Technische Universität München
Coding for Communications and Data Storage
Prof. Dr.-Ing. Antonia Wachter-Zeh

Master's Thesis

New ~~Achieved~~ sizes Gap For Generalized Combination Network

- New Gap sizes for the Gen. Comb. N.
- Vector Network Coding Gap sizes for the Gen. Comb. N.

Vorgelegt von:
Ha Nguyen

- other proposals?

München, 07 2019

Let's talk about it

Betreut von:
Sven Puchinger

2 Preliminaries

+ a bit text
+ references (eg textbooks)

2.1 Definition

~~**Definition 2.1** (Vector Space).~~ A vector space of dimension n over a finite field with q elements is denoted by $\mathbb{F}_q^n$.

← not a definition. This is just notation. Maybe write it outside the Def. environment

~~**Definition 2.2** (Gaussian coefficient).~~ Gaussian coefficient (also known as q-binomial), which counts the number of subspaces of dimension k in a vector space $\mathbb{F}_q^n$:

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = \prod_{i=0}^{k-1} \frac{q^n - q^i}{q^k - q^i}$$

Definition 2.3 (Grassmannian Code). A Grassmannian code is a set of all subspaces of dimension $k \leq n$ in $\mathbb{F}_q^n$, and is denoted by $\mathcal{G}_q(n,k)$. Due to being the set of all subspaces that have the same dimension k, it is also called a *constant dimension code*. The cardinality of $\mathcal{G}_q(n,k)$ is the Gaussian coefficient (also known as q-binomial), which counts the number of subspaces of dimension k in a vector space $\mathbb{F}_q^n$:

$$|\mathcal{G}_q(n,k)| = \begin{bmatrix} n \\ k \end{bmatrix}_q = \prod_{i=0}^{k-1} \frac{q^n - q^i}{q^k - q^i},$$

where $q^{(n-k)k} \leq \begin{bmatrix} n \\ k \end{bmatrix}_q \leq 4q^{(n-k)k}$.

this is already a statement
must be outside the def.

Definition 2.4 (Projective Space). The *projective space of order* n is a set of all subspaces of $\mathbb{F}_q^n$, and is denoted by $\mathcal{P}_q(n)$, i.e. a union of all dimension $k = 0, \ldots n$ subspaces in $\mathbb{F}_q^n$ or $\mathcal{P}_q(n) = \bigcup_{k=0}^{n} \mathcal{G}_q(n,k)$.

Definition 2.5 (Covering Grassmannian code). An $\alpha - (n,k,\delta)_q^c$ covering Grassmannian code (code in short) $\mathcal{C}$ is a subset of $\mathcal{G}_q(n,k)$ such that each subset of α codewords of $\mathcal{C}$ span a subspace whose dimension is at least $\delta + k$ in $\mathbb{F}_q^n$.

← Ref Etzion et al. 2019

Definition 2.6 (Subspace packing). A subspace packing $t - (n,k,\lambda)_q^m$ is a set $\mathcal{S}$ of k-subspaces or k-dimensional subspaces (called *blocks*), such that each t-subspace of $\mathbb{F}_q^n$ is contained in at most λ codewords of $\mathcal{C}$.

Be sure to add proper references everywhere. If you are not sure about some ref., please ask me!

Definition 2.7. $\mathcal{A}_q(n,k,t;\lambda)$ denotes the maximum size of a $t-(n,k,\lambda)_q^m$ code, where there are no repeated codewords.

~~**Definition 2.8** (Multigraph).~~ A graph is permitted to have multiple edges. Edges that are incident to same vertices can be in parallel.

~~**Definition 2.9** (Directed acyclic graph).~~ A finite directed graph with no directed cycles, i.e. it consists of a finite number vertices and edges, with each edge directed from a vertex to another, such that there is no loop from any vertex v with a sequence of directed edges back to the vertex again v.

Definition ~~2.10~~ ~~(Multicast~~). Multicast communication supports the distribution of a data packet to a group of users [ZNK12]. It can be one-to-many or many-to-many distribution [Har08]. In this study, we consider only one-to-many multicast network.

2.2 Theorem

Ref.!!!

Theorem 2.1. *If n, k, t, and λ are positive integers such that $1 \leq t < k < n$ and $1 \leq \lambda \leq \begin{bmatrix} n-t \\ k-t \end{bmatrix}_q$, then*

$$\mathcal{A}_q(n,k,t;\lambda) \leq \left\lfloor \lambda \frac{\begin{bmatrix} n \\ t \end{bmatrix}_q}{\begin{bmatrix} k \\ t \end{bmatrix}_q} \right\rfloor$$

Ref!!!

Theorem 2.2. *If n, k, t, and λ are positive integers such that $1 \leq t < k < n$ and $1 \leq \lambda \leq \begin{bmatrix} n-t \\ k-t \end{bmatrix}_q$, then*

$$\mathcal{A}_q(n,k,t;\lambda) \leq \left\lfloor \frac{q^n-1}{q^k-1} \mathcal{A}_q(n-1,k-1,t-1;\lambda) \right\rfloor$$

Please talk to me next week about properly using Def. environments

3 Network Coding

better write that Ahlswede et al. introduced the topic

3.1 What is Network Coding?

word rep.

The most general definition of Network Coding is defined in the seminal paper of Ahlswede *et al.* [ACLY00]. It refers to *coding* at a vertex in a network, where coding means an arbitrary combination of inputs for outputs. This combination is called a *network code*. Because this coding process does not only happen at the source but on any vertex in the network, the network codes are on *packets*. Packets can be messages of the sources or inputs of a vertex. Ahlswede *et al.* look at networks consisting of vertices interconnected by error-free point-to-point links, which is still applicable to present-day networks, e.g. error-free wireline networks. The study of "error-free links" in network coding distinguish itself from Channel Coding for noisy links.

uncertain. which output/input? You don't even say that there is an information transmission. From where to where do you transmit?

what is a wireline?

A particular form of network coding is *Random Linear Network Coding* - RLNC introduced in [HKM+03]. RLNC considers each vertex's outputs as a linear combination of its inputs, specified by independent and randomly chosen code coefficients from some finite field $\mathbb{F}_q$. These coefficients can be represented in a vector as extra information for an initial packets, which is able to be contained in packet headers of the present-day widely used network called *Packet Network*. Hence, in this study, we define *Network Coding* as coding at a vertex in a packet network, where data are divided into packets and network code is applied to the contents of packets. This concept is clearly explained in Section 4.2.1, Section 4.2.2, and Section 4.2.3, where we introduce scalar and vector network coding as a matrix channel.

~~say that packets~~ $\mathbb{F}_q$

I did not get this part. Please explain it better.

the Packet Network?

Please differentiate clearly between Random LNC and the network coding we consider here. Here, we can choose the coefficients

Maybe write in text instead of section

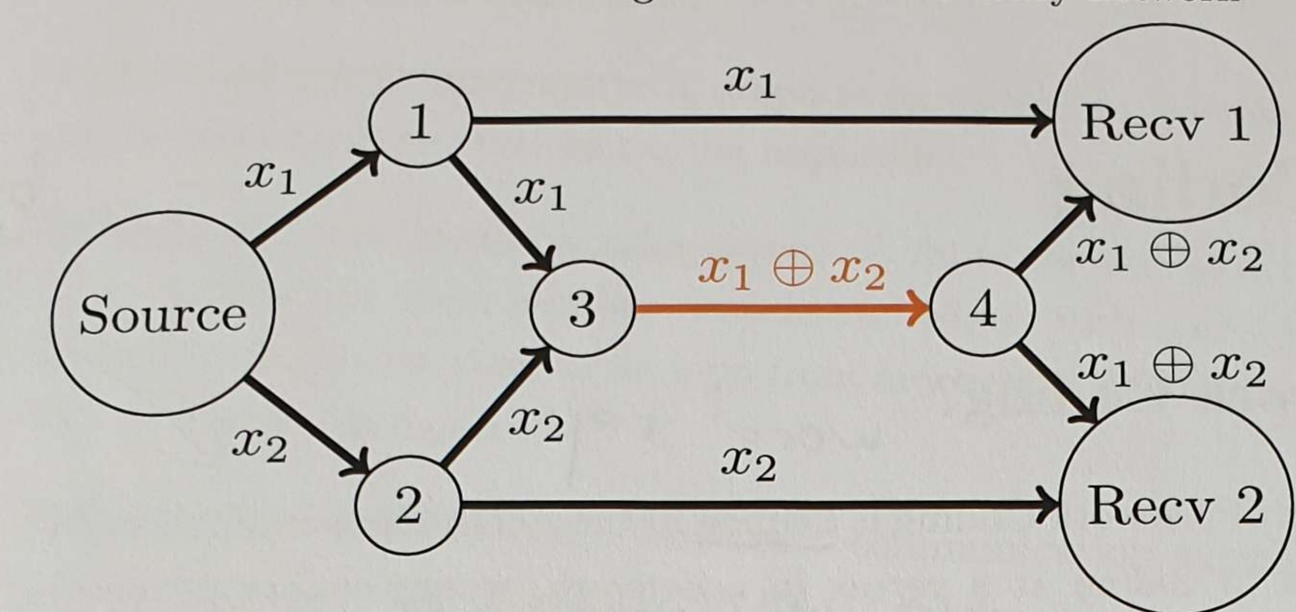
3.2 Advantages of Network Coding

3.2.1 Throughput gain and reduced complexity

Network coding gives a potential gain in throughput by communicating more information with fewer packet transmissions compared to the routing method. The butterfly network in [ACLY00] as a multicast in a wireline network is a standard example for an increase of throughput.

Better Explain Linear / Scalar Netw. coding (e.g. use ~~head to~~ (sub)section headline to stress it)

Figure 3.1: The butterfly network



like Section 6 the following section

Receiver 1 is a name → capital R
"the second receiver" is small

In Figure 3.1, we denoted a receiver by "Recv", which is used for all of figures in this study. With help of network coding, both Reveiver 1 and 2 can recover x_1 and x_2 by a bitwise XOR. Without network coding, an additional transmission between Vertex 3 and 4 must be supplemented to communicate the contents of 2 packets x_1 and x_2 from the source to Recv 1 and Recv 2, i.e. we must communicate x_1 or x_2 separately on this link twice under routing.

3.2.2 ~~Robustness~~

Packet loss is a particular issue in wireless packet networks due to several reasons, e.g. buffer overflow or communication failures. Sharing a common concept with Erasure Coding (EC) by exploiting a degree of redundancy to packets on any vertices in the network, the receivers are able to successfully recover the original packets from a large number of packet losses, e.g. $101 \circledast 10 \circledast 1$. The only difference is that packets are only encoded by the source in EC [FOG08]. This problem is dealed by acknowledgement messages in the mechanism of transmission control protocol (TCP).

this is too unimportant for this thesis to have sections about it. Better write it in the text.

make sure that the reader does not think this is part of the thesis

3.2.3 ~~Security~~

Network coding offers both benefits and drawbacks regarding to security. For example, node 4 is operated by an eavesdropper and it obtains only the packet $x_1 \oplus x_2$, so it cannot obtain either x_1 or x_2 and the communication is secure. Alternatively, if the eavesdropper controls node 3, it can anonymously send a fake packet masquerading as $x_1 \oplus x_2$, which is difficult to detect in network coding [HL08].

Vector Network Coding

3.2.4 ~~Increase the number of possible vertices~~

Data in the digital communication of today are represented in binary field, i.e. $\mathbb{F}_2$, so the coding coefficients are binary numbers. When we accomodate them into packets over $\mathbb{F}_2^t$ with $t \geq 2$, it has been studied in [EF11, EW18] showing that the so called vector network coding provides more possible vertices or more connected devices in a network compared to the linear network coding. In other words, network coding provides high adaptability for the worldwide network's requirement by extending the base field of packets.

why $\mathbb{F}_2$ only? Any $\mathbb{F}_q$!

3.3 Network Model

how do you mean that?

The "packet network" is practical, but difficult to accurately modelled. Hence, we introduce in next section the generalized combination network, where supports in varying a number of *connections*, specifically multicast (1 source to multiple receivers). Transmission rate is not regulated, so we do not consider congestion control in our network type [HL08].

Move many general defs from chapter 4 here! E.g. Vector/Scalar sol., $\mathbb{F}_{q^s}$, $\mathbb{F}_q^t$, Matrix form, gap size! ...

explain in much more detail

+ stress that we do not consider error correction here, only whether system is solvable or not!

is this really the reason why we consider the gen. comb. netw.?

Maybe something like ~~Sta~~ "Network Types Studied in this Thesis" or something similar

probably rather because gains are possible there & it is "easy" to analyze.

Please find out where the network could be found/how important it is → e.g. by reading works about the (g.) comb. netw.

* add a bit the "history" of the network (used anywhere? Why is it usually studied? Which papers have studied it? Back-track papers starting from Tuvi's & Antonia's paper & search keywords on google scholar.

titles are capital except for short words like "a", "to", "for", "the", ...

4 Generalized combination Networks $(\epsilon, l) - \mathcal{N}_{h,r,s}$

do you need this here?

4.1 Description

*

A generalized combination network $(\epsilon, l) - \mathcal{N}_{h,r,s}$ consists of 3 components from top to bottom: "Source" in the first layer, "Nodes" in the middle layer, and "Receivers" in the third layer. The network has a source with h messages, r nodes, and $\binom{r}{\alpha}$ receivers, which form a single source multicast network modeled as a finite directed acyclic multigraph [LYC03]. The source connects to each node by l parallel links and each node also connects to a receiver by l parallel links, which are respectively called a node's incoming and outgoing edges. Each receiver is connected by s links in total, specifically αl links from α nodes and ϵ direct links from the source, i.e. $s = \alpha l + \epsilon$. The combination network in [RA06] is the $(0, 1) - \mathcal{N}_{h,r,s}$ network and the $(1, 1) - \mathcal{N}_{h,r,s}$ network is called One-Direct Link Combination Network. Theorem 1 shows our interest of relations between the parameters h, α, ϵ and l.

layers?

everywhere!

was it first studied here?

use "ell" instead of l → better to distinguish from 1 etc. "one"

Move Fig. 4.1 here! (i.e. ~~at~~ on same page as paragraph above)

Reference!! You must always give proper refs.

Theorem 4.1. *The $(\epsilon, l) - \mathcal{N}_{h,r,s}$ network has a trivial solution if $l + \epsilon \geq h$, and it has no solution if $\alpha l + \epsilon < h$.*

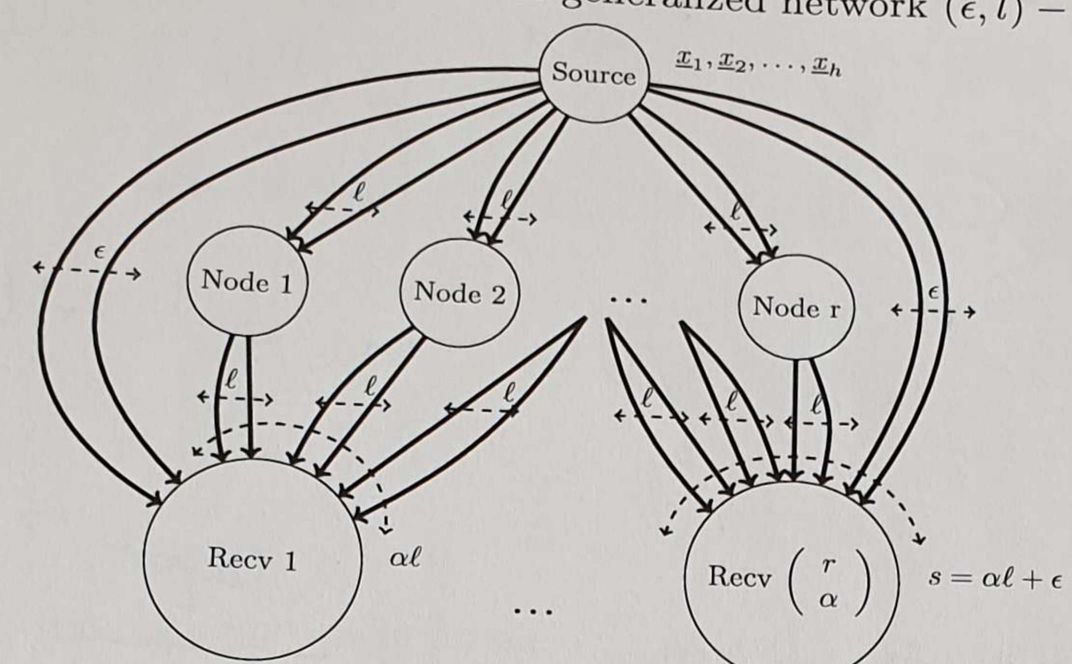
Proof: Following to the network coding max-flow min-cut theorem for multicast networks, the maximum number of messages from the source to each receiver is equal to the smallest min-cut between the source and any receiver. For our considered network, s links have to be deleted to disconnect the source from the receiver, which implies that the min-cut between the source and each receiver is at least s. Hence, $h \leq s \Leftrightarrow h \leq \alpha l + \epsilon$ □ ?

There exist at least $l + \epsilon$ disjoint links connected to each receiver. If $l + \epsilon \geq h$, each receiver can always reconstruct its requested messages on its links. Then we only need to do routing to select paths for the network. □

~~Remark 4.1.~~ Following to Theorem 4.1, we are interested in networks parameters satisfying this condition: $l + \epsilon + 1 \leq h \leq \alpha l + \epsilon$.

Only in text! maybe even there

Figure 4.1: The generalized network $(\epsilon, l) - \mathcal{N}_{h,r,s}$



nice!!

Table 4.1: Parameters of network coding

use \begin{center} \end{center} for the ~~[illegible]~~ table

h	The number of source messages
r	The number of nodes in the middle layer
$\binom{r}{\alpha}$	The number of receivers
ℓ	The source connects to each node by ℓ parallel links, and each node also connects to one receiver by ℓ parallel links
α	A receiver is connected by any α nodes in the middle layer
ϵ	The source additionally connects to each receiver by ϵ direct parallel links
s	Each receiver is connected by s links in total, with $s = \alpha\ell + \epsilon$.

also very nice!

4.2 Network Coding For This Network Family

4.2.1 Scalar network coding

$\mathbb{F}_{q_s}$ undefined (maybe already mention in chapter 3?

A message is equivalent to a symbol over $\mathbb{F}_{q_s}$. As a network of the multicast model, all receivers request the same packet? of h symbols at the same time [HT13]. A packet is a 1-dimentional subspace of $\mathbb{F}_{q_s}^h$?; hence, each receiver must obtain a subspace of $\mathbb{F}_{q_s}^h$, whose dimension is at least h, to be able to reconstruct the packet. Through ϵ direct links connected from the source to a receiver, the source can provide any required ϵ 1-dimensional subspaces of $\mathbb{F}_{q_s}^h$ for the corresponding receiver. Each receiver can accordingly

mention in more generality already in chapter 3

everywhere!

use \mathrm{s} for s since s is not a variable and might otherwise be confused with the network parameter s reconstruct the packet if and only if the linear span of α l-dimensional subspaces of $\mathbb{F}_{q_s}^h$ from the nodes is at least of dimension $h - \epsilon$. When this necessary condition is satisfied, the network is said to have a *solution* or to be *solvable*.

also ~~define~~ these notions already in general ~~for~~ in Chapter 3

Please recall briefly in preliminaries the parameters of an error-corr. code

Theorem 4.2. *The* $(0,1) - \mathcal{N}_{h,r,s}$ *network has a solution if and only if there exists an* $(r, |\mathbb{F}_{q_s}|\, h, r - \alpha + 1)$ $|\mathbb{F}_{q_s}|$*-ary error correcting code.* [RA06]

Theorem 4.3. *The* $(\epsilon, l) - \mathcal{N}_{h,r,s=\alpha l+\epsilon}$ *network is solvable over* $\mathbb{F}_q$ *if and only if there exists an* $\alpha - (h, l, h - l - \epsilon)_q^c$ *code with* r *codewords.* [EZ19]

not defined!

use \begin{theorem}[\cite{...}] ... \end{theorem}

4.2.2 Vector network coding

The messages are vectors of length t over $\mathbb{F}_q$, i.e. a vector solution is over field size q and dimension t. Such a vector solution has the same alphabet size as a scalar solution of field size q^t, and we denote $q_v = q^t$. A mapping from the scalar solution of field size q^t to a equivalent vector solution is represented in Example 4.1. Similarly with the scalar *linear* coding solution, each receiver can reconstruct its requested packet if and only if any α (lt)-dimensional subspaces span a subspace of dimension at least $(h - \epsilon)\, t$.

Theorem 4.4. *A vector solution for the* $(\epsilon, l) - \mathcal{N}_{h,r,s}$ *network exists if and only if there exists* $\mathcal{G}_q\,(ht, lt)$ *such that any* α *subspaces of the set span a subspace of dimension at least* $(h - \epsilon)\, t$*.* [EZ19]

Theorem 4.5. *The* $(\epsilon, l) - \mathcal{N}_{h,r,s=\alpha l+\epsilon}$ *network is solvable with vectors of length* t *over* $\mathbb{F}_q$ *if and only if there exists an* $\alpha - (ht, lt, ht - lt - \epsilon t)_q^c$ *code with* r *codewords.* [EZ19] — everywhere!

not defined!

Corollary 4.1. *The* $\alpha - (n = ht, n - k = ht - lt, \lambda = ht - lt - \epsilon t)_q^m$ *code formed from the dual subspaces of the* $\alpha - (n = ht, k = lt, \lambda = ht - lt - \epsilon t)_q^c$ *code yields the upper bound of* $\mathcal{A}_q\,(n = ht, n - k = ht - lt, \alpha; \lambda)$ *as maximum number of nodes for a vector network coding of the* $(\epsilon, l) - \mathcal{N}_{h,r,s}$ *network.*

not defined!

Nice overview in Sec. 4.2.1!

move to 4.3!

4.2.3 Network as a matrix channel

To formulate this description, the source has a set of disjoint messages referred to packets which are either symbols from $\mathbb{F}_{q_s}$ (scalar coding) or vectors of length t over $\mathbb{F}_q$ (vector coding). Each link in the network carries functions of the packets, and a *network code* is a set of these functions. The network code is called *linear* if all the functions are linear and nonlinear otherwise. Each receiver $R_j, j \in \{1, \ldots, N\}$ requests a subset of the source's length-h messages, and this subset is called a *packet*. Through all the functions on the links from the source to each receiver, the receiver obtains several linear combinations of the h messages to form a linear system of equations for its requested packets. The coefficients of a linear combination are called *global coding vectors* [SET03]. The linear equation system that any receiver R_j has to solve is as following:

nice !

$$\begin{array}{c|c} \textit{Scalar} & \textit{Vector} \\ \underbrace{\begin{bmatrix} y_{j_1} \\ \vdots \\ y_{j_s} \end{bmatrix}}_{\mathbb{F}_{q_s}^{s}} = \underbrace{\boldsymbol{A}_j}_{\mathbb{F}_{q_s}^{s\times h}} \cdot \underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_h \end{bmatrix}}_{\mathbb{F}_{q_s}^{h}} & \underbrace{\begin{bmatrix} \underline{y}_{j_1} \\ \vdots \\ \underline{y}_{j_s} \end{bmatrix}}_{\mathbb{F}_{q}^{st}} = \underbrace{\boldsymbol{A}_j}_{\mathbb{F}_{q}^{st\times th}} \cdot \underbrace{\begin{bmatrix} \underline{x}_1 \\ \vdots \\ \underline{x}_h \end{bmatrix}}_{\mathbb{F}_{q}^{th}} \end{array} \tag{4.1}$$

either use (i) $\underline{x}$, $\underline{A}$ or bold (ii) $\boldsymbol{x}$ & $\boldsymbol{A}$
I prefer (ii) in latex documents
((i) is good for blackboard)

The transfer matrix $\boldsymbol{A}_j$ contains the links' *global coding vectors*, which are combined by the coefficients of linear combinations on αl links from α nodes and ϵ direct-links to the corresponding receiver R_j:

move to ch. 3

$$\begin{array}{c|c} \textit{Scalar} & \textit{Vector} \\ \boldsymbol{A}_j = \begin{bmatrix} \underline{a}_{j_1} \\ \vdots \\ \underline{a}_{j_{\alpha l}} \\ \vdots \\ \underline{a}_{j_{\alpha l+\epsilon}} \end{bmatrix} & \boldsymbol{A}_j = \begin{bmatrix} \boldsymbol{A}_{j_1} \\ \vdots \\ \boldsymbol{A}_{j_{\alpha l}} \\ \vdots \\ \boldsymbol{A}_{j_{\alpha l+\epsilon}} \end{bmatrix} \end{array}$$

In general, the network is represented as a matrix channel for both scalar and vector coding:

Definition 4.1. Network As Matrix Channel

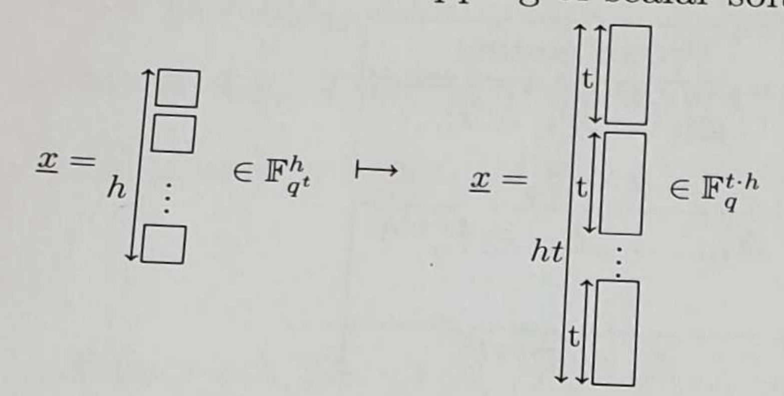
The channel output can be written as: $\boldsymbol{Y}_j = \boldsymbol{A}_j \cdot \boldsymbol{X}$

Because we reconstruct $\boldsymbol{X}$ with knowing $\boldsymbol{A}_j$, i.e. the network structure is known, our network is coherent. A network is *sovable* or a network code is a *solution*, if each receiver can reconstruct its requested messages or solve the system with a unique solution for scalars $x_1, \ldots, x_h$, or vectors $\underline{x}_1, \ldots, \underline{x}_h$. Therefore, we want to find global coding vectors such that the matrix $\boldsymbol{A}_j$ has full-rank for every $j = 1, \ldots, N$, and such that q_s or q^t is minimized. In Example 4.1, we provide a vector solution of field size q and dimension t, which has the same alphabet size as a scalar solution of field size q^t

Example 4.1. Given $h = 3, q = 2, t = 2$, we consider the extension field $\mathbb{F}_{q^t=2^2}$. The example shows how mapping messages from scalar coding to vector coding.

which example?

Figure 4.2: The mapping of scalar solution over $\mathbb{F}_{q_s=q^t}$ to the equivalent vector solution



nice
(use \begin{center} to center figure)

Example 4.2. We use the table of the extension field $\mathbb{F}_{2^2}$ with the primitive polynomial $f(x) = x^2 + x + 1$:

power of α	polynomial	binary vector
-	0	00
α^0	1	01
α^1	α	10
α^2	$\alpha + 1$	11

also here & everywhere in the following

For scalar coding, the messages are $x_1, \ldots, x_{h=3} \in \mathbb{F}_{2^2}$, and for vector coding the messages are $\underline{x}_1, \ldots, \underline{x}_{h=3} \in \mathbb{F}_2^2$. From the polynomial column, let's choose arbitrarily a scalar vector $\underline{x}_{scalar} = (x_1, x_2, x_3) = (1, \alpha, \alpha + 1)$. Then, we map it to $\underline{x}_{vector} = (\underline{x}_1, \underline{x}_2, \underline{x}_3)$ by using the binary vector column as following:

$$\begin{bmatrix} x_1 = 1 \\ x_2 = \alpha \\ x_3 = \alpha + 1 \end{bmatrix} \mapsto \begin{bmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \end{bmatrix},$$

nice!

where we use the following rule for mapping x_i individually: $a_0 \cdot \alpha^0 + a_1 \cdot \alpha^1 + \ldots + a_{t-1} \cdot \alpha^{t-1} \mapsto \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{t-1} \end{pmatrix}$.

To summarize the notations of both scalar and vector coding, we represent them as in ~~the~~ Table 4.2:

Table 4.2 is a name (no "the")

nice!

Table 4.2: Notations of network coding

	Scalar Coding	Vector coding
Source Messages/Packets	$x_1, \ldots, x_h \in \mathbb{F}_{q_s}$ $\underline{x} \in \mathbb{F}_{q_s}^h$	$\underline{x}_1, \ldots, \underline{x}_h \in \mathbb{F}_q^t$ $\underline{x} \in \mathbb{F}_q^{th}$
Global Coding Vectors Of Receiver R_j	$\underline{a}_{j_1}, \ldots, \underline{a}_{j_s} \in \mathbb{F}_{q_s}^h$	$\boldsymbol{A}_{j_1}, \ldots, \boldsymbol{A}_{j_s} \in \mathbb{F}_q^{t \times th}$
Transfer Matrix Of Receiver R_j	$\boldsymbol{A}_j \in \mathbb{F}_{q^t}^{s \times h}$	$\boldsymbol{A}_j \in \mathbb{F}_q^{st \times th}$
Packets On Receiver R_j	$y_{j_1}, \ldots, y_{j_s} \in \mathbb{F}_{q_s}$ $\underline{y} \in \mathbb{F}_{q_s}^s$	$\underline{y}_{j_1}, \ldots, \underline{y}_{j_s} \in \mathbb{F}_q^t$ $\boldsymbol{Y}_j \in \mathbb{F}_q^{st}$
Number of nodes	r_{scalar}	r_{vector}

center

move to ch. 3

~~# choices for coeffs~~

Remark 4.2. By using the vector coding, the upper bound number of solutions increases from q^{tkh} to q^{t^2kh}. Therefore, vector network coding offers more freedom in choosing the coding coefficients than does scalar linear coding for equivalent alphabet sizes, and a smaller alphabet size might be achievable [EF11]. By this advantage, we can have higher amount of receivers, i.e. higher amount of nodes, in vector network coding.

nice!

amount is more for uncountable objects, e.g. "amount of water". "number of nodes" is better.

4.2.4 Comparison between scalar and vector solutions by the gap size

The *gap* represents the difference between the smallest field (alphabet) size for which a scalar linear solution exists and the smallest alphatbet size for which we can construct a vector solution. In this study, we define a solvable vector network coding over the field size $\mathbb{F}_q^t$, and we conjecture the minimum amount of nodes such vector solution can achieve, i.e. $r_{vector} \geq f_1(q)$, with $f_1 : \mathbb{Z} \mapsto \mathbb{Z}$. Meanwhile, we has a scalar solution for the same network existing if and only if: $r_{scalar} \leq f_2(q_s)$, with $f_2 : \mathbb{Z} \mapsto \mathbb{Z}$. To find the field size q_s required for a scalar solution to reach the minimum achievable vector solution's nodes in this setting, we consider $r_{max,scalar} = f_2(q_s) = f_1(q) = r_{min,vector}$. Finally, we calculate the gap by $g = q_s - q_v = q_s - q^t$. Throughout this study, we show that vectors solutions significantly reduce the required alphabet size by this gap.

4.3 Special cases of generalized combination network

4.3.1 The $(l-1)$-Direct Links and l-Parallel Links $\mathcal{N}_{h=2l,r,s=3l-1}$

Why special cases?

This subfamily contains the largest number of direct links from the source to the receivers. For $l \geq 2$, this network $(\epsilon = l - 1, l) - \mathcal{N}_{h=2l,r,s=3l-1}$ yields the gap $q^{(l-1)t^2/l + O(t)}$

mention work of Antonia & Tuvi i.e. that they derived bounds on gap sizes of several special cases

which gap? who showed it? This is only a lower bound on the gap size, right?

among which networks?

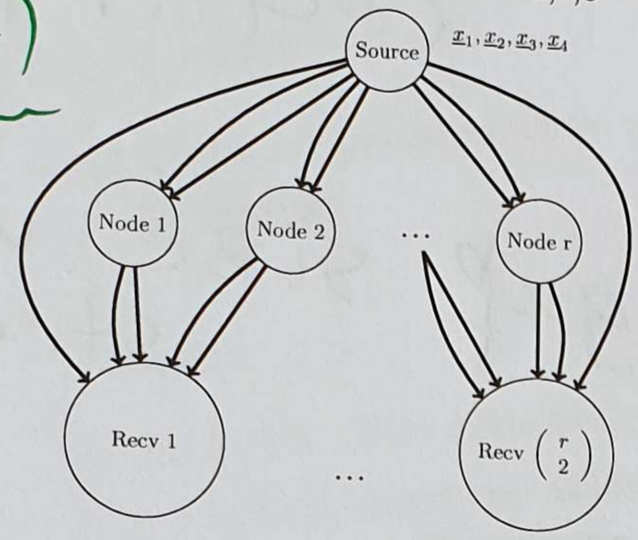
between vector solutions and optimal scalar solutions. The vector solution is based on an $\mathcal{MRD}\,[lt \times lt, t]_q$ code. Further, the gap tends to $q^{t^2/2+\mathcal{O}(t)}$ for large l.

~~what do you mean~~ ?

Lemma 4.1. +ref. *There is a scalar linear solution of field size q_s for the $(\epsilon = l-1, l) - \mathcal{N}_{h=2l,r,s=3l-1}$ network, where $l \geq 2$, if and only if $r \leq \begin{bmatrix} 2l \\ l \end{bmatrix}_{q_s}$.*

gap size (in dep. on l) as theorem

Figure 4.3: The $(1,2) - \mathcal{N}_{4,r,5}$ network as an example of the $(l-1,l) - \mathcal{N}_{2l,r,3l-1}$



center + larger (font very small)

4.3.2 The 1-Direct Link and l-Parallel Links $\mathcal{N}_{h=2l,r,s=2l+1}$

this is only the smallest for which an outperforming one was found so far!

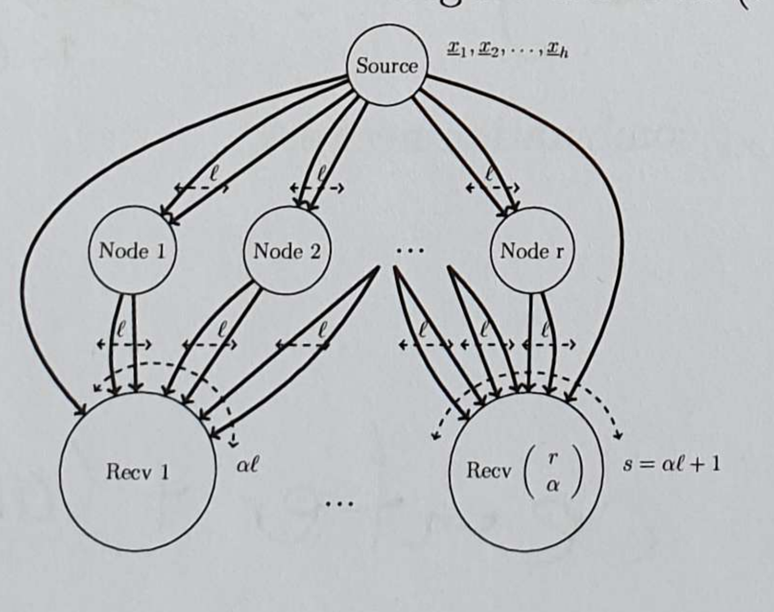
This is the smallest direct-link subfamily has an vector solution outperforming the optimal scalar solution, i.e. an vector solution outperforming the optimal scalar has not yet been found for the network $(0, l > 1) - \mathcal{N}_{h,r,s}$. Similar to the previous subfamily $(\epsilon = l-1, l) - \mathcal{N}_{h=2l,r,s=3l-1}$, when $l \geq 2$ or $h \geq 4$, this network yields the largest gap $q^{t^2/2+\mathcal{O}(t)}$ in the alphabet size by using the same approach with an $\mathcal{MRD}\,[lt \times lt, (l-1)t]_q$ code.

what do you mean by "smallest"?

Does not mean that there is a smaller one

gap size as theorem!

Figure 4.4: The $(l-1,l) - \mathcal{N}_{2l,r,3l-1}$ network



what does largest gap mean?

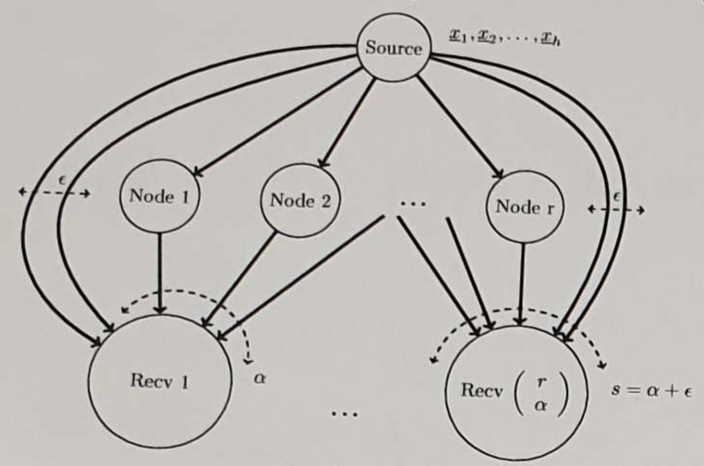
center + larger

4.3.3 The ϵ-Direct Links $\mathcal{N}_{h,r,s}$

? English

This subfamily is denoted as $(\epsilon \geq 1, l = 1) - \mathcal{N}_{h,r,s}$ and is the most focus topic on this thesis, because it motivates some interesting questions on a classic coding problem and on a new type of subspace code problem. In Section 5.1, we show our largest code set with low number of subspace codes for the network $(\epsilon = 1, l = 1) - \mathcal{N}_{h=3,r,s=4}$.

Figure 4.5: The $(\epsilon \geq 1, l = 1) - \mathcal{N}_{h,r,s}$ network



center + larger!

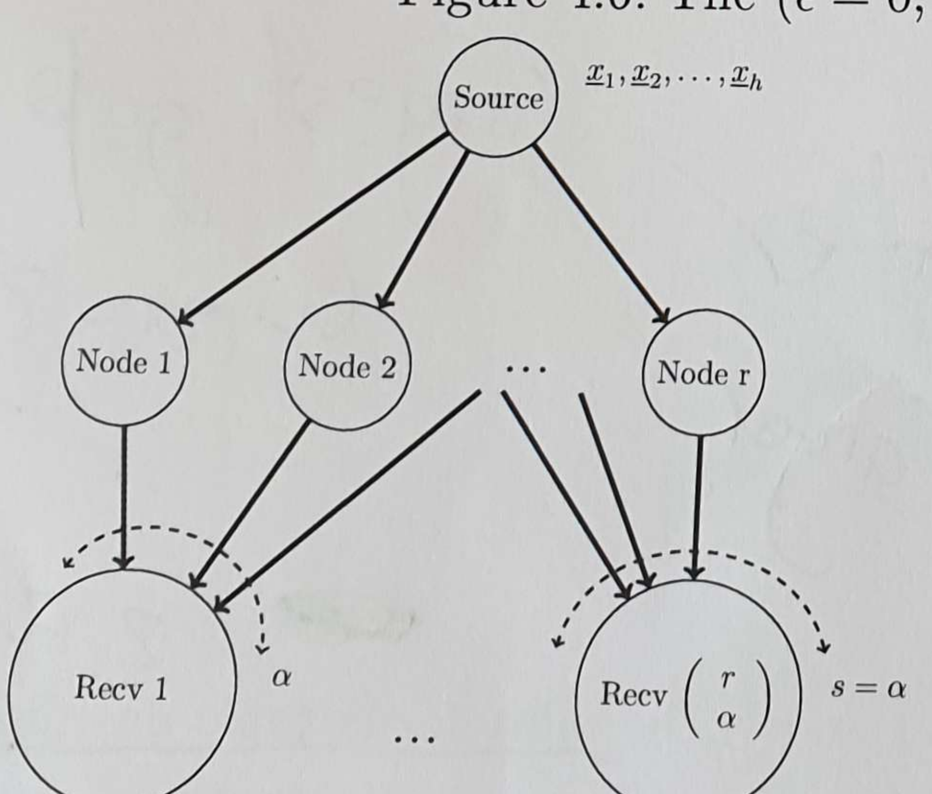
gap size unknown?
stress!

4.3.4 The $(\epsilon = 0, l = 1) - \mathcal{N}_{h,r,s}$ Combination Network

Since the scalar solution for the combination network uses an MDS code, a vector solution based on subspace codes must go beyond the MDS bound, i.e. Singleton bound $d \leq n - k + 1$, to outperform the scalar one. In paper [EW18], it is proved that vector solutions based on subspace codes cannot outperform optimal scalar linear solutions for $h = 2$, and they conjecture it for all h. Unfortunately, a vector solution based on an $\mathcal{MRD}\,[t \times t, t]_q$ code is also proved that it cannot outperform the optimal scalar linear solution.

English? Sentence does not make sense

Figure 4.6: The $(\epsilon = 0, l = 1) - \mathcal{N}_{h,r,s}$ combination network



center + larger

4.3.5 The largest possible gap between q_v and q_s in previous studies

$h \leq 2l$ and $\epsilon \neq 0$ very, very nice!

For this network, the number of direct links is at least 1, i.e. $\epsilon \geq 1$, and the number of parallel links is less than half of the number of source messages, i.e. $l \leq \frac{h}{2}$.

why two times even/odd resp.

h is even The above $(l-1,l)-\mathcal{N}_{2l,r,3l-1}$ network achieves the largest gap $q_s = q^{(h-2)t^2/h+\mathcal{O}(t)}$.

all or only h≥4?

h is odd The $(l-2,l)-\mathcal{N}_{2l-1,r,3l-2}$ network achieves the largest gap $q_s = q^{(h-3)t^2/(h-1)+\mathcal{O}(t)}$

all or h≥5?

$h \geq 2l$ and $\epsilon \neq h-2l$

h is even The same above $(l-1,l)-\mathcal{N}_{2l,r,3l-1}$ network achieves the largest gap $q_s = q^{(h-2)t^2/h+\mathcal{O}(t)}$.

h is odd The $(l-1,l)-\mathcal{N}_{2l+1,r,3l-1}$ network achieves the largest gap $q_s = q^{(h-3)t^2/(h-1)+\mathcal{O}(t)}$.

Remark 4.3. The achieved gap is $q^{(h-2)t^2/h+\mathcal{O}(t)}$ for any $q \geq 2$ and any even $h \geq 4$. If $h \geq 5$ is odd, then the achieved gap of the alphabet size is $q^{(h-3)t^2/(h-1)+\mathcal{O}(t)}$ [EW18].