HRPS: A Hybrid Routing Protocol for Space Based Networks

Alex Bordetsky, Joseph C. Horvath, Richard L. Tackett, and Cameron R. Yaste

***Abstract*—Just as terrestrial networks have shifted to routed, message centric networks, the satellite networking regime is beginning to transition in the same manner. An analysis of current networking protocols is conducted and determines that the unique mixture of static and dynamic links present in orbital constellations makes them equally unsuited for static network protocols and MANET protocols. The authors explore how Social Networking theories can be applied to the routing decision making process in the space environment. Utilizing aspects of Adaptation and the Weak Tie relationship, a new protocol model is proposed leveraging the predictability of orbital mechanics to decrease network overhead and increase network efficiency.**

***Index Terms*—Adaptation, Satellite Networks, Strong and Weak Ties.**

I. Introduction

THe use of satellites to relay communications has been in existence since the 1950's. Early systems were composed of single satellites with one or multiple frequencies that were used as repeaters to echo the signal back to the ground. While more complex than the first generation communication satellite, the current systems use onboard processing to boost the signal before retransmitting, but are still transponded systems. These current satellites can handle a high number of users through multiple transponder channels, but must send the signal to the ground for it to be routed before being sent back to the satellite for delivery to the final destination. This traditional bent-pipe system has been the standard architecture for many years but is analogous to the Plain Old Telephony Service (POTS) structure in that a discrete channel is required for each transmission. Just as the Internet opened new avenues for information to travel over terrestrial pathways, the use of routed communications in satellite networks will likely prove to be the new paradigm.

In order to provide routed communications in a constellation of networked satellites there are number of issues that must be contended with. Specifically, a new model must be created for how to route information in this unique topographical environment of orbital nodes separated by great distances and in constant motion. The goal of this paper is to apply two social networking theories to this problem: Adaptation and Strong/Weak Ties. First we will show that adaptation requires specific steps that can help to form a new routing model for satellites. Second we will explain how the use of specific orbital models can leverage the advantages weak ties play in strengthening networks.

II. Analysis of Existing Protocols

A. Research Question

Given a mixture of various satellite constellations, which type of existing protocol, proactive or reactive, is best suited to handle the predictably dynamic links between networked orbital nodes?

B. Methodology

Because designing, building, and launching a large constellation of satellites in order to test network architectures is unfeasible, a multidimensional modeling approach was used. The inherent complexity of the various phenomenology at work within both satellite communications architectures and within information networks drives the need for robust modeling and simulation.

In order to model and simulate the behavior of a space based information network, a high fidelity discrete event multi-resolution model must be developed. The requirement for fidelity is self-explanatory. A discrete event simulation allows the model to progress gracefully through time, allowing for human cognition to participate in the analysis since the model behaves as if it were real life and can be observed as such. Multi-resolution modeling is building a single model or a family of models that inform at various levels or complexity or within several interrelated but distinctly separate environments [1]. In this case, the physical laws which determine the behavior of a constellation of satellites in orbit around the Earth communicating via electromagnetic waves are distinct from the physical laws which govern the communication of nodes within the information network overlaid on that satellite constellation. The two systems are interdependent and yet distinct. This drives a need for a comprehensive modeling methodology that allows the interplay of these two systems in a discrete event simulation following a single time step and common modeling environment.

An additional motivation for multi-resolution modeling is that complex adaptive systems often exhibit emergent behaviors, or regular coherent behavior at the macroscopic level that are not readily apparent or understandable in terms of the microscopic laws that govern the system. These emergent behaviors make complex adaptive systems both difficult to experiment with and oftentimes counterintuitive at the macro level. Multi-resolution modeling allows for lower resolution models to be combined in a way that preserves the emergent behavior exhibited at the macro level without sacrificing the fidelity of the model as a whole. High resolution models tend to require such complexity that their scope must be constrained, making them unwieldy for experimenting with emergent phenomena in complex adaptive systems. In this way, it is preferable to use multiple specialized modeling tools aggregated into one multi-resolution model.

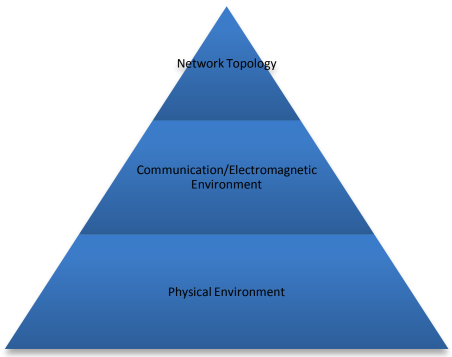

Fig. 1 Model Aggregation Hierarchy

Given the importance of the physical environment modeling, that is the type of modeling software that we focused on. So the requirements for the nascent modeling environment software are discrete event physics based modeling and simulation program with a robust capability for extensions into other forms of modeling such as communications and networking. We desired to have a single modeling environment with a common graphical user interface (GUI). As a result, we chose Analytical Graphics' (AGI) Satellite Toolkit (STK) for modeling the physical and communications environments and Qualnet to model the network behavior. Qualnet was chosen specifically due to its ability to interface with STK and run within a common GUI.

Using STK, four distinct constellation models were developed. The Geosynchronous (GEO) constellation consists of four satellites in geosynchronous orbit connected via inter-satellite links and represents the most static constellation model considered. The Low Earth Orbit (LEO) constellation consists of 66 satellites in the same configuration as the IRIDIUM constellation, which is an Adams-Rider constellation of optimally phased polar orbiting satellites. The LEO constellation, consisting of a single orbital regime, represents a dynamic constellation, but one with nearly all constant static links. The Hybrid constellation consists of the combination of the LEO and GEO constellations into a single constellation with 70 satellites and represents a more complex constellation. The Hosted constellation consists of a collection of satellites chosen for their available size weight and power (SWAP) that were launched in the past 10 years. The result of the hosted constellation was a model of seemingly randomly chosen satellites and represents the least deterministic approach.

Using QualNet to model the networking environment, four network protocols were tested in each of the four constellations. OSPFv2 was modeled as a non-MANET protocol and was representative of a static networking protocol. Then three MANET protocols were tested, specifically AODV, DYMO, and OLSR. For each protocol, traffic generators were used to emulate simple network traffic.

C. *Results*

The modeling effort generated 16 sets of data; four constellations each with data for four different protocols. Each set of data was then analyzed across multiple variables with an emphasis on network efficiency as measured by the ratio of data to overhead packets transmitted. This analysis focused on making relative comparisons between protocols within constellations and constellations within protocols. The network efficiency ratio was analyzed across each node.

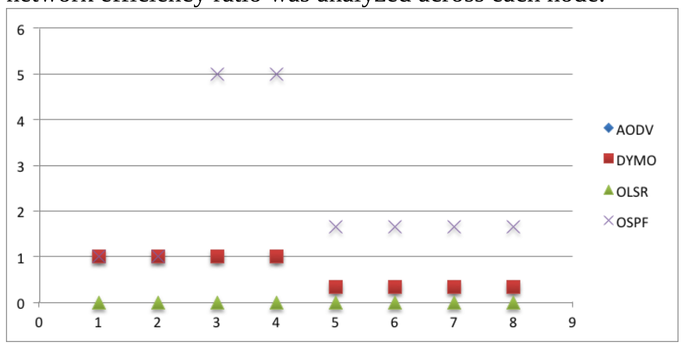

Table 1 Node Efficiency in the GEO Constellation

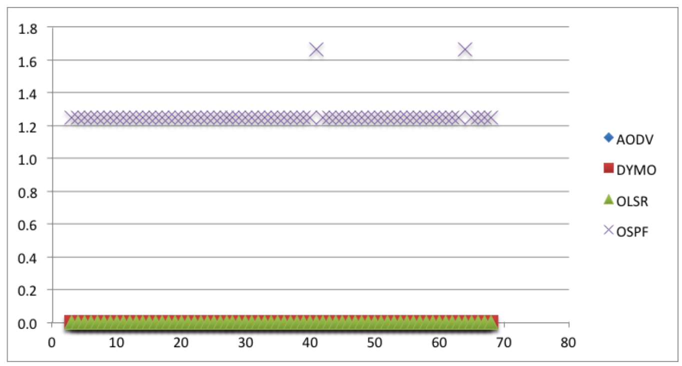

Table 2 Node Efficiency in the LEO Constellation

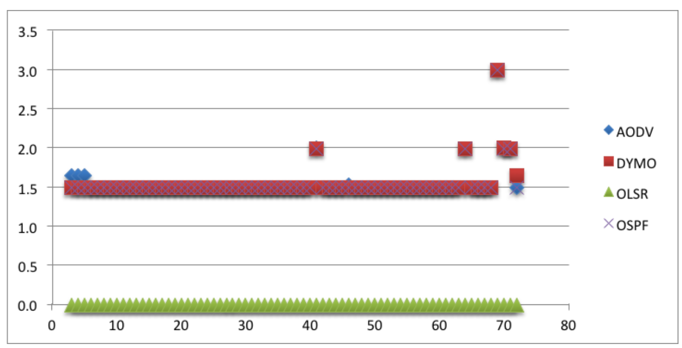

Table 3 Node Efficiency in the Hybrid Constellation

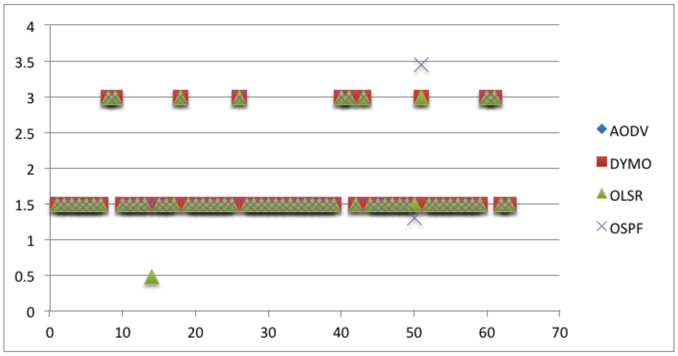

Table 4 Node Efficiency in the Hosted Constellation

D. Conclusions

The results, however incomplete, illustrate that no constellation was a single protocol ascendant and it is evident that none of the protocols are well suited to the space based network.

A network of satellites operates in a very unique combination of predictable static links and variable dynamic links in such a way that protocols designed to operate in static networks have trouble managing the dynamic links and protocols which are designed for dynamic networks, such as MANET's, are overly inefficient for the level of dynamic links in the network. Satellite networks also have the unique attribute of being predictably dynamic because their motion is based on orbital mechanics and are thus extremely predictable within their specific orbital regime. The dynamic links arise when links are made between satellites of differing orbital regimes.

A protocol which is optimized for a satellite based network should, with a minimum amount of network overhead so as to preserve the limited bandwidth of satellite communications links, be able to predict with some level of assurance the likely future network topology based on orbital dynamics of known satellites as well as to efficiently manage the highly dynamic links between satellites of differing orbital parameters. In order to achieve this, the network must become a more adaptive system.

III. Adaptation

A. Definition

The concept of adaptation is not easily definable and is used to imply a number of ideas. Holland defines it by saying, "[Adaptation] is a study of how systems can generate procedures enabling them to adjust efficiently to their environments." [2]. This definition involves the interaction between an entity and its environment and it includes qualitative assumptions about the process by using such words as "efficiently". While this definition is sufficient for *defining* and *identifying* adaptation, it is wholly insufficient for the purposes of *designing* adaptation into a complex system.

Adaptation is present in most emergent complex systems, but setting out to create a complex system that automatically and intelligently adapts has proven extremely difficult. In order to set upon the task of designing adaptation into a system, the process of adaptation must be examined and defined in such a way that it may be reproduced. The process of adaptation is comprised of six steps, each step being a prerequisite to the next and that each step adds additional levels of complexity to the system.

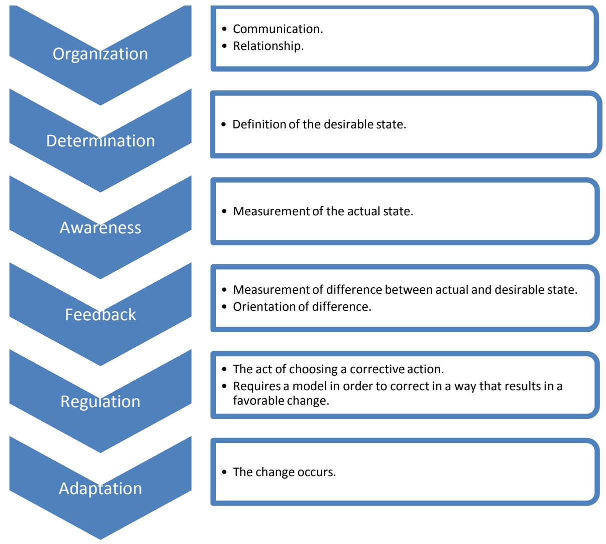

Fig. 2 The Adaptation Process

B. Organization

The first requirement for adaption is the presence of organization. The best definition of organization is provided by W. Ross Ashby [3] as a conditionality between interrelated entities, and that as soon as a relation between two entities exists such that one's state is conditional on the other's state, a component of organization is present. In this understanding of organizations, it is important to note, as Ashby points out, that the presence of organization makes no claims to the relative value of the organization. Whether the organization is a "good" one or a "bad" one is immaterial to the definition of organization for our purposes.

If two entities exist in perfect apathy to each other such that no relations can be drawn between them, or if two entities exist in perfect isolation such that no medium exists for them to communicate then there is no organization. Therefore, organization must exist and can only exist between any two entities that are both capable of communication and have a relationship between them. If two computers are designed to operate in concert but have no medium for communication, then no organization can exist since a change of state in the first computer can cause no change in the second computer.

The nature of this relationship must take the form of a natural law so that when a change in A is communicated to B, a resulting change in B must always take place and the change must be such that an infinite number of perturbations under identical circumstances will always result in the same change in B for any one change in A. This can be as simple as an array of hard drives that are set up to mirror data, or as complex as the interaction between the flux density of cosmic rays and the Earth's cloud cover. The point is that the interaction follows natural laws that allow results to be reproduced, or predicted.

C. Determination

The second step in the process of adaptation is determination. Ashby set out in 1962 to describe the principles of self-organization. He built upon the concept of organizations and defined two ways in which a system can be "self-organized." The first method of self-organization is the process of communications and relations forming between entities. As many individual and separate parts form relationships between each other, an organization emerges. This is the process that we understand today is responsible for the coalescing of life out of constituent chemicals. It is random, or at least, not directed and the only qualitative measure of such systems is whether they continue to exist. In natural systems of organisms, what is good or bad is simply the difference between what is replicating and what is dead. But, what biological systems and artificial systems that adapt share is that there is a definition of success. In an organism, the definition of success is continuing life. In an artificial system, success is a set of design constraints. Regardless of how success is defined and whether it is a natural emergent definition or a programmed objective, the act of defining the goal of the system is a critical step in the process of adaptation.

D. Feedback

The third and fourth steps are closely intertwined and together constitute traditional feedback. Step three is awareness. It is the ability to measure the actual state of the system. The fourth step is feedback, which Ramaprasad defines as "information about the gap between the actual level and the reference level of a system parameter [4]. Feedback represents the measurement of the orientation and difference between the desired state and the actual state. As an easy example, assume that you must monitor the temperature of water in a bathtub. Your definition of the desired state is "warm" and a thermometer placed in the tub measures your awareness. In order to consider your awareness as feedback, you must be able to translate the measured temperature into some state relative to your desired goal. So if you are unable to determine if 80 degrees Fahrenheit is above or below "warm" then you will be unable to classify such as feedback. Conversely, if your desired state is 80 degrees F, and your thermometer reads 75 degrees F, then you know that you must increase temperature. Therefore, for our purposes, feedback is the process by which measurement of actual state is translated into a form that is a relation to the desired state. If no relation can be drawn between desired and actual state, then no feedback exists.

E. Regulation

The fifth step is regulation. Conant and Ashby, through rigorous logical proof [5], concluded that in order for a regulator to succeed in regulating a system, the regulator must be a model of the system. The basic premise is that in order for a regulator to take action in the process of regulation, one must have a good idea of what actions will result in what outcomes. Therefore, whether preprogrammed into a system or built into its design, or injected with human interaction, a regulation to return a system to its desired state is predicated on the ability to model the outcomes of various actions. We have defined regulation as the act of choosing corrective action in response to feedback and so doing, requires a model to understand *how* to correct in a way that results in a favorable change.

As example, take the same bathtub. If your feedback is that the temperature must increase by 5 degrees, you would instinctively reach for the hot water tap, but assume for the sake of the example, that the taps are not labeled and so you do not know which is hot and which cold. You would have to test each to find out which one is hot. To add a level of complexity, imagine that you must do so remotely. So now you must run one faucet for a few minutes and observe the change in the temperature of the bathtub. You can measure the change, and then deduce which faucet is hot and which cold. This method is simple, and yet you should be able to get the temperature close to where you want it to be. If you were able to calculate such things as the heat transfer from the water to the air, the flow rate of the pipes, the heat transfer from the pipes into the walls, the amount of energy being input into the water in the hot water heater, and any number of other factors and could calculate the temperature of the tub for either faucet running for any time, then you would have a much better regulator of tub temperature.

The difference between the qualities of regulation is the fidelity of the model. The first example, just knowing which faucet was hot and cold represents a model that tells you the direction of a change vector in the temperature of the tub, but not the magnitude. It is also a binary control, on or off. The second is a much more complex model and truer to real life and thus makes a better regulator of the bathtub temperature system.

The simplest regulator is a binary control with feedback system and only knows one action to correct for deviation from the desired state. A more complex regulator might have a dozen or more available actions available to it and might be reacting to multiple forms of feedback but is still reacting. The most complex regulator is one that predicts changes in the environment and then models various actions to see which one will result in the appropriate action.

F. Adaptation

The sixth and final step is adaptation and it is the act of choosing a state and making the change. The act of adaptation can take one of three modes. As defined by Chakravathy [6], the three modes of adaptation are defensive, reactive and proactive. Each of these modes requires a different level of complexity in the overall system and in the decision-making process, whether human in the loop or automated regulator.

The most basic level of adaptation is defensive. This mode of adaptation attempts to insulate the system from the effects of the environment. Much of the actual regulation is done during the design of the system and emphasis is given to reliability in any anticipated environment. This is the kind of adaptation that is predominant in systems engineering and military systems. Much thought is given to the types of environment that the system might encounter and then the system is designed to avoid being affected in any detrimental way by that environment. This is also how modern satellite communications are designed. In a dynamic system such as the corporations that were the focus of Chakravathy, this mode of adaptation corresponds to a defend and hold mentality in business. If a business is not concerned with enlarging their markets or developing new products, then they are likely to focus on keeping what they have.

Reactive adaptation is more complex, and is the predominant mode of adaptation in automated regulatory systems. When a change in the system is detected, a response is automatically generated to bring the system back into acceptable parameters. These systems do not attempt to predict changes, they merely react changes. The strength of these systems is that they are relatively simple and are predicable. The weakness of these systems is that they are unable to respond to unforeseen changes in the environment. The bathtub example form earlier would be completely unable to respond to a loss of the water heater because such a thing is outside of its parameters.

Proactive adaption is the most complex. This mode of adaption requires models of high fidelity and the ability to model various actions to determine the best course even in situations that may not have been foreseen during design. These systems react to changes in the environment before they happen so that no loss of effectiveness of the organization is felt.

These principles of adaptation are appropriate whether discussing a communications system or a business. In terms of complexity, the systems that include humans are inherently much closer to operations like these simply because the human brain operates in this manner. When attempting to develop an automated system capable of intelligent adaptation, there is a much higher level of complexity involved. These kinds of systems will require a *feed-forward* system of proactive adaptation. This requires the ability to: sense or predict changes in the environment *before* they affect the organization, predict the effects of those changes, model changes in the organization to determine the outcomes of various courses of action, and the abilty to select the "best" adaptation to the situation.

G. Application

Based on Chakravathy's work in defining modes of adaptation, the following model is proposed as a basis for routing decision making in satellite networks.

Nature of Relation to Environment	Routing Decision Scheme	OSI Layer
~~Defensive~~	Transponded/Bent Pipe	Physical
Reactive	Link State	Physical
Proactive (Predictive)	Orbit Defined	Physical
Static	User Defined Priority Based	Application

Fig 3. Modes of Adaptation

In this model, the Defensive mode of adapation is dismissed as the traditional transponded satellite network model in which routing decisions are made on the ground. Instead a new, 3 stage, routing decision model is proposed based on Reactive, Proactive and Static modes.

The first stage is the Proactive or predictive mode. This stage of the model would update the routing table based on the expected position of the satellite using a sliding frame finite state machine. Once a satellite constellation is in space, the orbits are well defined and the crosslinks between the nodes can be predicted for any given time period.

The second stage is the Reactive mode. Based upon an expected state for each link from stage one, a metric would be used to identify which links are behaving as expected and adjust the routing table accordingly. It would also allow adjustments to be made to transmitter power or bandwidth as require to maintain link quality.

The final stage would be based on application priority as defined by the user. In this mode, routing decisions would match high priority packets with higher quality links and visa versa.

IV. Strong and Weak ties

A. Foundation

The definitive work on social networks and the interaction between the micro and macro levels is Mark Granovetter's *The Strength of Weak Ties* [7]. He defines Strong Ties as the connections within small groups of people. In other words, if subject A is in constant or close contact with subject B then there exists a strong tie between Subject A and B. He classifies these as people you have almost daily contact with i.e. family, close friends and co-workers. He goes on to define weak ties as connections between groups of people. For example, subject A is a member of group A and subject C is a member of group C. There does not exist a direct tie between subjects A and C. There does however, exist a tie between subject B (a member of group A) and subject C. Therefore there is a weak tie between groups A and C. It is these weak ties between groups that are the strength of the network and facilitate the flow of information from group to group.

Another example of the strength of weak ties was a study conducted by Stanley Milgram in 1967 in which he set out to define the social distance between two randomly selected people in the United States. His experiment was simple in design. First, two people were selected at random, one a resident of Wichita and the other a resident of Omaha. Once a participant received the folder they were to fill out the roster, detach one postcard and mail it back to the university (postage paid), if they knew the target person on a first name basis mail it directly to them, if not they were to mail the package to someone they thought might have a better chance of knowing the target, but they must know the person on a first name basis. Roughly half of the letters made it back to the university, the smallest distance was two hops while some had nearly a dozen but the median was 5.5, rounded up to 6 [8]. Obviously if the initial recipient and the target were on a first name basis with one another it was a strong tie between them. However it was the weak ties between groups that ultimately enabled the letter to reach its target.

The strength of weak ties was further examined in *Job Search and Network Composition: Implications of the Strength-Of-Weak-Ties Hypothesis* [9]. In this study

Montgomery interviewed business executives to determine how they found their current job. They were asked if close friends and family or an acquaintance that helped them find their job. He concluded that managerial workers found their current job through weak ties 27.8 percent of the time. He determined this was because close friends and families were likely to have access to the same information and in order to gain access to outside information they had to utilize weak ties.

B. Routing Schemes

Much work has been done to translate routing techniques of social networks to computer and communications networks, particularly Delay Tolerant Networks (DTN). Delay Tolerant Networks are ideal for translating and implementing social networking routing schemes due to their dynamic and volatile nature. With DTNs nodes are continuously joining and leaving the network. Additionally their respective link quality can vary throughout the life of the link. It is from these dynamic (weak) links that DTN's and mesh networks draw their strength. In traditional networks a link is relatively constant and always available for use unless it becomes too congested. Traditional networking algorithms like Open Path Shortest First (OPSF) route their traffic similar to the routing in the Milgram experiment, with the exception that routers advertise with whom they have connections to in order to route packets more efficiently. It isn't hard to think of packet switched network as a social network at all. Routers, and in some cases switches, can be equated to popular people in social networks. Along those lines the more friends a person has the more connections he or she is able to make. Likewise the more links a node has the more likely they are to form more links based on the fact that they are more likely to *know* the destination for a given message. It's slightly harder to correlate this example to DTN's because the topology of the network is constantly changing. Each node is likely to be in motion and therefore the nodes it is in contact with at any given time will change as well. At first glance the history of the contact may seem meaningless, but when collected, time predictive routing tables can be computed. It should be noted that these tables shouldn't be appended to indefinitely at the risk of being populated with contacts that are only made once and ultimately becoming unusable.

The authors of *Know Thy Neighbor: Towards Optimal Mapping of Contacts to Social Graphs for DTN Routing* [10]analyze and evaluate several DTN routing protocols based on social networks to map the nodes of a network based on different criteria. The paper evaluates SimBet and BubbleRap under several contact generation models and measure the performance of each protocol. Two approaches to time window based aggregations of contact, growing Time Window and sliding time window. Growing Time Window as contained in the original SimBet uses betweeness and similarity are calculated over a social graph where there is an edge between two nodes if there was at least one contact at any time in the past. Betweeness of a node is defined as the fraction of the shortest paths between each possible pair of nodes going through this node [10]. In sliding time window a limited time window is used for the centrality value calculations in BubbleRap [10]. Essentially, the time is split into six-hour windows and only the contacts in the last six hours form the edges of the graph. The sliding time window approach helps to regulate the graph to regular contacts and eliminate random contacts.

Another DTN routing Protocol is Socially Selfish Aware Routing (SSAR) [11]. In SSAR the willingness of a node to route packets for another node is based on the social tie between them. Furthermore, nodes will forward the highest priority packets first in order to increase their "social status". While it certainly has it's merits, SSAR may not prove viable in a space based scenario based on the fact that users may be required to manually assign values for "stranger" nodes via an interface. This would be difficult to do for a satellite constellation because of the limited access to the satellite from the ground for programming and configuration changes.

C. Application

The traditional approach to satellite communications has been and currently remains to be the "bent pipe" approach. In other words, signals are unprocessed and only transponder. This is an inefficient approach to routing data. Consider the following scenario: A user on the ground sends a message to a distant user, because there is no terrestrial infrastructure connection the two users, satellites are used to relay the message. The message goes from the user on the ground to the satellite where it is then multiplexed and transmitted to the ground where the signal is then de-multiplexed, the ground station then determines where it needs to go and could possibly send it on to its final destination via terrestrial infrastructure, but this is not always the case. Often, the message is then multiplexed again at the ground station and sent back up to the satellite to be retransmitted to its final destination. The reason behind this approach is cost and risk driven. Keeping the routing capability on the ground reduces the complexity of the design of the satellite, reduces the risk of having an on orbit failure rendering the satellite useless, an ultimately reduces the cost of the satellite. However with advances in software definable radios, small form factor routers, newer routing protocols and robustness of design, coupled with the ever-growing demand for access to large amounts of data by even the most disadvantaged user, the need to transition to packet switched satellite networks is becoming apparent.

The relatively long propagation times, coupled with other satellites constantly coming in and out of view requires a robust delay tolerant protocol for a packet switched satellite constellation to be viable. There are a number of approaches that can be utilized to provide worldwide coverage. One approach is a constellation of cross-linked satellites orbiting in low earth orbit. Another approach would be to place three or four large satellites at geo equipped with crosslinks. Another approach would be to deploy several cross-linked satellites in a highly elliptical orbit. Still another approach is a hybrid all the approaches. While each approach has it's own advantages and disadvantages, those discussions are beyond the scope of this paper.

Examining the low earth orbit (LEO) approach through the lens of social mapping of DTNs exposes some familiarities already discussed. Satellites in LEO orbit the earth at tremendous speeds, about 7.5 km/s on average. Of course this speed will vary depending on the altitude of the satellite. Satellites that are in the same plane have fixed positions relative to one another and therefore the connections between these satellites can be classified and strong ties. The tremendous speed of the satellites becomes a significant problem as satellites move above 60 degrees north latitude and below 60 degrees south latitude. Here the relative velocities between satellites create a Doppler shift in the transmit and receive frequencies so great that the radios can't compensate for it and the link disappears. Because these links between planes isn't constant these links are classified as weak ties. However, as demonstrated in both Granovetter and Montgomery's works, the strength of the network resides in the weak ties. Without the weak ties between satellites in different planes, routing would not be possible without the use of a significant buffer for a store-carry-forward approach or the use of a ground station which is counterintuitive to a packet switched satellite network. Furthermore the links between users on the ground and the satellites are weak ties because the satellites are in constant motion and the user switches from one satellite to another as they come in and out of view.

If geosynchronous (GEO) satellites are included in this network the connections between the GEO satellites and the LEO satellites can be classified as weak ties as well based on the fact that the connections are not persistent. Again, the strength of the network lies in the weak ties.

Coupling the strong and weak ties of a routed satellite network with a sliding time window as stated in [5] and the network now has the ability to predict links before they become available to increase the routing efficiency of a DTN.

To date there is only one satellite on orbit that is capable of performing routing onboard the satellite. INTELSAT 14 is a geosynchronous satellite that has a hosted payload, Internet Router In Space (IRIS) designed by Cisco on board. The inclusion of an internet router coupled with an IP modem on a commercial communications satellite allows for cross beam and cross band communications that are not seen in satellite communications. At least not without bouncing between the satellite and a ground station a few times. Because there is only one IP enabled satellite on orbit a true packet switched satellite constellation network does not yet exist.

V. HYBRID ROUTING PROTOCOL FOR SPACE

A. *Overview*

In order to more efficiently route network traffic through the unique space environment a new protocol is proposed, Hybrid Routing Protocol for Space (HRPS). HRPS is an adaptation based, hybrid routing protocol that leverages the unique predictability of orbital mechanics for use in a space-based network. The protocol is comprised of three layers a proactive, a reactive and a priority based decision matrix. These layers seek to maximize efficiency through the use of node location knowledge and minimize overhead required to maintain routes.

B. *Tier One: Proactive*

The Proactive layer is based on the premise that satellite orbits are well defined and the position of a satellite, and therefore the node itself, is well known at any given point in time. This can be predicted based on both the previous position and velocity data and verified through GPS or other telemetry measurements. The proactive layer takes a given satellite constellation and breaks it into a series of well-defined snapshots over time. These snapshots or Frames, are the basis for the predictability of Tier 1. For instance, a GEO satellite will see a LEO satellite come into view at a given reoccurring time. This LEO satellite will remain in view for just under half its orbit, and then break contact. This period of time that it is in view from first contact to break would be a frame with respect to that interface between GEO and LEO. Each type of interface from Figure (links figure) would have a different Frame period based on position and orbit of nodes.

The goal of this layer is to provide the nodes with an expectation at any given point in time of what other nodes should be within range and position to establish a link based solely on the orbital dynamics. From the basic orbital data, a Satellite Network Operating Center (SNOC) would periodically update the satellite constellation information to the satellites as required. This basic database is the Predetermined Satellite Database (PSD) that would reside on each node. From this, each node will send a Hello message to other nodes as they are expected to come into view as per the PSD. This Hello message will contain the Position (Pos), Period or orbit (Per), Time that further Hello message should be sent for that interface (Th/i) and a time stamp (Ts). Anytime a node receives a Hello message it will reply with an Acknowledgement (ACK) message containing Time received (Tr), Ts and Th/i for that node.

When a node receives a Hello message it stores the received information and uses it to update the PSD database, thereby correcting the satellite node positions due to any orbital effects and then computes a new routing table. When an ACK message is received by the original sender, it assigns a Trust Level (TL) value to the node it receives the message from. For each ACK message received this value is increased by one until a value of five is reached in which case further Hello messages are not required and the proactive algorithm backs off. The TL levels range from 0-5, with zero meaning no familiarity and five, full trust. If at anytime a Hello message is sent, but no ACK message received, then the TL goes back to zero. As well, when a Hello message is sent over an interface that has a TL of zero or one, then the Hello message will also include the TL, Pos and Ts for all one-hop neighbors. It should be noted that this discovery process is based on the interfaces in the topology, such that each node has an interface for all neighbors.

The use of the Trust Levels serves two purposes. One, as knowledge and predictability of an interface increases, the need for the discovery overhead is reduced. If an ACK message is not received then, the discovery process is re-initiated and the neighbor information can be used to help discover the topology. Two, the Trust Levels will be used in Tier 3 for further decision-making.

This Proactive portion of the HRPS protocol is a continual network discovery process that occurs, regardless of whether a message will be sent. It seeks to leverage the predictability of orbital mechanics in order to reduce the amount of overhead for network discovery.

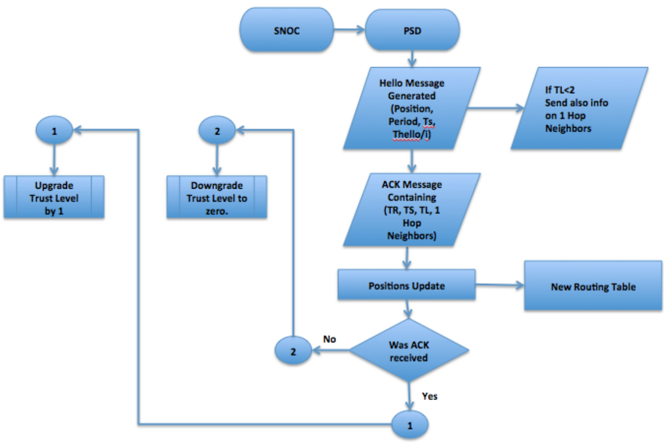

Fig 4. Proactive Layer

C. *Tier 2: Reactive*

The Reactive layer is the second tier in the routing decision process. In most reactive MANET protocols, the nodes flood the network with *route request* packets in order to establish links with neighbors before traffic can be sent. Instead of sending route requests for all possible nodes in the network, the second tier sends requests to all expected neighbors based on information passed from the first layer. In other words, in each frame, the reactive layer will search for neighbors based on expected positions and establish links with those it can receive. Because of variability in the space environment, satellite condition and other factors, a predicted link may not actually exist. Route request responses will be noted with link metrics for signal strength, data rate and delay.

The Reactive layer begins its process when a message is to be sent. The first step is to use the routing table from Tier 1 to choose the three shortest routes that could be utilized to transmit the message to its destination. At this point the sending node will send a Link State Request (LSREQ) to the destination across these paths. In response to this request, either the destination node or one of its neighbors who has the appropriate information will send a Link State Acknowledgement (LSACK) message with the following information for the interface paths: Bit Error Rate, Latency, Available Bandwidth and TL. This data is stored on the space routers in the form of a Management Information Base (MIB) and will be further discussed in a follow-on section. With the link-state information in hand, the algorithm will determine the best route for the packets to be forwarded.

Due to varying conditions within the space environment and changes in traffic loads across the network, the Reactive part of the protocol allows more advanced routing decisions. By utilizing the link-state metrics, HRPS becomes an adaptive protocol based on changing network dynamics in its route decision process. This could even extend to altering routes to take better advantage of weak ties or adjusting radio power or bandwidth as required.

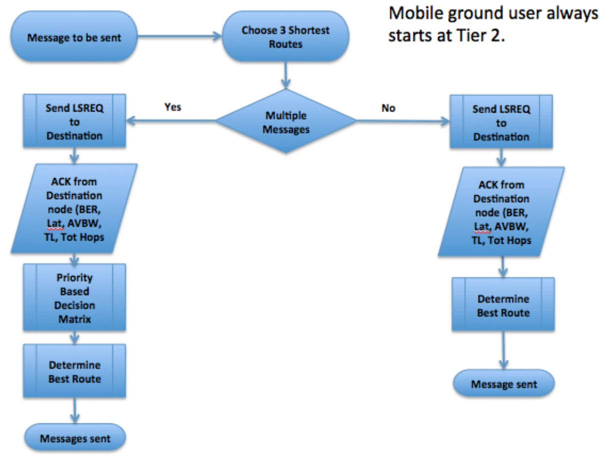

Fig 5. Reactive Layer

D. *Tier 3: Priority Based Decision Matrix*

The final layer in HRPS is the Priority Based Decision Matrix. This is a decision matrix that has been predefined in order to pass traffic over the most appropriate path. Because this protocol is designed for military use it will be described in that vein, but can easily be altered for commercial use. When multiple messages are to be sent at the same time, the third tier will be used to route the traffic based on a predetermined value system. This is done in order to use the links most efficiently based on priority of the message. The decision matrix will take into account the priority of the message and then choose the link based on the link metrics discovered in tier two and TL from tier one. For instance a high priority message for a Casualty Evacuation would be assigned the most robust link and route, as compared to a request for a routine weather update. The decisions matrix can be setup to choose priority based on user, message type, or other criteria. This element will operate at the Application layer as compared to the other tiers that operate at the Network layer.

E. *Space MIB*

In order to fully realize the potential of HRPS, a network management model must be introduced in order to allow remote monitoring and configuration of nodes on the network. This is commonplace in most networks today. For the proposed network configuration, the Simple Network Management Protocol (SNMP) is well suited for this requirement. The reason for this is that SNMP uses a Management Information Base (MIB) to store and exchange management information. There are a wide range of MIB types and sizes based on the requirements of the network and management model as defined in Internet Engineering Task Force (IETF) Request For Comment (RFC) documents. [12]

MIBs are stored as a virtual information tree on each node. The tree is broken into branching subgroups with managed objects within each subgroup. These pieces of data are collected and stored by the router and can be accessed by a NOC, other node, or an application. In the case of the space network, the MIB can be update to include the variables required for the HRPS protocol messages. Items such as the position, period and others can be pulled for the GPS unit, telemetry feed, and other sources and stored in the MIB. This allows network managers to access and update the MIB variables through SNMP as required to maintain the network as well. Other items not required by the HRPS protocol can also be included in the Space MIB in order to provide better diagnostics for the network health. The following table shows recommended variable for Space MIB.

MIB Variable
Node Position, Period, Th/I, Ts, Tr, TL
Satellite Battery State of Charge
Temperature of Transmitter
Single Event Latchups
Single Event Upsets

Table 5. Space MIB

F. HRPS Summary

The HRPS protocol with Space MIB was designed in order to take advantage of the uniquely predictable nature of a space based mobile ad-hoc network. As noted, the size, distance and extreme mobility of this environment is not well suited to the way traditional MANET protocols are designed. By combining Proactive and Reactive elements, this protocol seeks to maximize performance and reliability in this environment. The process of using Weak Tie links as they are available increases the strength of the network and allows for a large number of routes available as the topology changes as opposed to only utilizing fixed pipes in more traditional space networks. In the Reactive Tier, aspects of adaptation borrowed from social networking theory allow the protocol to receive feedback and adjust routing decisions to fully utilize the available links. The metrics for HPRS protocol merit include a unique combination of the delays in topology discovery, the completeness in discovery of weak and strong ties, proactive delay estimate, and throughput estimate associated with the weak and strong ties route composition.

VI. Conclusion

Theories used to explain how information flows in social networks can be used to improve network efficiency in satellite networks. The use of Weak Tie and Adaptation phenomena provide a method to tailor future protocols in order to best utilize the unique topology involving space-based networks. As well, the predictability of orbital mechanics further enhances such protocols by predicting future node positions and decreasing overhead in the network. Based on these theories, a Hybrid Routing Protocol for Space is presented. The next step will be to test this model in laboratory and real world experiments for verification and validation.

References

VII. Works Cited

[1] Paul K. Davis and James Bigelow, "Experiments in Multiresolution Modeling," 1998.

[2] John H. Holland, "Outline for a Logical Theory of Adaptive Systems," *Journal of the ACM*, pp. 297-314, 1962.

[3] W. Ross Ashby, "Principles of the Self-Organizing System," *University of Illinois Symposium*, 1962.

[4] Arkalgud Ramaprasad, "On the Definition of Feedback," *Behavioral Science*, vol. 28, pp. 4-13, 1983.

[5] W. Ross Ashby and Roger C. Conant, "Every Good Regulator of a System Must be a Model of that System," *International Journal of Systems Science*, vol. 1, no. 2, pp. 89-97, 1970.

[6] Balaji S. Chakravarthy, "Adaptation: A Promising Metaphor for Strategic Management," *The Academy of Management Review*, vol. 7, no. 1, pp. 35-44, January 1982.

[7] Mark S. Granovetter, "The Strength of Weak Ties," *American Journal of Sociology*, vol. 78, no. 6, pp. 1360-1380, 1973.

[8] Stanley Milgram, "The Small World Problem," *Psychology Today*, pp. 61-67, May 1967.

[9] James D. Montgomery, "Job Search and Network Composition: Implications of the Strength of Weak Ties Hypothosis," *American Sociological Review*, vol. 57, no. 5, pp. 586-596, 1992.

[10] T Hossman, T Spyropouls, and F Legendre, "Know Thy Neighbor: Towards Optimal Mapping of Contacts to Social Graphs for DTN Routing," in *INFOCOM, 2010 Proceedings IEEE*, Switzerland, 2010, pp. 1-9, 14-19.

[11] Li Qinghua, Zhu Sencun, and Cao Guohong, "Routing in Socialy Selfish Delay Tolerant Networks," in *INFOCOM*, 2010, pp. 1-9.

[12] Mani Subramanian, *Network Management: Principles and Practice*. Reading, MA: Addison Wesley Longman, Inc, 2000.

[13] J. J. Hopfield, "Neural Networks and Physical Systems with Emergent Collective Computational Abilities," *Proceedings of the National Academy of Sciences*, vol. 79, pp. 2554-2558, April 1982.

[14] Teuvo Kohonen, "The Self-Organizing Map," *Proceedings of the IEEE*, vol. 78, no. 9, Sep 1990.

[15] Lloyd Wood et al., "Adopting Internet Standards for Orbital Use," in *19th Annual AIAA/USU Conference on Small Satellites*, 2006.

[16] Espen Skjernvold, Trude Hafsoe, Frank T. Johnsen, and Ketil Lund, "Delay and Disruption Tolerant Web Services for Heterogeneous Networks," *IEEE*, p. Paper ID# 900190, 2009.

[17] Fan Xiaoguang, Xu Kuang, and Yang Guang-hua, "The

Effect of Communication Patterns on Opertunistic Mobile Networks," in *Consumer Communications and Networking Conference*, 2011, pp. 1016-1020.

[18] T. Hossmann, T. Spyropoulos, and F. Legendre, "Know Thy Neighbor: Towards Optimal Mapping of Contacts to Social Graphs for DTN Routing," in *INFOCOM Proceedings*, 2010, pp. 1-9.

[19] Linton C. Freeman, "Centrality in Social Networks Cenceptual Clarification," *Social Networks*, vol. 1, no. 3, pp. 215-239, 1979.

[20] Hui Pan, Jon Crowcroft, and Yoneki Eiko, "BUBBLE Rap: Social-Baed Forwarding in Delay Tolerant Networks," *IEEE Transactions on Mobile Computing*, pp. 1576-1589, 2011.

[21] Qualnet, "Qualnet Programmer's Guide," Los Angeles, 2011.

[22] Qualnet, "Qualnet 5.2 Wireless Model Library," Los Angeles, 2011.

[23] Qualnet, "QualNet 5.2 Model Library Index ," Los Angeles, 2011.

[24] Qualnet, "QualNet 5.2 Developer Model Library ," Los Angeles, 2011.

[25] Daniel Lang, "A comprehensive overview about selected Ad Hoc Networking Routing Protocols ," Department of Computer Science, Munchen, 2003.

[26] Aleksandr Huhtonen, "Comparing AODV and OLSR Routing Protocols," Telecommunication Software and Multimedia Library, Helsinki Unitversity of Technology, Aalto, 2004.

[27] Jan-Erik Ekberg, "Ad-Hoc Routing Protocols: The STAR routing protocol,".

[28] Tamara Dean, *Network+ Guide to Networks*. Boston, MA: Course Technology, Cengage Learning, 2010.

[29] Douglas Comer, *Compter Networks and Internets*. West Lafayette: Pearson Prentice Hall, 2004.

[30] "A Route Recovery Scheme in Dymo For Efficient Ad-oc," Department of Electrical Engineering, 2011.

[31] [Online]. http://www.software-engineer-training.com/category/higher-layer-network-protocols/

[32] Stanley Wasserman, "Social Network Analysis in the Social and Behabioral Sciences," *Social Network Analysis: Methods and Applications*, pp. 1-27, 1994.