Network Architecture and Network Stack

Roger Birkeland

June 2018

Note

This is a part of my thesis, titled *On the Use of Micro Satellites as Communication Nodes in an Arctic Sensor Network*, please do not distribute (yet).

1 Intro

This text introduces different concepts for space mission architectures, with different space segments, that may fulfill the needs of the mission requirements identified in Chapter **??**. The chosen space mission architecture must be integrated with the network architecture presented in Section **??** and the paper [1]. One space mission concept is chosen and topics for the network design are further discussed.

Through this chapter, topics on the network stack and network protocols will be discussed. A selection of possible network protocols is made. This network stack is the basis of simulations presented in the papers [2, 3] and in Chapter **??**.

While comparing satellite links with terrestrial links, satellite links have some very different properties compared to short-range radio links and cabled links, which have an impact on the network protocols. For example, the propagation delay due to increased distance is large, especially noticeable for GEO-satellites, which will have a propagation delay of at least 240 ms. For LEO-satellites the propagation delay will be shorter, but as the satellites move relative to the user, dynamic effects as Doppler shift, shadowing by obstructions, varying channels due to varying distance and propagation properties must be handled. Link-loss when the satellite is below the horizon, or handover between satellites, must also be addressed. This means that many well-known and commonly used protocols developed for more static channels may not be suitable, or must be used with caution.

Several published papers and standards discuss protocol stacks for satellites links [4, 5, 6]. The chosen network architecture must suit the particularities of the communication architecture when it comes to traffic and data type, number and dynamics of nodes, number of satellites, available link capacity, and Quality of Service (QoS).

Many traditional methods for media access are designed for data traffic properties as tight real-time demands[1] and continuous connections and data streams. In a remote sensor network the traffic need not have a strict real-time demand, compared to voice applications where the end-to-end delay cannot exceed a few hundred ms [5]. Depending on the nature of the sensor nodes, the sensors may not utilize the satellite link for a very long period of time; the traffic will be *bursty* by nature.

As for the number of nodes, this might be dynamic, but slowly changing. The communication system will be designed to fit one (or a few) end-users and their equipment. It will therefore be possible to register most types of sensor nodes in the network, at least nodes that require a return link. Small sensors like RF tags for tracking of animals or assets might be exempt from this general rule, should they be supported in the network at all.

The different users in the satellite network may have different traffic requirements. Some users might want to only transmit small data messages a few times a day, other users might wish to transfer larger amounts of scientific data for each and every pass. However, as the expected dynamics of the nodes are low (i.e. few and slowly moving nodes) prediction of the traffic should be possible.

The choice of network layer protocols is important, as this also has implications for lower layers. Use of the Internet Protocol (IP) (v4 or v6) can be desired in order to ensure interoperability of both terminals and dataflows to and from other networks.

A satellite network has its peculiarities, not present in the cabled types of networks. One example is the use of TCP on-top of IP. This topic has been addressed several times, for example in RFC2488 [7]. One issue is the long delay present in satellite networks, a second is that TCP will not see the difference if a packet is lost due to network congestion, or if a packet is lost due to erroneous transmission of a link-layer frame. TCP will in both cases reduce the data rate throughout the system [5, 8]. For an architecture with no real-time traffic, TCP will not be used, however, this case serves as an example of normal network protocols that not necessarily can be used directly when including a satellite component.

Throughout this chapter, the following definitions are useful:

Definition 1 *A structured chunk of data on the link-layer is called a* ***frame****.*

Definition 2 *A structured chunk of data on the network-layer is called a* ***packet****.*

Definition 3 *A structured chunk of data on the application-layer is called a* ***message****.*

2 Operational Considerations - Data Delivery and Data Flows

In general, two types of data must be transmitted between the satellite and the sensor nodes. One type of traffic is the user-data and/or commands corresponding to higher-level operations of the system. The second type of traffic is necessary signaling data that

[1]For example: Voice, video conferences, live remote control of equipment

must be allowed, for example, to allocate channel resources and coordinate data between sensor nodes and the satellite.

Definition 4 *Auxiliary traffic needed to allocate or change resources (frequency, modulation, time-slots and more) is called* ***signaling****. Also, "non-user data"-traffic necessary for higher-level operation.*

Figure 1 shows a simplified view of the dataflow in an operational context. Sensor data (-·-) will be collected by the sensor nodes (SN), stored in memory awaiting a satellite pass and then transmitted to the satellite (NWN). The satellite will store the data in its memory before transmission to the ground station (GW) during a pass. This system should act as a kind of delay-tolerant network. From the ground station, the data for various end users will be distributed over the Internet or other relevant networks according to agreements and concurrent system integration.

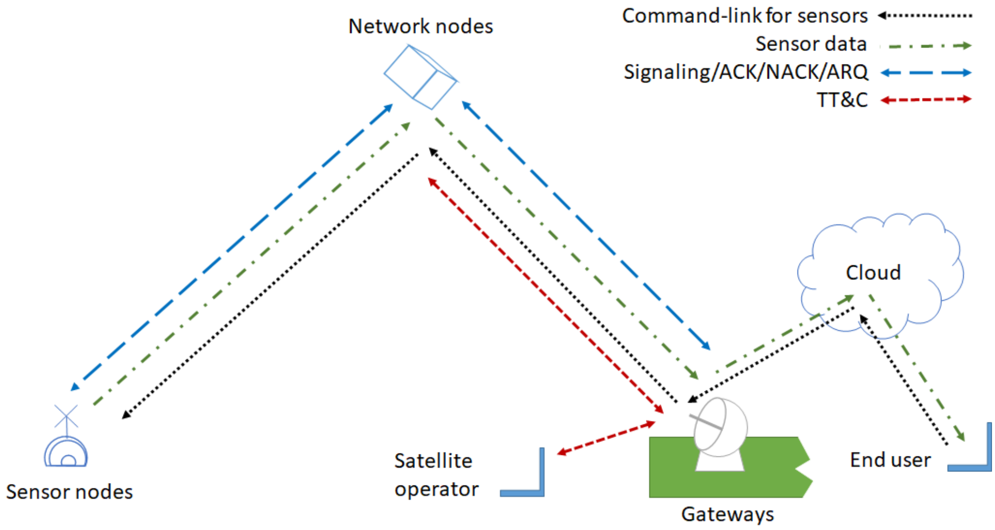

Figure 1: Model of the dataflow in the system. Sensor data flows from the sensor nodes to the end users (-·-). Control-messages flow from the end user to the nodes (···). These messages are relayed through the satellite. Each link has its own signaling channel (– – –) that handles message verification and so on. Also, the satellite operator uses the TT&C-link (- - -) to operate the satellite.

In addition to the sensor data, commands to the sensor nodes (···) will be transmitted from the SC&C, through the gateway and the satellite. This data will be delivered to the SC&C, represented by the end-user entity in the figure. Resource allocation and other signaling such as message verification, ACKnowlegde (ACK)/Negative ACKnowlegde (NACK)/Automatic Repeat reQuest (ARQ) will be resolved on a link-to-link basis (– – –). The satellite payload will therefore de-code all messages before they are stored and then forwarded on the downlink radio system. TT&C may be allocated to the satellite bus radio system (- - -), signaling, sensor node data and sensor node command links may be allocated to the payload radios. Also see Section ??.

In order to best exploit the limited bandwidth resources, sensor nodes must compress and code the data for transmission in the best way. It is a trade-off between (the need for) data processing in the sensor nodes vs. end-user data processing. Processing will reduce/condense the amount of data and hence save bandwidth capacity. With the availability of efficient but low-power processors today, as much processing as possible should be performed by the sensor node in order to reduce the data volume to transmit. However, original data should be re-constructable, or be available on request, in case the on-board processing malfunctions.

The degree of autonomy in the network is also open for discussion. The network architecture should serve a number of various sensor systems and end-users. Therefore, a fully operable communication system should be autonomous with respect to when and how aggregated data from sensor nodes is delivered to the end users. It should however be possible to pre-program the scheduling and sequencing if special needs arise, such as alarm events. A simpler scenario is considered in this thesis, where a request-response behavior is evaluated, see Section ??.

3 Satellite Mission Concepts

In Section ??, we presented the architecture chosen for this thesis work, but several types of satellite concepts may be chosen to complete this architecture. The following will present a few of the commonly chosen options.

All of these various solutions may fulfill the requirements for the scientific mission. In order to solve the communication challenges for a scientific mission in the Arctic, we can mention two candidates of satellite mission concepts. Either we propose a system based on small satellites in LEO orbit with more specialized communication payloads (concepts I & II), or we can rely on the proposed mega constellations from, for example, Starlink from SpaceX, and OneWeb[2] (concept III). For these two classes of concepts (dedicated mission or an architecture based on mega constellations), the space segment will differ significantly. Also, potential throughput, requirements and constraints for the ground systems may be different. Furthermore, the architecture based on small satellites can make use of a 3^{rd} party communication system such as VDES (concept II), or propose a new mission-dedicated system (concept I). Finally, the space segment may consist of one satellite or a swarm or constellation of satellites.

3.1 Concept I: Based on Dedicated Small Satellites

This concept is the main focus of the work presented in this thesis, and it stems from the research questions in Chapter ??. The satellite hardware market has matured over the years, and the cost of launching small satellites will be reduced as several organizations

[2]None of these are operating at the time of writing, but as they are supposed to be operating within a couple of years, they can be interesting candidates for an Arctic sensor program.

try to overcome the hurdle of efficient, timely and cheaper access to orbit. With this in mind, we should open up for very *mission centric* satellites, where a dedicated space segment can be a part of virtually any relevant project. The main part of this mission will be one or more small satellites, launched in one or more orbital planes providing coverage to the sensor nodes in question. Since this is a new design starting from scratch, most of the mission elements can be evaluated in a trade study.

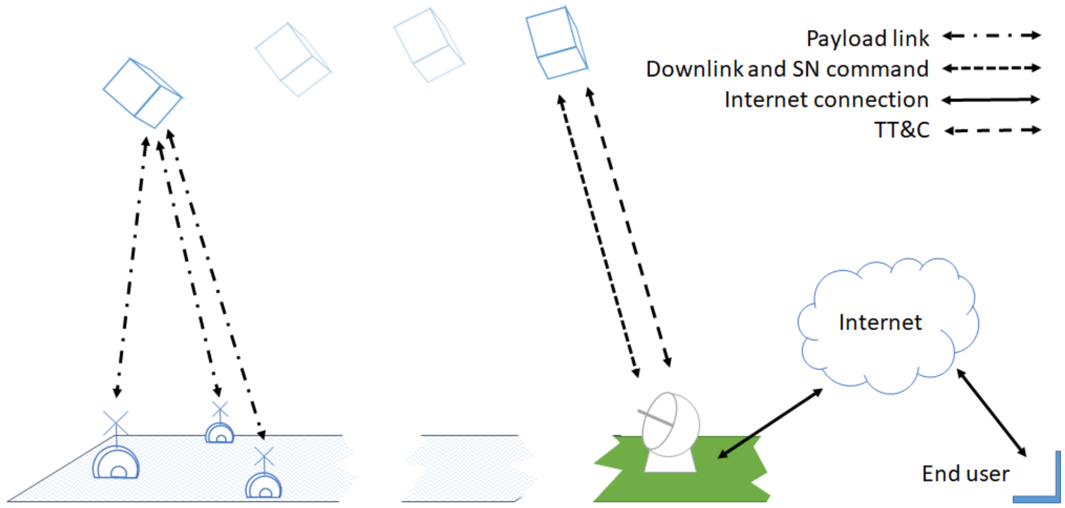

Figure 2: Concept I and II; small satellite UHF or VHF links (the use of these frequencies will be justified in Section ??).

Figure 2 shows conceptually how this can be set up. The TT&C-link is for telemetry and command between the ground station and the satellite. The UHF or VHF-links between satellite and sensor nodes as well as the downlink and sensor node commands will be used both for sensor data, command and housekeeping to and from the sensor nodes as well as network management for the communication system. All sensor nodes may have a connection directly to the space segment. The satellite(s) will store-and-forward the received data until the satellite(s) are within coverage of a ground station.

3.2 Concept II: Based on VDES

The VHF Data Exchange System (VDES) communication system [9] is a newly proposed service mainly for ship-ship and ship-shore communication. It will extend the Automatic Identification System (AIS)-service and be an integral part of e-navigation for ships. In addition to maritime service, this system could also provide the space segment infrastructure needed to support a sensor system in the Arctic. Since our satellite mission will be a secondary user of this infrastructure, the payload and satellites are also defined by others. Therefore, fewer of the mission elements are subject to possible tradeoffs in this concept. However, many of the generic studies performed here, for example, relating to different orbits, constellations and swarms are relevant for VDES. In some way, this concept is a particular implementation of Concept I. This is shown by Figure 2. The main differences between concepts I and II will be the implementation of an RF-link between the sensor nodes and the space segment, in addition to how data distribution and access is implemented.

3.3 Concept III: Based on Mega Constellations

In the past few years, the concept of *mega constellations* has gained quite some interest. Huge industrial actors as SpaceX, Facebook, Virgin and OneWeb are competing or teaming up to be the first provider of planet-wide internet connection through a plethora of satellites. This may be a game-changer of many dimensions, also for scientists wanting to get access to their scientific data from the Arctic. However, few technical details are known about any of these systems, and incorporating this space segment in an architecture proposal might be premature. On the other hand, the concept of these mega constellations and the game-changing service they may provide, is highly interesting and should be given some consideration.

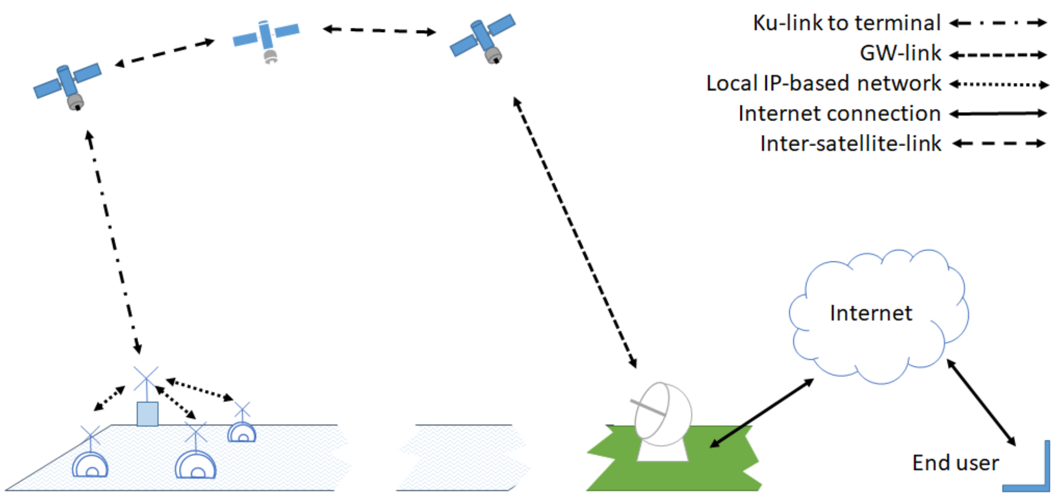

Figure 3: Concept III, based on mega constellations. All links will be defined by the mega constellation service.

For this concept, hardly any of the mission elements in Table **??** are subject to any real tradeoffs. The *subject* will however be subject for a tradeoff: For concept I and II, it is envisioned that each individual sensor should be able to connect directly to the space segment. For concept III, this might not be possible. OneWeb, for example, is said to only deliver infrastructure for cellular network providers, and access to the network must be granted through a roaming agreement. Also, the mega constellations will use higher frequencies and require a larger ground terminal that must act as a gateway base station for the sensor nodes. This means that this might be a viable solution for larger sensor nodes operating in clusters, rather than for individual sensors far away from each other, radio tags or various UVs. Such a solution is depicted in Figure 3.

3.4 Evaluation of the Mission Concepts

Table 1 shows the mission architecture elements and if these elements are subject to trade or not. These elements were introduced in Section **??**, and follow the definitions from [10, Chapter 2.2].

It is important to note that decisions on one element may enforce a decision on one or more other elements. For example, the decision on whether or not it is very important for an entity to operate and maintain control over its own system, may rule out the use of existing ground communication architecture. The decision will then lock or limit how other elements can be traded.

Table 1: Mission elements that may be traded.

Element of mission arch.	**Tradeable?**	**How and why**
Mission concept	Yes - limited	Should be open to a variety of sensor systems and space segments
Subject	Yes	Fixed/drifting sensor nodes, UVs
Payload	Yes - limited	Will depend on the system type/mission concept
Spacecraft bus	Yes - limited	Possible for a dedicated system
Launch system	Cost only	Choose minimum cost for near-polar orbit
Obit	Yes	Varying number of satellites and planes
Ground system	Yes	Several options depending on chosen system type
Comms. arch	Yes	Depending on system type, physical size and energy requirement
Mission ops	Yes	Level of autonomy and delay/timing in data delivery and data harvesting can be traded. Importance of own control over schedule and ops.

In Table 2, some of the mission elements are broken down into more detailed parts. For example, the *communication architecture* in Table 1 is split into several elements to visualize trade-offs for *frequency*, *coverage*, as well as *data delivery and access*. The three architecture options shown in Table 2, should all fulfill the main aspects of a scientific data recoding program. The point of this step in the design process is to be open and agnostic with respect to the space segment in the first design phase.

It is important to show that in order to bring forth the full communication infrastructure for use in the Arctic, several variants of the space segments can be utilized. Common for all three architectures is that they all are new; they do not exist at the time of writing. Also, architecture I and II may have a lot of commonality.

The different system types all have their potential strengths and weaknesses that can be used to evaluate them in the trade study. In addition to the space component, technology for the sensor nodes must be considered, including inter-communication, energy constraints, and throughput. The choice of architecture option will also impact these topics. In this thesis, the system type II and III serve as examples for comparison with the dedicated small satellite architecture, type I, that has been chosen as the architecture we focus on.

Table 2: Satellite architecture options

System type	Opt.	I: Custom	II: VDES	III: Mega const.
Orbit	A1	Single satellite	Single satellite	
	A2	Swarm	Swarm	
	A3	Constellation	Constellation	Constellation
Frequency	B1	EESS (UHF)		
	B2		VDES (VHF)	
	B3			Ku
Comm. arch (nodes)	C1	Direct to node	Direct to node	
	C2	Via GW-node	Via GW-node	Via GW-node
Ground system	D1	Dedicated		
	D2	Commercial		Commercial
	D3		VDES-infrastruct.	
Coverage	E1	Store and fwd.	Store and fwd.	
	E2	Near continuous	Near continuous	
	E3			Continuous
Data delivery and access	F1	Internet		Internet
	F2		VDES-infrastruct.	
Programmatic	G1	Full control		
	G2		Secondary user	
	G3			Commercial

From Table 2 a large number of options may be identified, but the true number of options might be both lower and higher. Lower, since some of the options are interlinked, such as that near continuous coverage will only be possible for systems I and II if a swarm or constellation of satellites is deployed. The number of options might also increase if other elements such as the make of the sensor nodes is included. A selection of a few options are made, and then their utility will be analyzed with the sensor nodes in mind (see Chapter **??**).

Selected Options As shown, only one combination of options exists for system III (A3-B3-C2-D2-E3-F1-G3), resulting in one possible mission architecture. For system type I (and II) more sets of options, leading to several possible mission architectures, are available. Some options are exclusively selected for one of the systems. All orbit options (A1, A2, A3) will be considered:
[single satellite, swarm, constellation]. In this thesis, the use of a UHF Earth-Exploration Satellite Service (EESS)-band (B1) will be considered, however VDES (B2) might be an alternative. It will be assumed that all sensor nodes are able to *connect directly* to the

space segment (C1), but this decision is in reality up to the owner of the sensors. Furthermore, the use of a *commercial ground system* (D2) is assumed, with (D3) as the alternative for VDES. Also, the communication payload will be of the *store-and-forward* type (E1) in either case, even if the forward delay can be close to zero.

In total, this gives three main options - depending on the orbit - with variations on the payload (system type I or II). In total, six mission architecture options (since options in B and D are linked) are available:
A{1,2,3}-B{1,2}-C1-D{2,3}-E1-F1-G1.

For the detailed study the three options A{1,2,3}-B1-C1-D2-E1-F1-G1 are selected.

4 Heterogeneous Networks

The network architecture presented in Section **??** is a heterogeneous network resembling the architecture type I from above. Unmanned Vehicle (UV)s and other vehicles as ships are also considered as Network nodes in this network. The space component is based on satellites in a polar orbit. In the following sections the network stack of such networks will be presented. The emphasis is on ease of interoperability between different nodes in the networks as well as interoperability between various networks. The network stack will be presented with that as the primary driver, however the specific *implementation* of the network stack and protocols is given less consideration.

5 The Network Stack

The design and functions of the network stack are important choices for the communication system. Figure 4 shows a simple layout of a IoT-compatible network stack model, corresponding to the Open Systems Interconnection - OSI model [11].

The definition and implementation of a network stack is a complex topic. The use of existing protocols (for example, as used for IoT) potentially gives easy interoperability with excising systems, but has to be traded against custom implementation that might be more efficient and better to use for a narrowband satellite link. Discussions and work presented here may also be found in the papers [1, 2, 3]. There it is argued that all sensor nodes should be reachable through the Internet, and therefore common internet protocols should be used. The use of a network stack as described below also means that the satellites must be fully regenerative; as all nodes in the network must be able to handle the messages and interact with neighboring nodes.

5.1 Cross Layer Design

Traditionally, one layer only interacts with the layers above (if any) and below (if any) through a defined set of interface functions. However, in order to best adapt the total system yield to changing conditions, and changing user needs, some degree of *cross layer* design will be needed [4, 5]. This means that one layer interacts with a layer more than one place above or below in the network stack.

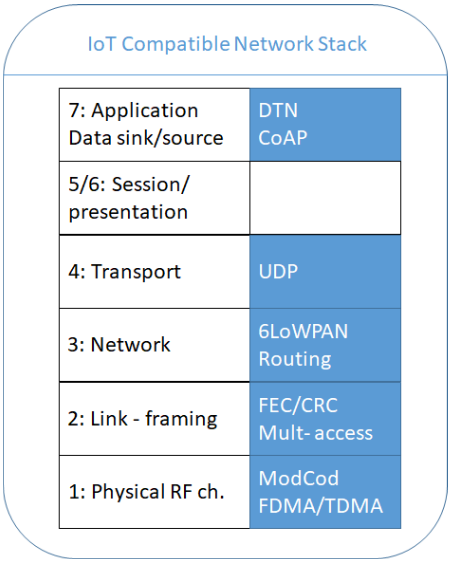

Figure 4: Simple layered network stack model with example functions and protocols, derived from the commonly known OSI model.

One simple example is if we want the Physical Network Layer (PHY) to adapt to changing conditions of the radio channel, using methods for Adaptive Coding and Modulation (ACM). There can be several reasons for wanting this: One is that the link budget is often designed in a way so the link will function also when the link conditions are worse than normal, in order for the link not to drop (see Chapter **??** on Link Budgets). During normal operations, this leads to under-estimation of the link margin and poor exploitation of the link, hence valuable throughput is wasted. This means that the link should be capable of adapting to changing conditions. Also, a node might want to transfer a type of data that must be error-free, or sometimes the node wants to transfer data where (small) losses are acceptable. In all cases, this calls for interaction between the PHY, Media Access Control (MAC) and application layers. If link quality is good, the PHY might instruct the application-layer to generate more data. In the opposite case, the PHY might want to instruct the application to generate less data, in order to prevent buffer overflows and losses due to that. When transitioning from data with a reliability requirement to data without this requirement, the application layer may instruct MAC to ease requirements for ACK/NACK/ARQ, which potentially could waste both time and channel resources.

A further argument for employing a cross layer design, or even remove some of the traditional layers, is that each layer in the network stack may add additional overhead to the system. One could then argue that, for example, avoiding network and transport layers (see Figure 4) and further more combine layers 5 to 7, it will be possible to remove the overhead introduced by network protocols. However, this may lead to a higher implementation cost in the higher layers; as the application must take care of everything; from network, routing and keeping track of delivered and non-delivered messages. The gain of relying on standard implementations for network protocols is also lost. The increased overhead from standard protocol implementations must be weighed against the higher implementation cost. The overhead from standard protocols is further discussed in Section **??** and in [3].

5.2 Layer 1 - Physical Layer

Layer one is often denoted Physical Network Layer (PHY), and is in our case the radio link. Its properties are the selected frequency, modulation, bandwidth, power levels and so on. Layer two is the link or Media Access Control (MAC) layer. Here, the data frames transmitted over the radio channel will be created. Coding/de-coding, modulation and de-modulation takes place here. Layer 3 & 4 are responsible for addressing, routing and reliable transport of the data packets to the end user. Higher layers present, in the defined format, data to the end user.

Some of the properties of the physical layer, such as power levels, modulation(s), frequency and allowed frequency bandwidth, are covered in Chapters **??** and **??**. Adaptive links are discussed in Section **??**.

Other Layer 1 related topics include how to divide the physical channel resources among users, by means of multiple access methods such as Time Division Multiple Access (TDMA), Frequency Division Multiple Access (FDMA), Code Division Multiple Access (CDMA) or use of spot beams for one or several users. The mentioned access methods can be combined in different manners that vary with system, application and time. These methods are Layer 1 properties, but in some way they can be said to implement functionality (granting user access to the channel), normally thought of as a Layer 2 task.

5.3 Layer 2 - Link - Medium Access Control-Layer

This layer has two main tasks: Number one is granting users access to the radio resources. The second is ensuring that the traffic on the radio channel is transmitted from sink-to-source error free in the most efficient manner. The total data throughput capacity in a small satellite such as a CubeSat is fairly limited. Also the pass duration is relatively short, on the order of 10 minutes, so efficient use of the radio resources is important. Therefore, the use of link-layer (or channel) coding is common.

The traffic from the nodes can be said to be *bursty*, meaning the traffic from one node only occurs for a short period of time during a day, or a period of coverage. This implies that a fixed allocation of channel resources, either by time-slots or different frequency channels may lead to waste of resources, as they are not used for most of the time.

5.3.1 Medium Access and Random Access Methods

The sensor nodes may need to transmit various forms of messages. Both longer messages containing scientific data and short status messages (telemetry) will be needed. Both types of messages can in addition be *regular* or *random*, with regards to when and how often the node wishes to transmit a message. The mode of operation for the node might also change during the period of operation, as a function of time of day or year, time passed, events occurring and so on.

Based upon how sensor data is generated, two modes are defined:

Definition 5 *A node producing larger amounts of data on a regular basis can operate in* ***assigned mode****.*

Definition 6 *A node producing smaller amounts of data on non-regular (random) basis can operate in* ***random mode****.*

Random Mode In order to accommodate the nodes operating in *random mode*, as well as new nodes, the system must allow for Random Access (RA). However, since much of the traffic in this kind of system can be predicted, not all traffic should to be initiated over an RA-channel, as this will waste time during setup and channel assignment.

A well-known RA-method is ALOHA. This method allows the nodes to speak when they have data to transmit. It is shown that a method as slotted ALOHA will only yield a channel efficiency of around 36% [12], meaning valuable channel resources are wasted. Several articles discuss this topic, as well as other suitable random access methods [6, 13, 14, 15].

Assigned Mode The nodes operating in *assigned mode*, could have fixed and predetermined access to channel resources, either by assigning them a fixed frequency channel (Frequency Division Multiple Access (FDMA)), a fixed time-slot (Time Division Multiple Access (TDMA)) or a combination of these (separate FDMA-channels which in turn are divided by time slots). If all nodes use their assigned resources all the time, the channel can be utilized efficiently and congestion free. However, if the nodes are not using their assigned resources, these are wasted. Therefore, this way of allocating channel resources is best when the load is split in a defined way between the nodes.

Another way of designing a (near) congestion-free system is to base the traffic management on the satellite polling the nodes. This will be like a client/server-based request-response system, where the node can be viewed as the server (data source), and the satellite as the client (data user). This way, we will not allow for any random access to the channel. The consequence is that this will impede deployment of new nodes, as new nodes should be able to register to the network.

A middle ground is to split the time between a signaling, or random access channel, and strict allocation of the channel resources. A given fraction of the channel is set aside for random access or other signaling. The duration of the random channel frame must be decided so it is long enough for the random access period to be usable, and short enough for the random access period to repeat itself a sufficient number of times during a pass over a sensor node.

During a pass, the satellite will use its acquired knowledge about the sensor nodes, as well as requests received during the signaling-period, to allocate resources to the sensor nodes. This can be done in two ways: The first method will be to use fixed time-slot allocations similar to a dynamic TDMA system, and then grant the sensor nodes access to a given number of time-slots. If the data transfer is obstructed for some reason, and the transfer of data is not complete within the allocated number of frames, the node must wait until the next pass.

In the second method, the satellite will poll each sensor node, and instruct it to send a fixed amount of data. The sensor node then has access to the channel until the data is transferred, until a timeout is reached or until the satellite has moved out of sight. A timeout mechanism is necessary if an error should occur during data transfer, in order to not block other nodes during the whole pass. When data from one node is transferred to the satellite, the satellite will issue a polling request to the next sensor node.

At the end of each pass between the satellite and the sensor node, the next contact should be agreed, in order to minimize the load on the random-access channel. This way, only new nodes, or new and un-expected events have to be announced through the random-access channel.

5.3.2 Link-Layer Coding (FEC)

The radio channel will never be able to transfer data error-free. Noise will impede the signal, and bit errors will be introduced. These errors must be detected and corrected for. Data produced by the sensor nodes should reach its destination unchanged and eventually error free. In the following we will use the term "error-free" for the data obtained after error correction, meaning they are "error-free" to the extent of meeting the data quality requirements of the end user.

A decision must be made on where, and how, consistent data transmission checks should be performed. In order to ensure integrity and consistency of data, the link layer should at least employ an integrity check of received frames by using Cyclic Redundancy Check (CRC). In addition, to increase the probability of receiving an error-free frame (or to save transmitted power), additional coding can be added, commonly called Forward Error-correction Code (FEC). These codes work by adding redundant bits to a link-layer-frame, in order to enable recovery of some erroneous bits in the received frame, and then decode the whole frame correctly. The strength of the code depends upon how much redundancy is added. This indicates that it exists a trade-off between receiving a frame error-free and the usable bit rate, as the information rate is reduced due to introduction of redundant data. In combination with FEC, different varieties of ARQ can be used in order to request any missing frames due to link outages or FEC failure. In the ideal case, the link layer should provide near loss-free communication. This in order to ease the requirements for higher order protocols, such as IP. This also implies that all nodes in the network, including the satellite, must be fully re-generative. All messages must be de-coded and checked in order to fulfill necessary error handling within acceptable time limits.

Discussions on how the coding influences the link budget are presented in Chapter **??** and in Section **??**. Coding will increase the link margin, but it also reduces the end user goodput.

5.3.3 Confirmation of Message Delivery

The implementation of message integrity checks will result in the need for ACKnowlegde (ACK)/Negative ACKnowlegde (NACK) messages. Should this be implemented on the link-layer or on the network/transport layer or on the application layer? Should it be an end-to-end check or link-to-link (node-to-node) check? If network layer frames (See Definition 1) are evaluated on end-to-end transmission over a multihop network (for example, sensor node to satellite to gateway to end-user), an error introduced on any of those links will cause an erroneous transmission. Due to the nature of the network architecture and coverage; end-to-end transfers may take a long time; sometimes several hours. Then, the NACK must also be transmitted back the same or a similar route. A re-transmission must be issued and the process will start over again. This leads to potentially very long delays (minimum three end-to-end transmissions) as well as the need for buffering of old data on the sensor nodes.

A more efficient procedure is to evaluate the traffic on a link-to-link basis, meaning that a receiver node demodulates and decodes the received packet from its neighbors in the multi-hop-link. This calls for a "smarter" receiver in the satellite, but the end-to-end transmission times will be significantly reduced. The transmitted message can be deleted from buffers when reception in the next network node is confirmed. Since the architecture must have delay tolerant properties; meaning that any node on the route must take custody of the data regardless if it is decoded or not; link-to-link based FEC and ARQ resolving will not increase the need for buffers in intermittent nodes significantly. Decoding of data is common in highly integrated radio systems for small satellites, and radios based on Software Defined Radio (SDR) will intrinsically decode any received message. This is further discussed in Section **??**.

Short Messages The use of link-layer codes will be efficient if data from nodes is on the form of short independent packets. In this case, each and every packet must be transmitted and received error-free.

Long Messages Different strategies to code larger chunks of data exists. Which strategy to use also depends on the radio channel. If the noise can be modeled as Additive White Gaussian Noise (AWGN), link-layer coding can be sufficient. However, if the link experiences various fading phenomena that causes loss of many bits in individual frames, then, for example, several IP-packets may be interleaved. Then, bit-errors will be split over multiple IP-packets, with the possibility of reconstructing them.

Also, in general, a data network may also lose packets due to other errors than bit errors. The network can experience congestion and full buffers, causing packet drops. If a NWN experience a re-start or power-off while it holds data, this data may be lost. Various strategies may be enforced to mitigate this. This encompasses both various coding schemes, but it also includes how to logically handle such situations. Detailed discussion and analysis of these topics are outside the scope of this work.

5.4 Layers 3 & 4 - Network and Transport Layers

Layer 3 and 4 take care of the routing and further transport of the data to its end-point, possibly through several jumps and several different physical channels. All nodes in the network must be able to take care of their routing function so they can select which packets to send forward on which channel.

By using common implementations of well-known protocols, for example, IoT-protocols such as 6LoWPAN, sensor nodes and satellites can interact easily with other agents in the connected network(s). Which protocols to use must be selected carefully. New IoT protocols are designed for networks with low data rates and where the link availability is not constant. These implement functionality that helps the already stated short-comings of, for example, TCP [5, 8, 7]. Proposed protocols, implications and implementations are discussed further in the papers [2, 3].

5.4.1 Transmission Planning, Orbit Propagators and Smart Routing

In order to minimize the energy consumption at the node, the contact times between the sensor node and the satellite must be coordinated. This coordination, including the smart routing, can be implemented in several ways. Two examples follow: One option is to let the node have its own orbit propagator, and receive Two-Line Elements (TLE)s [16] once in a while from all satellites[3]. Based on the TLEs and the node's geographic position, the propagator will calculate the next satellite pass, taking into account the total message propagation time to the gateway (the smart routing), and the radio can be woken up in due time.

A second option is to let the gateway or the satellite perform the calculation of the best transmit times. The satellite will then broadcast the time of the next passes during the current pass. This can be a broadcast message to all sensor nodes, transmitted at regular intervals. For the satellite, this requires that the satellite itself knows its own TLE and therefore its place in the orbit as well as the node positions, so this time can be adjusted as the satellite moves over the ground.

Another version of the last option can be to let the node and the satellite directly negotiate when the next contact will be. However, this will require much more radio traffic and will be a more complex process compared to a broadcast message. The pitfall with the broadcast option is that if the sensor node loses contact with the satellite for one or two passes where it should have had contact, it can get completely out of sync.

In that case, the node should have its own orbit propagator anyway, or it must keep its receiver on until it successfully receives a new link availability message from the satellite. It is assumed that the advanced sensor nodes will have sufficient computational power to successfully run the orbit propagator. There are readily available open source propagator libraries for Python, so this can be run on any microprocessor running Linux.

Smart Routing In order to best plan in which order the nodes should be assigned which resources and when, the satellite should have knowledge about the geographical position

[3]TLEs are usually valid one to two weeks at the time.

of each and every sensor node. It can be shown that when the network has members that are not seen by all satellite revolutions, considerable reduction of the end-to-end delay may be gained if the node employs a *smart routing* method for selecting which satellite to route its data through [3]. In short, this method is forward-looking and aims to calculate which of the satellites in a swarm or constellation a sensor node or gateway should connect to, in order to minimize the delay between transmission and reception of a message. This is done so that the sensor node (or gateway) simulates the next satellite passes, form its own viewpoint and also at the target node. Based on when these passes will take place, the source node calculates which satellite can relay a message to the target quickest. This, in contrast with the *naive* routing, which will send its message to the first satellite available.

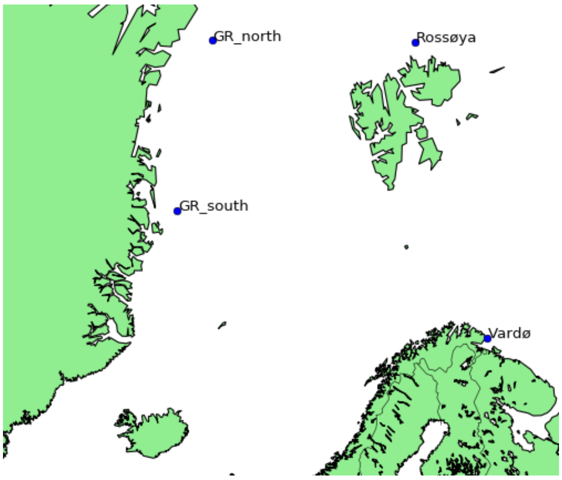

Figure 5: Node placements

For a case with nodes deployed as shown in Figure 5 with the gateway located in Vardø, it can be shown that employing the smart routing method will lead to improvement in 15% of all message transfers. The average improvements in delay will be up to 93 minutes, which is significant and correspond to making contact between the sensor node and the gateway one orbit earlier. Table 3 shows a sum-up for the overall performance. In the case studied, the message to be transmitted was 512 B, and the links must last for at least 30 seconds to be considered usable.

In Table 3, the results between the *naive* routing (picking the first possible satellite) and the *smart* routing are compared. The line *First Satellite* shows the average number of seconds the transmitting node must wait in order to send its message to a satellite. Note that the table is only showing the cases where the *smart* and *naive* methods differ, so the total average wait time is much lower. The line *Start to Finish* shows the time it takes for the message transfer to complete.

Two things can be seen from this: First, by using the *smart* routing, the wait-time until the correct satellite is available is naturally longer. This is because the *naive* method always picks the first satellite available. However, it is seen that the time it takes until the message is successfully received, is lower when relying on the *smart* routing. The average improvement corresponds to 71 minutes in the case shown here.

Table 3: Performance Comparison – Overall

	Naive	**Smart**
First Satellite (s)	4031	4427
Start to Finish (s)	5853	5401
Same Choice (%)	–	85.2
Improvement (s)	427	4281
Total Losses (%)	2.8	3.3

The network emulation (the method is described in Section **??**) implements a real networking stack. Therefore, unexpected behaviors due to congestion or delays led to the *naive* approach being better for some message requests. These occurrences correspond to less than 3% when both routing approaches selected a different satellite.

5.5 Layers 5, 6 & 7 - Higher Layers

Even if the details of the data sinks and sources are beyond scope of this thesis work, it is important to maintain some knowledge about the implications enforced by these layers. Also, the requirement to have a delay-tolerant, error-free, data delivery leads to constraints on how lower layers should be designed and implemented. Delay-Tolerant Network (DTN)-like protocols, also CoAP, are running on the application level (level 7), so the network functionality itself encompasses all layers in the network stack.

6 Dynamic Network Management

The main events that could call for network adaptivity, are mainly extreme link conditions such as solar storms, antenna icing or antenna drowning. In those cases, the risk of link loss is imminent and the link will either be present or not. Another likely case is related to the throughput requirement from a node. This can change with regards to the type of data recorded, as detection of events can trigger more data to be aggregated or that the importance of the data changes due to an event. This could call for the need of dynamic capacity allocation and/or dynamic scheduling.

If the data to be transmitted is not urgent, the sensor node could request a slot large enough to fit its data and then the satellite (or a central network manager) could allocate the capacity. This could lead to the situation where the node will not be allowed to transmit its data during the current pass, it might have to wait until the next pass. If the data is urgent, other nodes may have to yield their capacity.

In situations where there are several routes to reach the network (or if it is possible to make capacity requests during one pass and do the actual data transmission the following pass), one could envision a central network controller (not in the satellite) handling the routing within the network. This controller could utilize all network resources, ground/air-based and satellite. The controller's decision could then be transmitted to the network, either through the satellite or through other paths. This could lead to less service and signaling traffic over the links and therefore more capacity for user data. The network, and the links therein, will only change when needed and otherwise be static.

7 Network Design Proposal

The discussion above leads to two distinct design choices: Alternative 1 is to implement a network stack as discussed above, implementing support for IoT-like IP networks based upon 6LoWPAN. Alternative 2 is to implement a very simple stack, with no real network functionality, giving the need for network gateways to translate between the satellite links and the outside world. With option 2, network overhead over the satellite links can most likely be reduced a great deal; restricting the overhead to only a simple custom link-layer addressing scheme and other required header fields. Then, for example, overhead from 6LoWPAN and Constrained Application Protocol (CoAP) as well as network service traffic will be eliminated.

All network nodes or sensor nodes interacting with other networks must then have a gateway router to translate between the networks. Depending how important a seamless operation between multiple heterogeneous networks is, this added implementation cost – and reduced overhead gain – will be a subject to trade. The layer 3 smart-routing functionality could still be implemented. Also, standard network protocols are well proven and they have intrinsic security and authentication measures. This loss of functionality must then also be traded against the increased overhead.

In the remainder of this study, it is assumed the use of an IoT network stack, due to the functionality gain this implementation gives. Section **??** gives more details on estimation of expected overhead and throughput from this implementation.

Selected network stack

Layer 1 A few FDMA ch.s + dynamically allocated "slow" TDMA
Layer 2 Suitable frame format. Network emulations use 802.15.4
Layer 3 6LoWPAN
Layer 4 User Datagram Protocol (UDP)
Layer 7 CoAP

7.1 Hand-Over Between Satellites

In this study, a link to a satellite is considered active if a sensor node or gateway sees the satellite. When the satellite is beyond line of sight, the link does not exist. This also implies that there is no direct hand-over to the next satellite; even if that satellite might be visible before the first satellite goes below the horizon; as can be the case for some configurations of a freely drifting swarm. The links are only handled on individual basis, and the DTN functionality will take care of transporting the message to its final destination.

7.2 Identification and Security

Cybersecurity is an increasingly important topic today, but details of this topic is considered beyond the scope of this work. For an Arctic sensor network, it is expected that all network nodes must be uniquely identified, and access must be restricted to authenticated users. For an IoT-based implementation it is expected to make use of inherit functionality.

7.3 Transfer of Data to Earth Station

The satellite(s) in the network will have a store-and-forward payload that collects data from sensor nodes, aggregates this data and then downloads it to the gateway station for distribution to the end user. If the sensor-node and the gateway are both within sight from the satellite, the delay time for this forward might be on the order of milliseconds, if the satellite simply forwards the data to the gateway immediately.

References

[1] Roger Birkeland, David Palma, and Artur Zolich. Integrated smallsats and unmanned vehicles for networking in remote locations. In *Proceedings of The 68th International Astronautical Congress*, 2017.

[2] Roger Birkeland and David Palma. Freely-drifting small-satellite swarms for sensor networks in the arctic. In *Proceedings of the Third International Congress on Information and Communication Technology*, 2018.

[3] David Palma and Roger Birkeland. Enabling the Internet of Arctic Things with Freely-Drifting Small-Satellite Swarms. *IEEE GlobeComm (submitted)*, 2018.

[4] Giovanni Giambene and Sastri Kota. Cross-layer protocol optimization for satellite communications networks: a survey. *International Journal of Satellite Communications and Networking*, 24(5):323–341, 2006. ISSN 1542-0981. doi: 10.1002/sat.853. URL http://dx.doi.org/10.1002/sat.853.

[5] Paolo Barsocchi et al. Radio resource management across multiple protocol layers in satellite networks: a tutorial overview. *International Journal of Satellite Communications and Networking*, 23(5):265–305, 2005. ISSN 1542-0981. doi: 10.1002/sat.820. URL http://dx.doi.org/10.1002/sat.820.

[6] Riccardo De Gaudenzi, Oscar Del Rio Herrero, Gennaro Gallinaro, Stefano Cioni, and Pantelis-Daniel Arapoglou. Random access schemes for satellite networks, from vsat to m2m: a survey. *International Journal of Satellite Communications and Networking*, pages n/a–n/a, 2016. ISSN 1542-0981. doi: 10.1002/sat.1204. URL http://dx.doi.org/10.1002/sat.1204. sat.1204.

[7] M. Allman, D. Glover, and L. Sanchez. RFC 2488 - Enhancing TCP Over Satellite Channels using Standard Mechanisms, January 1999. URL https://tools.ietf.org/html/rfc2488.

[8] N. Kuhn, E. Lochin, J. Lacan, R. Boreli, and L. Clarac. On the impact of link layer retransmission schemes on tcp over 4g satellite links. *International Journal of Satellite Communications and Networking*, 33(1):19–42, 2015. ISSN 1542-0981. doi: 10.1002/sat.1064. URL http://dx.doi.org/10.1002/sat.1064.

[9] *Recommendation ITU-R M.2092-0 - Technical characteristics for a VHF data exchange system in the VHF maritime mobile band.* ITU, 2015. URL https://www.itu.int/rec/R-REC-M.2092.

[10] James R. Wertz and Wiley J. Larson, editor. *Space Mission Analysis and Design*. Microcosm Press and Springer, third edition, 1999.

[11] International Organization for Standardization. ISO/IEC 7498-1:1994 Information technology - Open Systems Interconnection – Basic Reference Model: The Basic Model, November 1994. URL `https://www.ecma-international.org/activities/Communications/TG11/s020269e.pdf`.

[12] A. Meloni and M. Murroni. Random access in dvb-rcs2: Design and dynamic control for congestion avoidance. *IEEE Transactions on Broadcasting*, 60(1):16–28, March 2014. ISSN 0018-9316. doi: 10.1109/TBC.2013.2293920.

[13] O. Del Rio Herrero and R. De Gaudenzi. High efficiency satellite multiple access scheme for machine-to-machine communications. *IEEE Transactions on Aerospace and Electronic Systems*, 48(4):2961–2989, October 2012. ISSN 0018-9251. doi: 10.1109/TAES.2012.6324672.

[14] H. C. Bui, J. Lacan, and M. L. Boucheret. An enhanced multiple random access scheme for satellite communications. In *Wireless Telecommunications Symposium 2012*, pages 1–6, April 2012. doi: 10.1109/WTS.2012.6266094.

[15] C. Kissling and A. M. Dlr. On the integration of random access and dama channels for the return link of satellite networks. In *2013 IEEE International Conference on Communications (ICC)*, pages 4282–4287, June 2013. doi: 10.1109/ICC.2013.6655237.

[16] Dr. T.S. Kelso. Frequently asked questions: Two-line element set format, 1998. Available at https://celestrak.com/columns/v04n03/.

Acronyms

6LoWPAN IPv6 over Low power Wireless Personal Area Networks. 15, 18

ACK ACKnowlegde. 3, 10, 14

ACM Adaptive Coding and Modulation. 10

AIS Automatic Identification System. 5

ARQ Automatic Repeat reQuest. 3, 10, 13, 14

AWGN Additive White Gaussian Noise. 14

CDMA Code Division Multiple Access. 11

CoAP Constrained Application Protocol. 18

CRC Cyclic Redundancy Check. 13

DTN Delay-Tolerant Network. 17, 18

EESS Earth-Exploration Satellite Service. 8

FDMA Frequency Division Multiple Access. 11, 12, 18

FEC Forward Error-correction Code. 13, 14

GEO Geostationary Orbit. 1

GW Gateway. 3

IoT Internet-of-Things. 9, 15, 18, 19

IP Internet Protocol. 2, 13, 14, 18

LEO Low Earth Orbit. 1, 4

MAC Media Access Control. 10, 11

NACK Negative ACKnowlegde. 3, 10, 14

NWN Network node. 3, 9, 14

OSI Open Systems Interconnection. 9

PHY Physical Network Layer. 10, 11

QoS Quality of Service. 1

RA Random Access. 12

SC&C Science Command and Control. 3

SDR Software Defined Radio. 14

SN Sensor node. 3

TCP Transmission Control Protocol. 2, 15

TDMA Time Division Multiple Access. 11, 12, 18

TLE Two-Line Elements. 15

UDP User Datagram Protocol. 18

UV Unmanned Vehicle. 7, 9

VDES VHF Data Exchange System. 4, 5, 8, 9

List of Definitions

ALOHA A method for granting random access to a physical channel. A user will send the data it has, when it has data to send. It will listen for data from other users, and if it receives data when it transmits, it will decide it was a collision and therefore re-send its data later. 12

assigned mode A node producing (larger) amounts of data on a regular basis can operate in assigned mode. 12

CDMA Use of spread spectrum in order to share the available RF bandwidth between users. Users can access the channel at the same time, using the full bandwidth. Will also uniquely identify messages meant for one recipient. 11

CubeSat A much-used form-factor for small satellites. Based on unit cubes of 10 cm^3, which equals 1 U. Typical sizes are 1, 2, 3, 6 and 12 U. The mass of 1 U can be 1.33 kg. 11

FDMA The RF resource is divided into a set of independent channels using a partition of the total available bandwidth allocated for the system. Users can operate simultaneously. 11

frame A structured chunk of data on the link-layer is called a frame. 2

goodput The goodput of a system is the rate of transmitted application level data. This is lower than the systems throughput, as goodput excludes any protocol overhead or re-transmissions. 13

Ku Ku-band (K-under) is a radio communication band that spans the 12 - 18 GHz frequency range (IEEE definition). 8

MAC Functions for how to divide the channel resource and grant access to individual users. Placed on Layer 2 of the OSI-model. 10, 11

message A structured chunk of data on the application-layer is called a message. 2

OSI model A much used model that describes the functions of the network stack of a communication system. 9, 10

packet A structured chunk of data on the network-layer is called a packet. 2

random mode A node producing (smaller) amounts of data on non-regular (random) basis can operate in random mode. 12

signaling Auxiliary traffic needed to allocate or change resources (frequency, modulation, TDMA-slots and more). Also, "non-user data"-traffic necessary for higher-level operation. 3, 4

TDMA The RF resource is divided into time-slots, where individual users are allocated a pre-defined set of one or more slots in which the user is granted access to the full bandwidth of the RF-channel. 11

UHF Ultra High Frequency. A radio communication band that spans 300 - 3000 MHz (ITU definition). Usually used to denote frequencies from 300 - 1000 MHz (IEEE definition). 5, 8

VHF Very High Frequency. A radio communication band that spans 30 - 300 MHz (ITU and IEEE definition). 5, 8