

# Chapter 1 – Introduction

The aim of this chapter is to provide a background to topics which are relevant to the subject of this thesis. The motivation for writing a thesis regarding the topic is also given, together with the goal I wanted to reach.

## 1.1 Background

### 1.1.1 Wireless networks

Telecommunication networks have taken the world by storm, and revolutionized the way we live and think. Most people have become so accustomed to using it, that it is considered a necessity in businesses and daily life.

Wireless networks specifically have grown considerably in terms of capability and implementation because of its connection flexibility [1] because of its connection flexibility. It is this characteristic that pushes technology and research even further in search of new ways to accommodate the ever-increasing needs of people and applications.

#### 1.1.1.1 Wireless ad hoc networks

A wireless ad hoc network [2] is a network consisting only of wireless nodes, meaning no other infrastructure is needed besides the nodes themselves. The entire network consists only of the nodes, and they are used to forward data from one user to another rather than using a fixed network infrastructure, as is mostly used in today's world. Depending on the routing protocol used, a path is created from the sender to the receiver node on demand, or using a previously established routing table to determine a route. Data is passed on from one node to another along the route to the final destination. Within the network different methods of data transmission can be used, including sending or casting data to a single (uni) destination, also known as unicasting, sending to multiple (multi) destinations, also known as multicasting, and finally broadcasting, which refers to sending to all possible destinations.

Wireless ad hoc networks and MANETs (Mobile ad hoc networks [3]) are increasing in popularity because the nodes themselves are all that is needed to create a network. This makes these networks

more flexible and reliable, and possibly cheaper as well. Previously these networks were limited to military applications but are now starting to see more public uses. [4,5].

## 1.1.2 Linear network coding

Network coding [6] is a technique used to increase the capacity of a network by combining messages sent over the network. The combined messages can be separated by using a sufficient number of original messages in conjunction with an algorithm that is designed in accordance to the algorithm used to combine the messages. From [7] the Min Cut, Max Flow theorem states that the maximum flow of information in a network is equal to the sum of the cut of the link capacities. In [8] it is shown that linear codes are sufficient to reach this multicast capacity, and therefore making more complicated nonlinear codes redundant. Linear network coding is easier to understand and implement, and can still make a network utilize its maximum capacity.

Linear network coding [9] uses an algorithm that combines messages so that the result is a linear combination of the original messages. Linear network coding can easily be implemented, using the exclusive OR operation to add messages without the result being larger than the constituting messages. The original messages can be extracted from the combined message if enough linearly independent equations of the original messages are known, as follows:

Two messages A and B combined using XOR to find a combined message C:  $A \oplus B = C$

Using original message A to decode the combined message C and find B:  $C \oplus A = B$

Network coding can be implemented in wired, as well as wireless networks. The broadcast nature of wireless networks make them ideal for network coding since messages sent from one node to another can often be overheard, and the overheard messages can then be used to decode combined messages[10,9].

The following example explains how network coding can be used to increase the throughput, using XOR calculations:

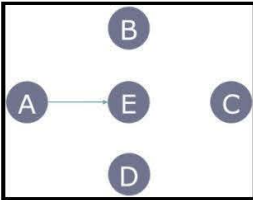
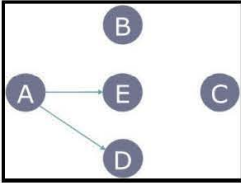
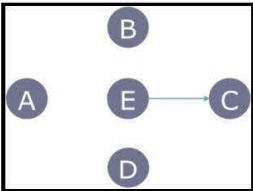
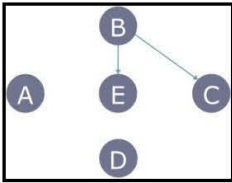
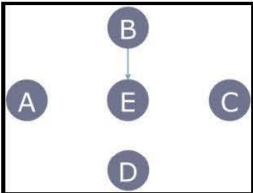
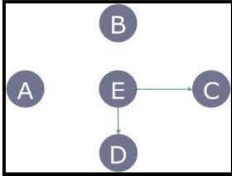
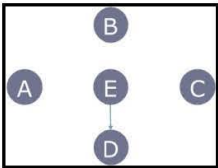
Traditional routing without network coding.	Routing with network coding
<p>Step 1: Node A sends a message to C via node E</p> 	<p>Step 1: Node A forwards a message intended for node C to nodes E and D</p> 
<p>Step 2: Node E forwards the message from A to C</p> 	<p>Step 2: Node B forwards a message intended for node D to nodes E and C</p> 
<p>Step 3: Node B sends a message to D via node E</p> 	<p>Step 3: Node E combines the messages from nodes A and B using a XOR calculation and forwards the same resulting message to nodes C and D. Each has already received a message from either node A or B. Since both these nodes now have sufficient linearly independent combinations of messages from A and B (with an uncombined message also counting as a linearly independent message), they can now find the two parts of the message from E. Each now has the original message intended for them.</p> 
<p>Step 4: E forwards the message from B to D</p> 	<p>Using network coding clearly uses one less time slot than the method not using it. Implementing network coding however does cause some extra processing power at the nodes. Less time is used to send the same amount of data over the network, thus the throughput in the network can increase if network coding is used.</p>

Figure 1-1: Explanation of network coding.

This can be formulated in a more formal form. A network is represented by a graph  $G = (V; E)$  in network coding theory [11]. The vertices  $V$  of  $G$  are the termination or end points in the network of the links. The edges  $E$  are the links between the terminals. Naturally  $E \subseteq V \times V$ . Each link  $e \in E$  then has the capacity  $C(e)$  which is a non-negative number, and a vertex at its start  $v$ , and one at its end  $v'$ . This is written as  $e = (v', v)$ .  $\Gamma_I(v)$  is defined as a set of edges that ends at a vertex, and  $\Gamma_O(v)$  is a set that originates at a vertex. This is the information necessary to define the topology of any network. Further one can define messages or symbols at each vertex that have to be sent over a relevant link, the relevant connection between the source and sink vertex, as well as the relevant rate that the message(s) can be sent at. Each data unit is represented by a symbol from a finite field  $F_{2^m}$ , with  $m$  some integer. The connection and rate are functions of the message(s) that have to be sent. As an example, a node  $v$  has symbols  $A_1, \dots, A_K$  in a buffer which it has previously received. Symbols  $A_1, \dots, A_K$  can also represent symbols that originated at node  $v$ , and were saved in the buffer after they have been sent. Node  $v$  now receives symbols  $B_1, \dots, B_L$  from incoming edges. With linear network coding the symbol  $C_e$  sent by  $v$  over some  $e$  will be a linear combination of the messages  $A_1, \dots, A_K$  and  $B_1, \dots, B_L$ :

$$C_e = \sum_{i=1}^K \alpha_i A_i + \sum_{j=1}^L \beta_j B_j \quad (1)$$

Where:  $\alpha_i, \beta_j \in F_{2^m}$  are the linear coefficients that are used to encode the data.

A possible issue with network coding is that a coded message can only be decoded if the decoder has received a sufficient number of the original consisting messages in the coded message. Therefore, networks with multiple transmission paths are necessary for network coding to be profitable, so that a receiver can receive different linearly independent coded combinations of the original messages, with the different combinations being a result of different transmission paths. Typically ad hoc or multi-hop networks are favourable for the implementation of network coding. The following are two methods to ensure that sufficient combinations of messages have been received to decode coded messages:

### 1.1.2.1.1 Random network coding

In a multi-hop network, each forwarding node makes a combination of messages it has received, and forwards it to the destination. The destination must then wait until it has received a sufficient number of linearly independent combinations of the original messages destined for it to decode the received combinations. For example:

$$\begin{aligned} \text{Received combinations:} & \quad (A \oplus B \oplus C) \quad (A \oplus C) \quad (B \oplus C) \\ \text{Decode:} & \quad (A \oplus B \oplus C) \oplus (B \oplus C) = A \\ \text{Decode:} & \quad (A \oplus C) \oplus A = C \\ \text{Decode:} & \quad (A \oplus B \oplus C) \oplus A \oplus C = B \end{aligned}$$

The problem a system, using random network coding has, is that it has to wait for enough information before it can decode everything that has been sent. The time it will take to receive enough messages is not constant; moreover' the time it takes to receive enough messages is not known before the recipient has finished receiving enough messages. Theoretically the receiver sometimes might not receive enough combinations at all. The probabilities for receiving enough information are presented in [12].

### 1.1.2.1.2 Deterministic network coding

Deterministic network coding predetermines how the coding will take place before the actual network coding is done. This means that each node in the system has to have knowledge of the entire network topology which will be involved in the data transmission, as well as which nodes in the network have received which data before it can code and send data through the network. This is an effective form of network coding in the sense that the data can be combined at the coding nodes by using a polynomial time algorithm [13]. By doing this, the decoder does not have to wait for a random period, but rather it will receive as few combinations possible whilst still making decoding possible.

The problem with this approach is all the nodes have to keep track of what the other nodes in the network have sent to be sure to make the right network coding combinations. This additional information that has to be sent, limits the expandability of a network using deterministic network coding. Update messages have to be sent by all the nodes in the network, to all other nodes in the network regarding the data they sent. The bigger a network is, the more update messages would have to be sent, and an expanding network would eventually come to a point where it has so many update messages being sent on it, that there would remain no bandwidth for other data.

#### 1.1.2.1.2.1.1 Logical network coding

Logical network coding [8] simply implies that the messages are coded by means of a logical or arithmetical formula. Mostly this form of network coding is referred to as linear network coding, or simply network coding. It is named logical network coding here to clarify the difference between it

and analogue network coding, which is theoretically also linear coding. If Equation (1) (presented in section 1.1.2) is applied to messages to provide a coded message, this would be a typical example of logical network coding.

#### 1.1.2.1.2.1.2 Analogue network coding

Analogue or physical layer network coding [14] differs from logical network coding in that, rather than combining two messages by using some linear or nonlinear logical function, it uses positive or negative interference in the electromagnetic transmission waves as a method of combining messages[15]. The following figure explains the gain of physical layer network coding over logical network coding in an Alice and Bob topology. Nodes A and C have to interchange data via node B, and either logical or analogue network coding can then be done at node B:

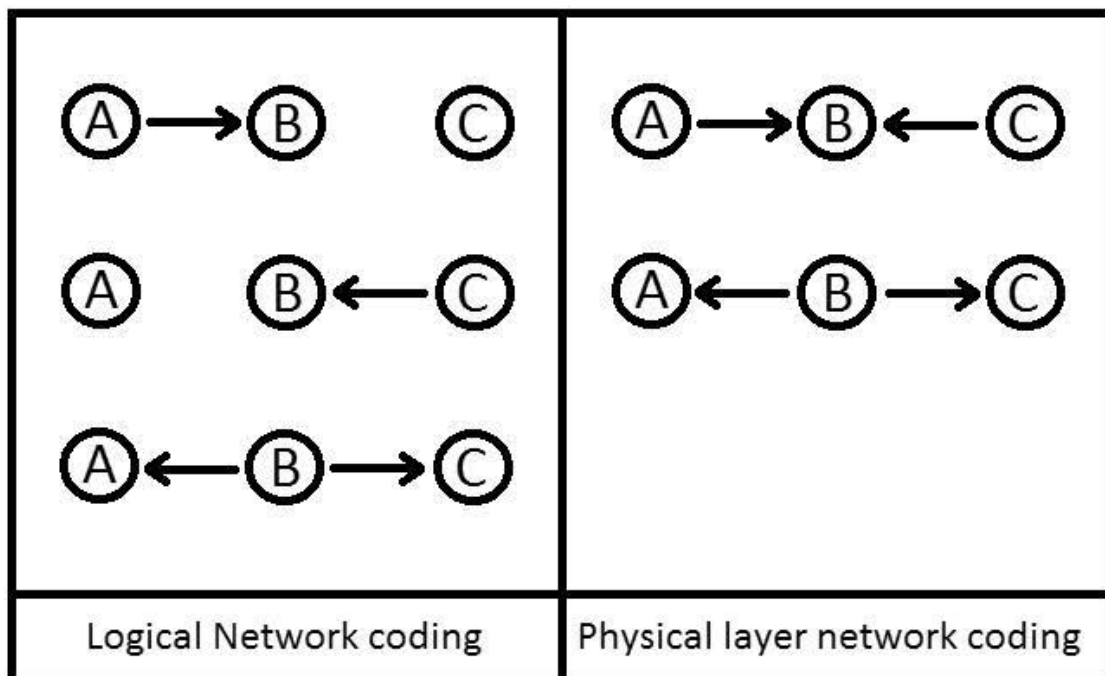


Figure 1-2: Logical and physical network coding compared

#### 1.1.2.1.2.2 Scheduling

The scheduling of a network establishes which node's turn it is to transmit at which time. Scheduling [16,17,18] is a major concern in all deterministic network coding systems because one has to let all relevant nodes know who sends what messages at which time intervals to perform optimal network coding. Scheduling with physical layer network coding [14] provides an even greater challenge, because the nodes which have to transmit, have to be synchronised to send at exactly the same time for the receiving node to receive the correct coded message. From the scheduling problem arose a need to be more opportunistic in applying network coding. These schemes [19,20] propose methods of opportunistic scheduling for implementing network coding, and do away with deterministic

network coding over the whole network. They rather keep track of only the data the surrounding nodes have, and use this knowledge to exploit network coding opportunities.

### **1.1.2.2 Network coding benefits**

The following are a few examples of how network coding can benefit a network:

#### **1.1.2.2.1 Security**

Because linear combinations of data, rather than the data itself, are being transmitted over the network, it can be considered a more secure network in some instances [9,21]. If an intruder in a network listens in long enough it would be able to gather enough data to perform decoding when random network coding is used. Using deterministic network coding is more secure because it plans the network coding so that the receiver, rather than the intruder, can decode efficiently. But there is still no guarantee that someone listening in, cannot obtain enough data to perform decoding, even if it is less likely [22].

#### **1.1.2.2.2 Throughput**

If one uses network coding, one can send the same amount of data over a shorter period of time than when traditional routing is used. Sending more data in the time won, results in an increased throughput in a network using network coding [8,9,23]. In today's world, networks are always in need of increased throughput and the fact that the use of network coding offers increased throughput in a network without necessarily changing the hardware, could lead to widespread usage of network coding.

#### **1.1.2.2.3 Robustness**

The robustness of a network describes the network's ability to continue functioning in the case of sudden failure of network components. Sometimes a forwarding node or link fails, and a network has to make provision for such failures. Most protocols must start looking for an alternative route or have some way of dealing with this. The problem remains that a session is still interrupted, and usually needs to be re-established, causing a delay in the transmission. When random network coding is used, no specific path is used, and a single link of failing node would not interrupt a session. This makes random network coding robust to sudden failures in the network [24,9,25,21].

#### 1.1.2.2.4 Adaptability

With some networks, parts of the network or the whole network is mobile. This requires constant new route establishment in the network. Random network coding does not use specific routes, but rather the whole network, to send data; therefore, it is not affected much when the connectivity in a network changes. This makes a mobile or changing network using random network coding much more adaptable [15,9], and the problem of constant route establishment is nullified.

#### 1.1.2.3 Network coding vulnerabilities

Since packets are mixed and propagated through a network, one faulty packet could infect the whole network, and cause many other packets to be decoded incorrectly and spread, leading to more errors [21]. This fact makes network coding particularly vulnerable to attacks from inside the network coding network. A single unintended error could also have the same effect, so error management is crucial.

### 1.1.3 TCP/IP and OSI models

When data is sent over a network, metadata is sent along to provide information regarding the message sent. This is necessary because the data itself does not necessarily provide enough information regarding its destination network, node or application. The metadata in computer networks can be likened to the address written on an envelope of a letter. Since a computer network is vastly more complex than a standard postage system, more metadata is required for different processes in the network. A layered approach is used in networks today to simplify the process of creating and understanding protocols [26,27,28]. The two main models used to describe how protocols interact, are the following:

<b><u>TCP/IP Model</u></b>	<b><u>OSI Model</u></b>
Application	Application
	Presentation
	Session
Transport	Transport
Internet	Network
Link	Data link
	Physical

**Figure 1-3: TCP/IP and OSI models**

The TCP model is used to describe the internet protocols. It has thus become a standard for most protocols. The OSI model describes the functions of a protocol in more detail and fits into the TCP/IP model as well. From the OSI model the functions of various protocols can be clearly seen, and when a network is created, each layer has to be considered and regarded separately.

### **1.1.4 Suitable wireless ad hoc topologies for network coding**

The following table contains a few typical experimental network topologies suitable for experimentation, to illustrate or observe the workings of network coding in a wireless ad hoc network. The black circles represent nodes, and the arrows represent the connectivity between the nodes.

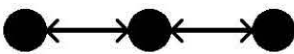
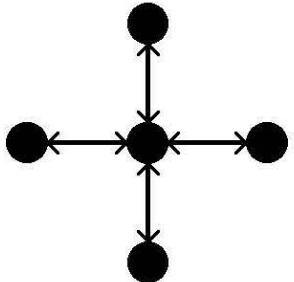
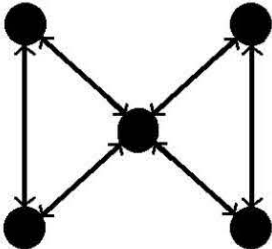
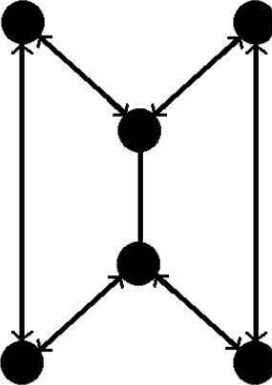
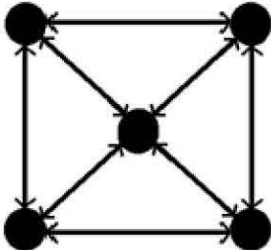
Topology	Description
Three node chain	
Cross	
Bow-tie	
Butterfly	
Five node <b>b</b> Box	

Figure 1-4: Different network topologies

These networks merely illustrate the inner workings of transactions in greater networks like MANETs or wireless ad hoc networks. These networks show what happens in a simplified form so that one can follow step by step how the transactions take place.

## 1.1.5 Different transmission types

When sending data over a network, one would often like to reach different people or groups of people. Unicasting, multicasting and broadcasting can be used to transmit the data. Each of the above-mentioned layers has special functions that make provision for each type of data transmission. The following figure explains more about these three concepts. The blue circles represent nodes, the blue square hubs, and the yellow blocks messages being sent and received:

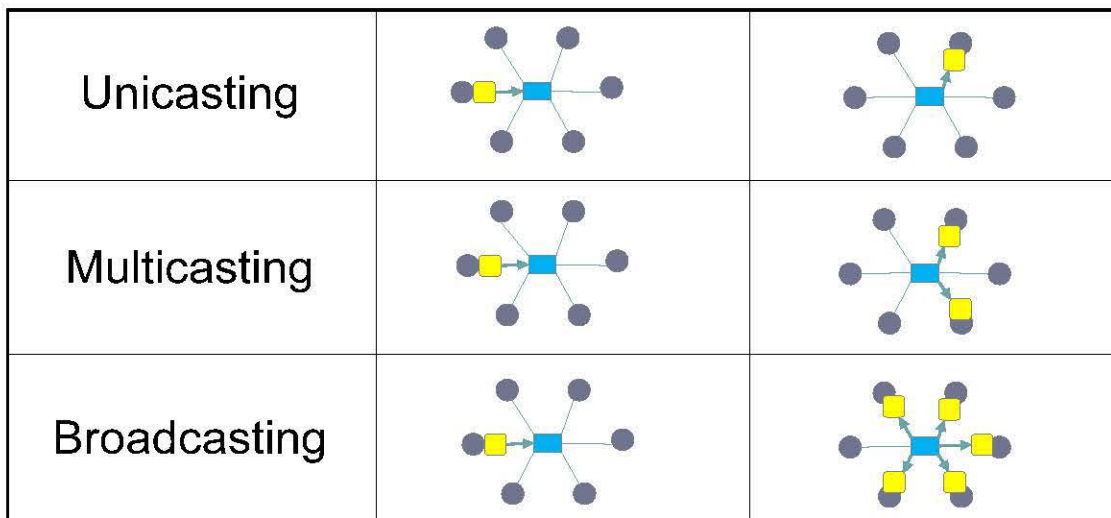


Figure 1-5: Different data transmission methods.

## 1.2 Motivation

The following two quotes from a paper [21] on the pros and cons of network coding done in 2008, reveal that there is still much work to be done in research on network coding, specifically regarding the implementation thereof in a layered network approach:

- “Although there is evidence of considerable gains in terms throughput and robustness, particularly in flooding and broadcast services for wireless networks, it is probably too early to say whether network coding will one day be as ubiquitous as routing protocols. However one thing is clear — there is still a lot of research work to be done.”
- “It is certainly reasonable to question whether or not the current layered architecture makes sense under the network coding paradigm and to ask if cross-layer methods lead to better performance trade-offs for instance in wireless networking scenarios. Although certainly pertinent, these questions are beyond the scope of this paper and shall be dealt with elsewhere.”

The work reported in this thesis is an attempt at addressing this problem and to come closer to the final goal of characterising and rating the implementation of network coding in different layers of the OSI stack.

## 1.3 Research goal

The goal of this thesis is to implement a wireless node model with network coding in the MAC layer and evaluate the performance characteristics of reference networks that implement the new node model. This will serve as the first step of a greater goal, namely finding the most favourable position in the OSI stack to implement network coding. To realize this greater goal, network coding must be implemented in all of the layers of the stack so that comparisons can be drawn between the characteristics of network coding in the different layers. Since this is a great research task, the research shown in this thesis will only serve as the first step of a greater whole.

One cannot draw decisive comparisons between previous network coding implementations since there are a variety of network parameters, which first have to be standardised, before a fair comparison can be made. The work done in chapter 3.2.3.1 hopes to create a standard network on which network coding can be implemented on different layers, and then be characterised and compared.

Previous implementations of network coding in the different layers of the OSI stack have also been studied (refer to section 2.2), and each layer's characteristics are described and discussed in the following chapter. Alternatives to implementing network coding in one of the OSI layers will also be looked at, together with other practical issues in a wireless network.

## 1.4 Conclusion

In this chapter, an introduction to wireless networks, network coding, common layered network models, suitable topologies for network coding is given, and basic transmission types are briefly explained. With this introduction we hope to provide a platform for the reader to fully understand the motivation for the work done in this thesis, as well as the goal of the research. The goal of this thesis is to characterise network coding which is implemented in the MAC layer of the OSI stack