snoop and WTCP Protocols, Computer Communications, Vol. 28, Issue 6, pp. 702 – 711.

[5] Raid Zaghal, Sandeep Davu, and Javed Khan. An Interactive Transparent Protocol for Connection Oriented Mobility - Performance Analysis with Voice Traffic. Third International Symposium on Modeling and Optimization in Mobile, Ad-hoc and Wireless Networks – WiOpt'05, Riva Del Garda, Trentino, Italy. April 2005.