Universidad de Valladolid

Facultad de Ciencias

Trabajo Fin de Grado

Grado en Matemáticas

Seguridad en códigos de red

Autor: Diego Martín Garzón

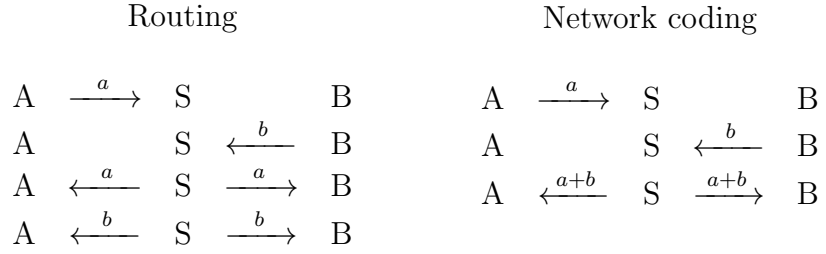
Tutor: Diego Ruano Benito

Índice general

Introducción

El nacimiento del concepto de 'Network coding' o codificación de red podemos situarlo en el año 2000, con la publicación de Ahlswede, Cai, Li, y Yeung [Ahls]. Para empezar a explicar en qué consiste la codificación de red, consideremos una red de comunicación compuesta por nodos y aristas o enlaces, en la que se quiere enviar un mensaje desde un nodo fuente o emisor, y hacerlo llegar a un conjunto de nodos receptores. Podemos preguntarnos qué requisitos es necesario pedir a la red para que la comunicación sea exitosa (los receptores reciban correctamente el mensaje enviado) y eficiente.

En las redes informáticas existentes, por ejemplo el actual Internet, la información es enviada por la fuente y transmitida a lo largo de los nodos intermedios de la red, los cuales almacenan los paquetes de datos y reenvían una copia a los siguientes nodos de la red (proceso conocido como 'routing') a través de las conexiones sin necesidad de realizar un procesamiento de los datos (más allá del reenvío). En cambio, lo que propone la codificación de red, es permitir a los nodos intermedios que combinen los paquetes de datos que les llegan, para transmitir el paquete resultante. Si esta combinación es lineal, hablamos de codificación de red lineal, que será el área que cubriremos en este trabajo.
En el problema de codificación de red que estudiaremos, se quiere enviar una cierta cantidad de mensajes simultáneamente, que serán elementos de un cuerpo finito (podemos pensar en el cuerpo $\mathbb{F}_2$, si queremos transmitir una serie de bits). La capacidad de cada canal será unitaria, lo que significa que por cada canal de la red solo podrá viajar un mensaje o paquete a la vez por unidad de tiempo. Además, se quiere que todos los receptores reciban todos los mensajes, esta situación se hace llamar multidifusión. En estas condiciones el routing no es óptimo para la transmisión en comparación con la codificación de red, que supone una mejora en la eficiencia y la robustez de esta. El siguiente esquema es un simple ejemplo que ilustra lo anterior.



A y B quieren intercambiar los mensajes a y b via el nodo intermedio S. A (resp. B) envía a (resp. b) a S, el cual transmite $a + b$ en vez de enviar a y b por separado en dos pasos. Tanto A como B pueden recuperar el mensaje que esperaban ($(a + b) - b = a$ o $(a + b) - a = b$) y se ha ahorrado una transmisión.

A la explicación de esta técnica se dedica el Capítulo 1 del trabajo, en el que nos hemos servido principalmente de las referencias [Casp], [HoLun] y [GeTh]. Se presentan los conceptos que proporcionan la caracterización algebraica del network coding. Una red de comunicación vendrá representada por un grafo dirigido y acíclico, en cuyos nodos se combinan los paquetes. Las diferentes combinaciones que se realizan y la forma de la red, vendrán descritas por unos coeficientes, recogidos en matrices, que determinan un problema de network coding. La resolución del problema, este es, determinar si con ciertos coeficientes se consigue una comunicación exitosa, pasa por comprobar si el llamado polinomio de transferencia, descrito en términos de dichos coeficientes, se anula o no. En el mismo capítulo, se trata el método de obtención de coeficientes conocido como random network coding o codificación de red aleatoria, que consiste en asignar los coeficientes de la red de forma aleatoria. Esta opción es realmente útil en la práctica, pues permite abordar el problema de comunicación sin necesariamente tener que conocer la red por la que se transmite, o en redes que son cambiantes en el tiempo, que son las situaciones más realistas. Es inmediato darse cuenta de que este método no tiene por qué resultar siempre en una comunicación satisfactoria, pues puede haber elecciones de coeficientes que no produzcan una solución al problema. Por ello, se muestran condiciones necesarias y suficientes (tamaño del cuerpo finito, existencia de flujos...) para lograr la transmisión de información así como algún ejemplo explicativo, y se prueba una de las posibles cotas de probabilidad de éxito con el random network coding. Se presenta también el Algoritmo de Jaggi-Sanders y un ejemplo de su uso, por el cual dada una red, se le asignan coeficientes que consigan una comunicación exitosa.

Previamente a enviar un mensaje por una red, se quiere codificarlo para proporcionar seguridad y fiabilidad a la comunicación. Esto se hará por medio de códigos correctores dotados con la métrica de rango, los cuales describimos en el Capítulo 2 del trabajo, habiéndonos apoyado en los escritos [GoRav], [Gabi] y [Sil] principalmente. Un código lineal no es más que un subespacio vectorial $\mathcal{C} \subseteq \mathbb{F}_q^n$, y nuestros códigos para la métrica de rango son códigos lineales construidos sobre alguna extensión de cuerpos $\mathbb{F}_q \hookrightarrow \mathbb{F}_{q^m}$. Al ser $\mathbb{F}_{q^m}$ un espacio vectorial de dimensión m sobre $\mathbb{F}_q$, una palabra $x \in \mathcal{C}$ puede verse como una matriz $X \in \mathbb{F}_q^{n\times m}$ cuyas entradas son las coordenadas de las componentes de x en una base de $\mathbb{F}_{q^m}$. Y esto es lo que nos permite introducir la métrica de rango, que mide la distancia de dos palabras, vistas como matrices, como el rango de la matriz diferencia. En términos de dicha distancia se enuncia la Cota de Singleton, que es alcanzada por los códigos llamados de máxima distancia de rango o MRD, entre los que destacan los códigos de Gabidulin. Durante el desarrollo de lo anterior se exponen ejemplos y explicaciones para apoyarlo. Se finaliza el capítulo introduciendo los polinomios linealizados y algunas de sus propiedades, que muestran un símil entre la construcción de los códigos Reed-Solomon y los códigos de Gabidulin, y tienen importancia en la descodificación de estos últimos.

La seguridad de la que queremos dotar la comunicación se estudia desde el punto de vista de la teoría de la información en el Capítulo 3, el cual ha sido realizado basándonos en [LoVei] y [CoJo]. En él, se explican los conceptos de entropía e información mutua, y otras definiciones partir de ellos, además de aplicar los conceptos y resultados en ejemplos concretos. La entropía nos sirve como medida de incertidumbre sobre una variable aleatoria discreta, y depende solo de la distribución de probabilidades. Así, la entropía de una variable aleatoria cuya distribución asigne más probabilidad a unos sucesos que a otros, será menor que la de otra cuyos posibles valores son equiprobables. La información mutua entre dos variables aleatorias discretas se define en términos de entropías, y mide cuánta información posee una variable sobre la otra. Así, aplicando estos conceptos a nuestro modelo de comunicación en el que se quiere que un intruso no obtenga información sobre el mensaje original a partir de sus observaciones, esto es equivalente a que la información mutua entre el mensaje y las observaciones del espía sea nula.

Una vez explicado en qué consiste el network coding, la clase de códigos que se usan para codificar el mensaje antes de enviarlo, y la perspectiva de seguridad con la que queremos que se realice la

comunicación, se procede a presentar en el Capítulo 4 el esquema de comunicación que reúne todo lo tratado. Para ello nos hemos basado en el artículo [SiKsch] y en [Sil], ampliando las explicaciones y aplicándolas en ejemplos. También se han añadido resultados necesarios para una completa comprensión del texto, así como completado varias demostraciones. Se comienza explicando la situación más simple de nuestro problema, conocida como Wiretap Channel II: un emisor codifica un mensaje en n paquetes y lo envía a un receptor por una arista, la cual es pinchada por un espía que observa $\mu < n$ de los paquetes. Se demuestra que la comunicación de $k = n - \mu$ paquetes es segura si se utiliza una codificación por cogrupo basada en un código MDS. Esta codificación se apoya en el hecho de que el mensaje a enviar determina un cogrupo del código, del cual se obtiene la palabra codificada. Este modelo lo extendemos a una red en la que se aplica network coding, y se comprueba que la seguridad no está garantizada pues existe una dependencia entre el código usado al inicio de la comunicación y el diseño del network coding. Por ello, se sustituyen los códigos MDS por códigos para la métrica de rango MRD, y se prueba que con la codificación por cogrupo la comunicación es segura si y solo si se usan esta clase de códigos, y la longitud de los paquetes transmitidos es al menos n.

Capítulo 1

Network Coding

1.1. Introducción del problema

Para presentar el problema de codificación de red, fijaremos una notación y definiremos los conceptos básicos sobre los que trabajaremos. En la realización de este primer capítulo han sido de gran ayuda los escritos [Casp], [HoLun] y [GeTh].
Representaremos una red, o red de comunicación, como un grafo dirigido finito y acíclico (esto es, que no contenga ciclos dirigidos). El hecho de que el grafo sea dirigido, es lo que proporciona a la comunicación una orientación: los mensajes comienzan su marcha siguiendo una trayectoria definida en un único sentido, sin vuelta atrás. En cuanto a trabajar con grafos sin ciclos, esto permite considerar un orden en las aristas, siguiendo el cual, un nodo codifica solo cuando ha recibido toda la información necesaria por sus aristas entrantes. Con ciclos en la red dicho orden no existe, y constituye otro tipo de problema sobre el que se puede consultar en [Yeung].
El grafo que representa la red se expresa de la forma $G = (V, E)$ donde E denota el conjunto de aristas (canales de comunicación), y V el conjunto de vértices (nodos). Las aristas toman la forma (u, v) siendo u y v nodos, aunque por comodidad las nombraremos como números naturales.

Denotamos por $S = \{s_1, ..., s_{|S|}\} \subseteq V$ al conjunto de nodos emisores, estos son los nodos donde se generan los mensajes, y $R = \{r_1, ..., r_{|R|}\} \subseteq V$ al conjunto de nodos receptores, donde queremos que lleguen los mensajes. En este trabajo, solo contemplaremos la existencia de un único nodo emisor s, pues si existiesen varios, podríamos extender la red y obtener un problema equivalente pero con un solo emisor. Esto lo explicaremos más adelante.

Llamamos vector mensaje al vector $x = (X_1, .., X_h)$ cuyas entradas, llamadas mensajes, toman valores en un cuerpo finito $\mathbb{F}_q$. Como ya comentamos anteriormente, por un canal solo podrá viajar un mensaje a la vez, y este lo hará en una unidad de tiempo.

Para cada nodo $v \in V$, sea $in(v)$ el conjunto de aristas que entran en v, es decir de la forma (u, v), y $out(v)$ el conjunto de aristas que salen de v, es decir de la forma (v, u). De igual manera dada una arista $j = (u, v)$ denotaremos $in(j) = in(u)$ y $out(j) = out(v)$.
Al considerar grafos acíclicos, siempre existe un orden ancestral, no necesariamente único, el cual asumiremos. Este es un orden tal que si para un par de aristas i, j se tiene $j \in out(i)$, entonces $j > i$.
La Figura 1.1 muestra un ejemplo de red sobre el que ver la notación introducida.

La clave de la codificación de red consiste precisamente en permitir a los nodos de la red que combinen la información que les llega para transmitirla. Por ello se definen a continuación las variables que explican la codificación que se lleva a cabo en los nodos, y que contienen la información que se transmite por cada arista de la red.

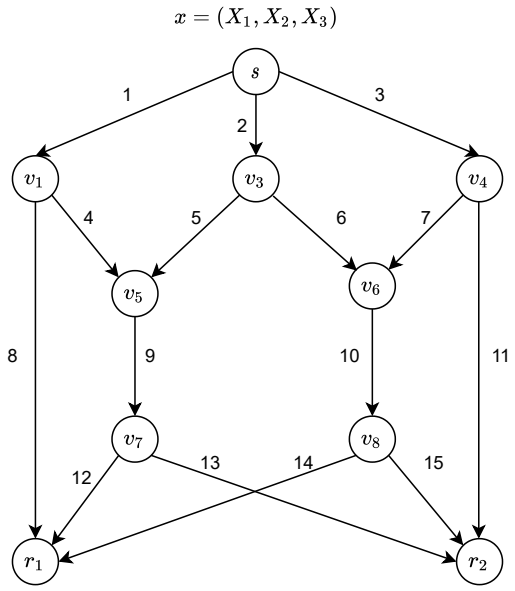


Figura 1.1: Red en la que el emisor s envía los mensajes X_1, X_2, y X_3 a los receptores r_1 y r_2

Para cada arista $j = (u, v) \in E$ consideramos la variable $Y(j)$ que toma valores en $\mathbb{F}_q$. La variable $Y(j)$ depende por medio de una función de codificación f_j del resto de las variables $Y(i)$ donde $i \in in(j)$ y de los mensajes X_k que se generen en u. Esto es:

$$Y(j) = f_j((Y(i) \mid i \in in(j)), (X_k \mid X_k \text{ se genera en } u)).$$

Además, consideramos la llamada función demanda D que aplica el conjunto de receptores R en el conjunto de subconjuntos ordenados de $(X_1, .., X_h)$ que a cada receptor $r \in R$ le asocia el conjunto de mensajes que quiera recibir, esto es $D(r) = (X_{i_1}, .., X_{i_t})$. En el caso que nos ocupa, el de multidifusión, todos los receptores han de recibir todos los mensajes, luego para todo $r \in R$ tendremos $D(r) = (X_1, .., X_h)$.
Generalmente los receptores no recibirán los mensajes deseados explícitamente, ya que estarán codificados. Por ello para explicar la descodificación que se produce en los receptores, introducimos las siguientes variables. Para cada receptor r consideramos las variables $Z_1^{(r)}, ..., Z_h^{(r)}$ tomando valores en $\mathbb{F}_q$, que dependen de las variables asociadas a las aristas entrantes a r vía las funciones de decodificación $d_j^{(r)}$

$$Z_j = d_j^{(r)}(Y(i) \mid i \in in(r)) \text{ con } j = 1, ..., h$$

Especificando una red $G = (V, E)$ con un conjunto de emisores S, un conjunto de receptores R y un vector mensaje $x = (X_1, ..., X_h)$ se define un problema de codificación de red. Dar una solución a este problema consiste en proporcionar un cuerpo finito $\mathbb{F}_q$, y unas funciones de codificación f_j y de decodificación d_j de tal manera que cada nodo receptor reproduzca perfectamente los mensajes que quería recibir. Esto es

$$(Z_1^{(r)}, ..., Z_h^{(r)}) = D(r)$$

se ha de cumplir para todo $r \in R$. En nuestro caso de multidifusión $D(r) = (X_1, ..., X_h)$. Un problema de codificación de red para el que existe solución, se dice que es resoluble. Cuando las funciones de codificación f_j y de decodificación d_j son lineales en el cuerpo, se habla de codificación de red lineal, que es el contexto en el que situamos este trabajo.

La existencia de un orden ancestral en la red nos permite definir las variables $Y(j)$ siguiendo dicho orden. La codificación de red lineal para una arista $j = (u, v)$ puede ser representada por la ecuación

$$Y(j) = \sum_{i \in in(j)} f_{i,j} Y(i) + \sum_{\substack{u \in S \\ 1 \leq i \leq h}} a_{i,j} X_i \tag{1.1}$$

y la decodificación en cada nodo receptor r, con $j = 1, ..., h$

$$Z_j^{(r)} = \sum_{i \in in(r)} b_{i,j}^{(r)} Y(i) \tag{1.2}$$

donde $a_{i,j}$, $f_{i,j}$ y $b_{i,j}$ son elementos de $\mathbb{F}_q$. Los $a_{i,j}$ y $f_{i,j}$ son llamados *coeficientes de codificación* (locales), y los $b_{i,j}$ *coeficientes de decodificación*. En ocasiones por comodidad, nos referiremos ellos con la notación $(a, f, b^{(r)})$.

Puesto que todas las operaciones de codificación en la red son operaciones lineales de la forma (1.1), se deduce de manera inductiva que para cada arista j, la variable $Y(j)$ puede verse como una combinación lineal de los mensajes X_i. La deducción se sigue del hecho de que las variables $Y(j)$ se definen de forma recursiva con combinaciones lineales en el cuerpo, y las primeras variables en definirse son las correspondientes a las aristas que salen inmediatamente del nodo emisor, que son solo combinaciones lineales de los mensajes X_i. Esta forma de ver las variables $Y(j)$ la podemos escribir como sigue

$$Y(j) = \sum_{i=1}^{h} c_{i,j} X_i \tag{1.3}$$

donde los coeficientes $c_{i,j} \in \mathbb{F}_q$ son función de los coeficientes de codificación $a_{i,j}$ y $f_{i,j}$, y el vector

$$c_j = (c_{1,j}, ..., c_{h,j}) \in \mathbb{F}_q^h$$

lo llamaremos *vector de codificación global* de la arista j.
De lo anterior se deduce que un vector de codificación global se puede escribir en función de sus predecesores en la red de la forma siguiente

$$c_j = \sum_{i \in in(j)} f_{i,j} c_i + \sum_{1 \leq l \leq h} a_{l,j} e_l$$

donde e_l denota el vector canónico con un 1 en la posición l-ésima.

Por ejemplo, para cada arista j saliente del nodo emisor de la red, es obvio que su vector de codificación global será de la forma $c_j = (a_{1,j}, ..., a_{h,j})$ (pues no hay términos en el primer sumatorio de (1.1)). Igualmente, las variables $Z_i^{(r)}$ (es decir, los mensajes descodificados que proporcionan los receptores como salida) resultan de operaciones lineales con los mensajes X_i, por lo que para cada nodo receptor r, también tenemos una forma de expresar el vector decodificado $z^{(r)} = (Z_1^{(r)}, ..., Z_h^{(r)})$ en función del vector mensaje $x = (X_1, ..., X_h)$ mediante un sistema de ecuaciones lineales

$$z^{(r)} = xM^{(r)} \tag{1.4}$$

La matriz $M^{(r)}$ es la denominada *matriz de transferencia* para el receptor r, y en ella se recoge toda la información relativa a la red. La definimos a partir del producto de las matrices formadas por los coeficientes de codificación y descodificación que hemos venido introduciendo hasta ahora, del siguiente modo

$$M^{(r)} = A(I - F)^{-1} B^{(r)} \tag{1.5}$$

donde

- $A = (A_{i,j})$ es una matriz $h \times |E|$ cuyas entradas son de la forma

$$A_{i,j} = \begin{cases} a_{i,j} & \text{si } j \in out(s) \\ 0 & \text{en cualquier otro caso} \end{cases}$$

  Es decir, las entradas no nulas son los coeficientes que aparecen combinándose con los mensajes X_i en (1.1) para generar las varibles $Y(j)$ correspondientes a las aristas que salen del nodo emisor. O dicho de otro modo, si una arista j no se origina en el nodo emisor, entonces la j-ésima columna de A es nula.

- $F = (F_{i,j})$ es una matriz $|E| \times |E|$ cuyas entradas son de la forma

$$F_{i,j} = \begin{cases} f_{i,j} & \text{si } j \in out(i) \\ 0 & \text{en cualquier otro caso} \end{cases}$$

  O lo que es lo mismo: la entrada (i, j) será $f_{i,j}$ si la arista j se origina en el nodo donde termina la arista i. Como consideramos un orden ancestral, F será triangular superior con ceros en su diagonal.

- $B^{(r)} = (B^{(r)}_{i,j})$ es una matriz $|E| \times h$ para cada receptor $r \in R$. Sus entradas vienen definidas de la siguiente manera

$$B^{(r)}_{i,j} = \begin{cases} b^{(r)}_{i,j} & \text{si } i \in in(r) \\ 0 & \text{en cualquier otro caso} \end{cases}$$

  Es decir, para cada $r = 1, ..., |R|$, si la arista i está conectada con el receptor r, la entrada (i, j) será $b^{(r)}_{i,j}$ y 0 en cualquier otro caso. Así las filas nulas de $B^{(r)}$ corresponden a las aristas que no están conectadas con el receptor r.

Se ilustra a continuación con un ejemplo el proceso de network coding, y las matrices asociadas al problema.

Ejemplo 1. La red de la Figura 1.2a es la red mariposa, y es el ejemplo más representativo y simple del uso del network coding. Se quiere enviar los mensajes X_1 y X_2 a los receptores r_1 y r_2. El mensaje X_1 es enviado a través de la arista 1, y reenviado por el nodo v_1; mientras que X_2 viaja por la arista 2, y el nodo v_2 lo reenvía. El problema surge cuando confluyen ambos mensajes X_1 y X_2 en el mismo nodo v_3, y de este solo sale una arista por la que puede pasar un solo mensaje a la vez. Si decidimos que el nodo transmita X_1, el nodo v_4 retransmitirá dicho mensaje por las aristas 8 y 9 hasta los receptores. Así, al receptor r_2 llegarán los mensajes deseados por las aristas 9 y 7, sin embargo el receptor r_1 no recibirá ninguna información sobre X_2. De manera análoga, si decidimos que sea X_2 el mensaje transmitido por v_3, no llegará a r_2 información sobre X_1. En cualquier caso, la comunicación de una sola vez no es posible, pues no todos los receptores reciben todos los mensajes.
Es aquí donde la codificación de red ofrece la solución: combinar los mensajes X_1 y X_2 en el nodo v_3 para formar el mensaje $X_1 + X_2$. De esta forma, como se muestra en Figura 1.2b al receptor r_1 llegarán X_1 vía la arista 5 y $X_1 + X_2$ vía la arista 8, por lo que puede recuperar X_2 restando ambos mensajes $(X_1 + X_2) - X_1 = X_2$. Lo mismo sucede en el receptor r_2, al que llegan los mensajes X_2 por la arista 7 y $X_1 + X_2$ por la arista 9, pudiendo recuperar el mensaje X_1 calculando $(X_1 + X_2) - X_2 = X_1$.

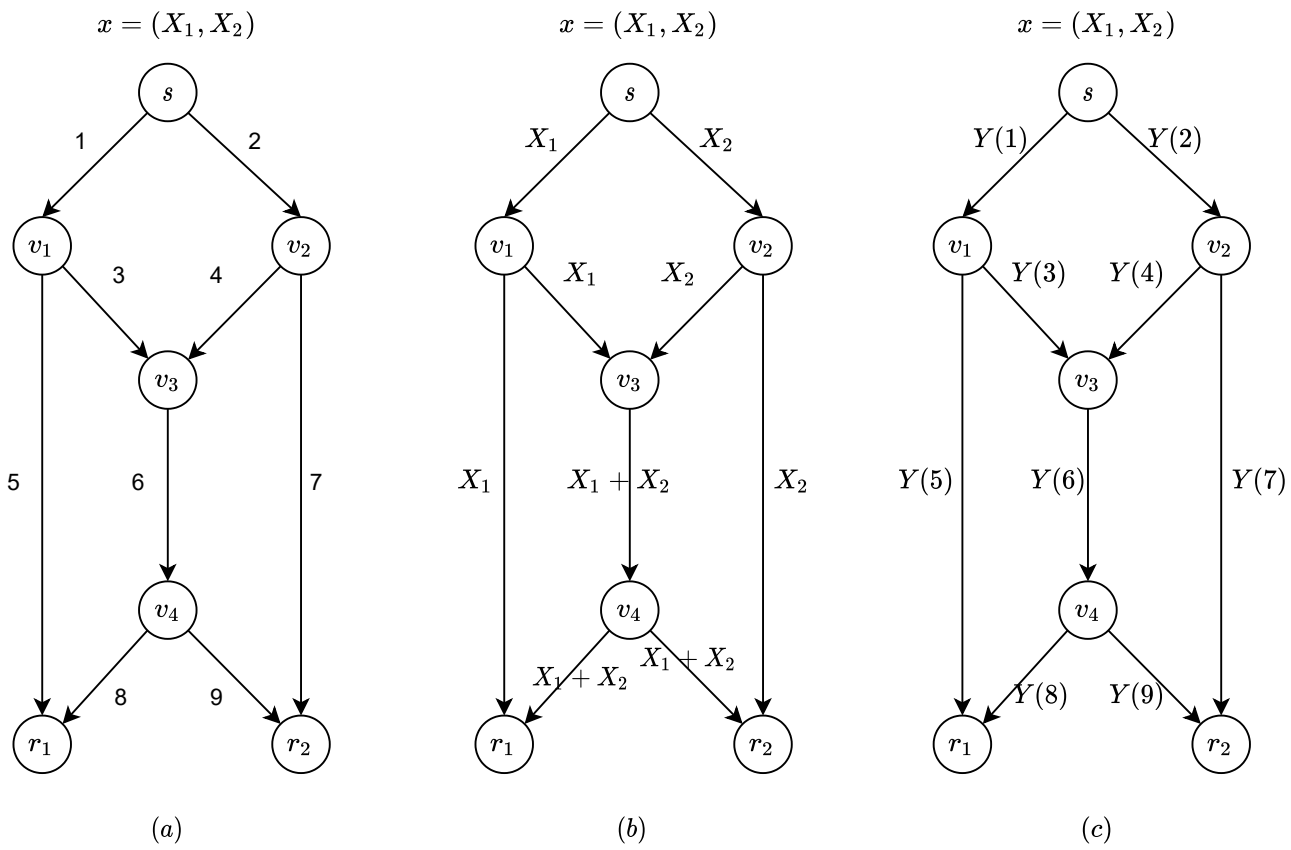


Figura 1.2: Red mariposa

Sobre este problema, considerado en el cuerpo $\mathbb{F}_2 = \{0, 1\}$, veamos cómo son los coeficientes de codificación y descodificación que construyen la solución explicada.
Para los coeficientes de codificación $a_{i,j}$, nos fijamos en qué aristas i salen del nodo emisor s, y qué mensaje viaja por cada una de ellas. Así, por la arista 1 se empieza a transmitir el mensaje X_1, luego $a_{1,1} = 1$; lo propio ocurre con la arista 2 y el mensaje X_2, luego $a_{2,2} = 1$. El resto de $a_{i,j}$ son nulos. Sobre los coeficientes $f_{i,j}$, serán no nulos aquellos en los que $j \in out(i)$. Observando la red tenemos

$$f_{1,3} = 1 \qquad f_{1,5} = 1 \qquad f_{2,4} = 1 \qquad f_{2,7} = 1 \qquad f_{3,6} = 1$$
$$f_{4,6} = 1 \qquad f_{6,8} = 1 \qquad f_{6,9} = 1 \qquad a_{1,1} = 1 \qquad a_{2,2} = 1$$

Y las matrices de codificación A de tamaño 2×9 $(= |h| \times |E|)$ y F de tamaño 9×9 $(= |E| \times |E|)$

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ a_{2,1} & a_{2,2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$F = \begin{bmatrix} 0 & 0 & f_{1,3} & 0 & f_{1,5} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & f_{2,4} & 0 & 0 & f_{2,7} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & f_{3,6} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & f_{4,6} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & f_{6,8} & f_{6,9} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Siguiendo la fórmula dada en (1.1), calculemos ahora las variables $Y(j)$ que representan la información que viaja por cada arista como se muestra en Figura 1.2c

$$\begin{aligned}
Y(1) &= a_{1,1}X_1 + a_{2,1}X_2 = 1 \cdot X_1 + 0 \cdot X_2 = X_1 \\
Y(2) &= a_{1,2}X_1 + a_{2,2}X_2 = 0 \cdot X_1 + 1 \cdot X_2 = X_2 \\
Y(3) &= f_{1,3}Y(1) = 1 \cdot X_1 = X_1 \\
Y(4) &= f_{2,4}Y(2) = 1 \cdot X_2 = X_2 \\
Y(5) &= f_{1,5}Y(1) = 1 \cdot X_1 = X_1 \\
Y(6) &= f_{3,6}Y(3) + f_{4,6}Y(4) = 1 \cdot X_1 + 1 \cdot X_2 = X_1 + X_2 \\
Y(7) &= f_{2,7}Y(2) = 1 \cdot X_2 = X_2 \\
Y(8) &= f_{6,8}Y(6) = 1 \cdot (X_1 + X_2) = X_1 + X_2 \\
Y(9) &= f_{6,9}Y(6) = 1 \cdot (X_1 + X_2) = X_1 + X_2
\end{aligned}$$

Esta información también puede ser escrita en forma de vectores, con los ya presentados vectores de codificación global, que en este problema tienen la forma

$$\begin{aligned}
c_1 &= \begin{pmatrix} 1 & 0 \end{pmatrix} & c_2 &= \begin{pmatrix} 0 & 1 \end{pmatrix} & c_3 &= \begin{pmatrix} 1 & 0 \end{pmatrix} \\
c_4 &= \begin{pmatrix} 0 & 1 \end{pmatrix} & c_5 &= \begin{pmatrix} 1 & 0 \end{pmatrix} & c_6 &= \begin{pmatrix} 1 & 1 \end{pmatrix} \\
c_7 &= \begin{pmatrix} 0 & 1 \end{pmatrix} & c_8 &= \begin{pmatrix} 1 & 1 \end{pmatrix} & c_9 &= \begin{pmatrix} 1 & 1 \end{pmatrix}
\end{aligned}$$

En cuanto a la descodificación, recordemos cómo se lleva a cabo en cada receptor r_t. En el receptor r_t se tiene que cumplir que $X_j = Z_j^{(r_t)} = \sum_{i \in in(r_t)} b_{i,j}^{(r_t)} Y(i)$. Explícitamente en nuestro ejemplo, en r_1 ha de ocurrir

$$\begin{aligned}
X_1 = Z_1^{(r_1)} &= b_{5,1}^{(r_1)}Y(5) + b_{8,1}^{(r_1)}Y(8) = b_{5,1}^{(r_1)}X_1 + b_{8,1}^{(r_1)}(X_1 + X_2) \\
X_2 = Z_1^{(r_1)} &= b_{5,2}^{(r_1)}Y(5) + b_{8,2}^{(r_1)}Y(8) = b_{5,2}^{(r_1)}X_1 + b_{8,2}^{(r_1)}(X_1 + X_2)
\end{aligned}$$

Y análogamente en r_2

$$\begin{aligned}
X_1 = Z_1^{(r_2)} &= b_{7,1}^{(r_2)}Y(7) + b_{9,1}^{(r_2)}Y(9) = b_{7,1}^{(r_2)}X_2 + b_{9,1}^{(r_2)}(X_1 + X_2) \\
X_2 = Z_1^{(r_2)} &= b_{7,2}^{(r_2)}Y(7) + b_{9,2}^{(r_2)}Y(9) = b_{7,2}^{(r_2)}X_2 + b_{9,2}^{(r_2)}(X_1 + X_2)
\end{aligned}$$

Al estar considerando el problema en $\mathbb{F}_2$, los coeficientes $b_{i,j}^{(r)}$ que resuelven estas ecuaciones se deducen fácilmente y estos son

$$\begin{aligned}
b_{5,1}^{(r_1)} &= 1 & b_{8,1}^{(r_1)} &= 0 & b_{5,2}^{(r_1)} &= 1 & b_{8,2}^{(r_1)} &= 1 \\
b_{7,1}^{(r_2)} &= 1 & b_{9,1}^{(r_2)} &= 1 & b_{7,2}^{(r_2)} &= 1 & b_{9,2}^{(r_2)} &= 0
\end{aligned}$$

Recopilando estos coeficientes de descodificación, se tienen las matrices $B^{(r_1)}$ y $B^{(r_2)}$ de tamaño 9×2 $(= |E| \times |h|)$

$$B^{(r_1)} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ b_{5,1} & b_{5,2} \\ 0 & 0 \\ 0 & 0 \\ b_{8,1} & b_{8,2} \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \qquad B^{(r_2)} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ b_{7,1} & b_{7,2} \\ 0 & 0 \\ b_{9,1} & b_{9,2} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 1 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}$$

Habiéndo definido ya las matrices que intervienen en el proceso de codificación lineal de una red de comunicación, veamos con más detalle la construcción de las matrices de transferencia $M^{(r)}$ dada en (1.5) y la ecuación (1.4).

Empecemos fijándonos en que podemos ver la matriz F como matriz de adyacencia del grafo G que representa la red. El número de caminos diferentes de longitud $n+1$ (llamamos longitud al número de aristas que contiene el camino) partiendo de la arista i para llegar a la arista j, es la entrada (i,j) de la matriz F^n para $n = 1, 2, ...$ (la demostración puede verse en [Ros]). Explícitamente

$$F^n_{i,j} = \sum_{\substack{(i=j_0,j_1,...,j_n=j) \\ \text{es un camino en G}}} f_{i=j_0,j_1} f_{j_1,j_2} \cdots f_{j_{n-1},j_n=j}$$

.

Como estamos considerando G acíclico, no existirán caminos en la red de longitud infinita. Esto implica que F es nilpotente y para $n \in \mathbb{N}$ suficientemente grande se cumple $F^n = 0$. Por tanto, la entrada (i,j) de la matriz

$$I + F + F^2 + \cdots + F^{N-1} = (I - F)^{-1}$$

nos proporciona información sobre todos los posibles caminos de i a j. Como la matriz F es triangular superior, $(I - F)$ también lo es, y es invertible. Ahora bien si damos valores a los coeficientes de codificación $a_{i,j}$ y $f_{i,j}$ se cumple

$$(Y(1), ..., Y(|E|)) = (X_1, ..., X_h)A(I + F + \cdots + F^{N-1}) \tag{1.6}$$

y fijando valores para los coeficientes de decodificación $b^{(r)}_{i,j}$, la salida que proporciona cada receptor r es, como vimos anteriormente en (1.2)

$$(Z^{(r)}_1, ..., Z^{(r)}_h) = (Y(1), ..., Y(|E|)B^{(r)}$$

así, hilando toda esta información tenemos

$$\begin{aligned}(Z^{(r)}_1, ..., Z^{(r)}_h) &= (Y(1), ..., Y(|E|))B^{(r)} \\ &= (X_1, ..., X_h)A(I + F + \cdots + F^{N-1})B^{(r)} \\ &= (X_1, ..., X_h)A(I - F)^{-1}B^{(r)}\end{aligned}$$

O lo que es lo mismo, coherente con (1.4)

$$(Z^{(r)}_1, ..., Z^{(r)}_h) = (X_1, ..., X_h)M^{(r)} \tag{1.7}$$

Observación. Habiendo visto cómo se construyen las matrices que determinan la comunicación en una red, y cómo se ha definido el vector de codificación global de una arista (este es $c_j = (c_{1,j}, ..., c_{h,j})$ si $Y(j) = c_{1,j}X_1 + c_{2,j}X_2 + \cdots + c_{h,j}X_h$), es interesante notar que el vector de codificación global de una arista j es precisamente la columna j-ésima de la matriz $A(I - F)^{-1}$. Esto lo podemos ver fácilmente fijándonos en (1.6), pues se aprecia que cada $Y(j)$ es el producto del vector mensaje $(X_1, \ldots, X_h)$ por la columna j-ésima de $A(I - F)^{-1}$, lo que coincide con la definición de vector de codificación global.

No olvidemos que nuestro objetivo es conseguir que la comunicación se lleve a cabo de manera correcta, y para ello debemos ser capaces de que los receptores, al decodificar la información recibida, reproduzcan exactamente la información original enviada por el emisor. Esto se traduce en que el vector mensaje $x = (X_1, ..., X_h)$ y los vectores decodificados $z^{(r)} = (Z_1^{(r)}, ..., Z_h^{(r)})$ han de coincidir para todo $r \in R$. A la vista de (1.7), si pedimos $M^{(r)} = I$ para cada r, es claro que obtenemos una comunicación exitosa, sin embargo, como vemos a continuación es suficiente con que $det(M^{(r)}) \neq 0$ para cada receptor r. Este último requisito, que podemos escribir de forma equivalente

$$\prod_{r \in R} detM^{(r)} \neq 0 \tag{1.8}$$

es válido, pues si consideramos unos coeficientes de codificación y decodificación $a_{i,j}$, $f_{i,j}$ y $b_{i,j}^{(r)}$ que lo verifiquen, entonces podemos sustituir los $b_{i,j}^{(r)}$ por unos nuevos coeficientes $\tilde{b}_{i,j}^{(r)}$ para que se cumpla $M^{(1)} = \cdots = M^{(|R|)]} = I$. Esto es, dados unos coeficientes $(a, f, b^{(r)})$ tales que para cada $r \in R$ se cumpla $detM^{(r)} \neq 0$, considérese la matriz $\widetilde{B}^{(r)} = B^{(r)}(M^{(r)})^{-1}$, bien definida pues $M^{(r)}$ posee inversa. Entonces si se toman como nuevos coeficientes de descodificación las entradas de dicha matriz, la nueva matriz de transferencia correspondiente a los coeficientes $(a, f, \tilde{b}^{(r)})$ verificará $\widetilde{M}^{(r)} = A(I - F)^{-1}\widetilde{B}^{(r)} = M^{(r)}(M^{(r)})^{-1} = I$, lo que se traduce en una comunicación exitosa.

Cuando los $a_{i,j}$, $f_{i,j}$ y $b_{i,j}^{(r)}$ son desconocidos, podemos pensar en ellos como variables, en cuyo caso el polinomio $\prod_{r \in R} detM^{(r)}$ es llamado *polinomio de transferencia*, y tiene una relevancia vital en la resolubilidad del problema de comunicación.

Como hemos visto, es de gran importancia determinar si $\prod_{r \in R} detM^{(r)}$ es nulo o no. Por ello, es de utilidad introducir la denominada *matriz de Edmond* $E^{(r)}$ para un receptor r, que es una matriz $(h + |E|) \times (h + |E|)$ definida de la forma

$$E^{(r)} = \begin{bmatrix} A & 0 \\ I - F & B^{(r)} \end{bmatrix}$$

La siguiente proposición muestra la relación entre el determinante de la matriz de Edmond y el determinante de la matriz de transferencia de un receptor.

Proposición 2. *Considérese una red de comunicación cuyas matrices de codificación y descodificación son $(A, F, B^{(r)})$, sea h el número de mensajes a enviar, y $|E|$ el número de aristas en la red. Sea la matriz de transferencia $M^{(r)} = A(I - F)^{-1}B^{(r)}$ y la matriz de Edmond $E^{(r)}$ para cada $r \in R$, entonces se verifica*

$$detM^{(r)} = (-1)^{h(1+|E|)} detE^{(r)}$$

Demostración. Fijémonos para empezar en que se cumple

$$\begin{bmatrix} I & -A(I - F)^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A & 0 \\ I - F & B^{(r)} \end{bmatrix} = \begin{bmatrix} 0 & -A(I - F)^{-1}B^{(r)} \\ I - F & B^{(r)} \end{bmatrix}$$

La matriz $\begin{bmatrix} I & -A(I - F)^{-1} \\ 0 & I \end{bmatrix}$, al ser triangular superior con unos en la diagonal, tiene determinante 1, luego $detE^{(r)} = det \begin{bmatrix} 0 & -A(I - F)^{-1}B^{(r)} \\ I - F & B^{(r)} \end{bmatrix}$. Trabajemos en este último determinante para llegar al resultado querido, haciendo uso de las propiedades de los determinantes. Si denotamos por $(a_1, \ldots, a_{|E|}, b_1, \ldots, b_h)$ las columnas de la matriz, intercambiemos las $|E|$ primeras columnas por las h siguientes para conseguir $(b_1, \ldots, b_h, a_1, \ldots, a_{|E|})$. Primero cambiamos $a_{|E|}$ por b_1, luego $a_{|E|}$ por b_2, y continuamos el proceso hasta llegar a $(a_1, \ldots, a_{|E|-1}, b_1, \ldots, b_h, a_{|E|})$. Esto conlleva h

intercambios, y repitiéndolo para las columnas $a_{|E|-1}, \ldots, a_1$ llegamos a la configuración deseada en un total de $h|E|$ cambios. Teniendo en cuenta que cada intercambio de filas o columnas en una matriz cambia el signo de su determinante, y que el determinante de una matriz triangular por bloques (de tamaños adecuados) es el producto de los determinantes de las matrices de la diagonal, se tiene

$$\begin{aligned} det\begin{bmatrix} A & 0 \\ I-F & B^{(r)} \end{bmatrix} &= det\begin{bmatrix} 0 & -A(I-F)^{-1}B^{(r)} \\ I-F & B^{(r)} \end{bmatrix} \\ &= (-1)^{h|E|} det\begin{bmatrix} -A(I-F)^{-1}B^{(r)} & 0 \\ B^{(r)} & I-F \end{bmatrix} \\ &= (-1)^{h|E|} det(-A(I-F)^{-1}B^{(r)}) det(I-F) \\ &= (-1)^{h|E|}(-1)^{h} det(A(I-F)^{-1}B^{(r)}) det(I-F) \\ &= (-1)^{h(1+|E|)} detM^{(r)} \end{aligned}$$

Donde el último paso se ha dado sabiendo que $det(I-F) = 1$ pues F es triangular superior con ceros en la diagonal. □

Gracias a la anterior proposición, para saber si $\prod_{r\in R} detM^{(r)}$ se anula o no, es equivalente comprobar si lo hace $\prod_{r\in R} detE^{(r)}$. Nos fijamos en la siguiente observación que nos servirá más adelante.

Observación. Para cada $r \in R$ en el polinomio $detE^{(r)}$, cada uno de los $a_{i,j}$'s y $f_{i,j}$'s aparecen como mucho con exponente 1 (esto se puede ver si calculamos el determinante desarrollando por filas, por ejemplo, ya que cada coeficiente aparece en una sola entrada). Mientras que cada $b_{i,j}^{(r)}$ aparece únicamente en uno solo de los polinomios $detE^{(r)}$ para algún $r \in R$. Lo que implica que en el polinomio de transferencia, al ser producto de $|R|$ polinomios, los $a_{i,j}$'s y $f_{i,j}$'s aparecerán con exponente $|R|$ como mucho, y los $b_{i,j}^{(r)}$'s con exponente 1 como mucho.

Los valores de A, F y $B^{(r)}$ con $r \in R$ determinan un código de red lineal. Esto es, hallar una solución equivale a encontrar valores para $a_{i,j}$, $f_{i,j}$ y $b_{i,j}^{(r)}$ tales que $A(I-F)^{-1}B^{(r)} = I$.

Habiendo establecido ya un marco en el que se construye el problema de codificación de red lineal, podemos empezar a hacernos cuestiones sobre su solución, en qué condiciones podemos encontrarla o cómo construirla.

Con vistas a esto, conviene definir el concepto de flujo y sistema de flujos en una red.

Definición 3. *Dado un problema de codificación de red multidifusión, sea $r \in R$ y h el número de mensajes a transmitir. Un flujo F para r (de tamaño h) es un conjunto de h caminos disjuntos (esto es que no compartan aristas) que empiecen en el nodo emisor s y terminen en r.*

Nota. Aunque un flujo sea un conjunto de caminos, pensaremos en él como un conjunto de aristas, así una arista j pertenece a F si esta aparece en algún camino del flujo.

En las condiciones anteriores

Definición 4. *Un sistema de flujos $\mathcal{F} = \{F_1, \ldots, F_{|R|}\}$ (de tamaño h) es un conjunto de $|R|$ flujos (de tamaño h), donde F_i es un flujo para el receptor r_i con $i = 1, ..., |R|$.*

Usemos la red de la Figura 1.1 para ilustrar los flujos y sistema de flujos de un grafo . En Figura 1.3a se muestra un flujo de tamaño 3 para el receptor r_1 formado por los caminos$\{(1,8),(2,5,9,12),(3,7,10,14)\}$, y en 1.3b el flujo $\{(1,4,9,13),(2,6,10,15),(3,11)\}$ de tamaño 3 para el receptor r_2. Ambos flujos forman un sistema de flujos de la red.

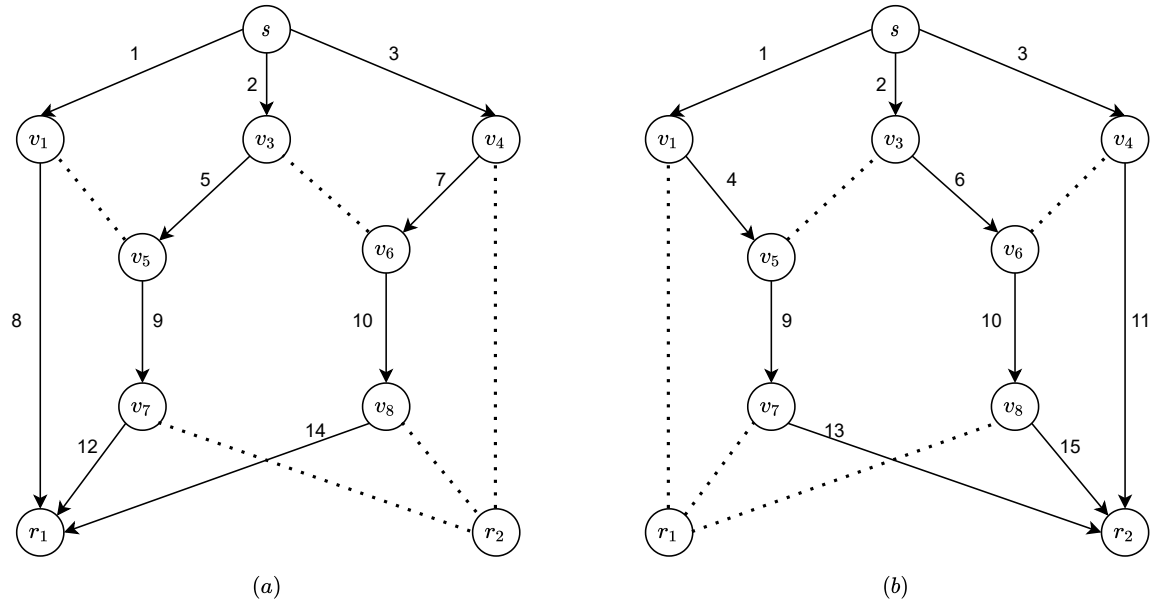


Figura 1.3

En teoría de grafos, una versión del teorema de Menger prueba la igualdad entre el tamaño del flujo máximo y el tamaño del corte mínimo (que es, sin entrar en detalles y en nuestro caso con aristas de igual capacidad, el mínimo número de aristas que hay que eliminar para desconectar emisor y receptor) entre dos vértices de un grafo como los que estamos considerando. En [LSRYC] haciendo uso del teorema del flujo máximo-corte mínimo (una generalización del teorema de Menger), se establece que, usando codificación de red, el número de mensajes que puede recibir un receptor coincide con el flujo máximo para ese receptor (y con el corte mínimo).

Es obvio que una condición necesaria para que nuestro problema sea resoluble, es la existencia en la red de un sistema de flujos. Debe existir un flujo para cada receptor, pues si no es así no habría caminos suficientes por los que transmitir todos los mensajes, se formarían cuellos de botella, y los receptores no recibirían información suficiente para llevar a cabo la decodificación.

En esta línea, mostramos a continuación los principales resultados que relacionan los flujos de una red, el polinomio de transferencia, y la resolubilidad del problema.

Necesitaremos el siguiente lema

Lema 5. *Sea f un polinomio no nulo en las variables $x_1, x_2, ..., x_n$ sobre $\mathbb{F}_q$, y sea d el máximo grado de f con respecto a cualquiera de las variables. Entonces, existen valores $a_1, a_2, ..., a_n \in \mathbb{F}_q$ tales que $f(a_1, a_2, ..., a_n) \neq 0$, para cualquier m tal que $q = p^m > d$.*

Demostración. Razonaremos por inducción sobre el número de variables. Para $n = 1$, $f \in \mathbb{F}_q[x_1]$ es un polinomio de una variable de grado d, luego posee como mucho d raíces en $\mathbb{F}_q$. Como $|\mathbb{F}_q| = q > d$ existirá un elemento $a_1 \in \mathbb{F}_q$ que no es raíz de f, como queríamos. Supongamos ahora que la proposición es cierta para $n-1$ variables y probémosla para n.
Consideremos f como un polinomio en las variables $x_1, ..., x_{n-1}$ con coeficientes en $\mathbb{F}_q[x_n]$. Como los coeficientes de f son polinomios de grado como mucho d, no son divisibles por $x_n^q - x_n$ (cuyas raíces son todos los elementos de $\mathbb{F}_q$) pues $q > d$. Así pues, existe algún elemento $a_n \in \mathbb{F}_q$ tal que f no se anula cuando $x_n = a_n$. Concluyendo, $f(x_1, \ldots, x_{n-1}, a_n)$ es un polinomio de $n-1$ variables, y por hipótesis existen $a_1, a_2, ..., a_{n-1} \in \mathbb{F}_q$ tal que $f(a_1, a_2, ..., a_n) \neq 0$, como queríamos. □

Usamos el siguiente teorema referente al caso en el que se considera un único receptor en la red,

para facilitarnos el camino hacia el caso con varios receptores. En su demostración nos servimos del algoritmo de Ford-Fulkerson, que sirve para encontrar un flujo máximo (una solución) en una red con un emisor y un receptor, tal que todos los coeficientes $a_{i,j}$'s y $f_{i,j}$'s son unos y ceros.

Teorema 6. *Sea r el nodo receptor de la red y h el número de mensajes a transmitir. Entonces existe un flujo para r de tamaño h en la red, si y solo si, el determinante de la matriz de transferencia $M^{(r)}$ es no nulo en el anillo de polinomios $\mathbb{F}_q[a, f, b^{(r)}]$.*

Demostración. Si existe un flujo en la red, se puede usar el algoritmo de Ford-Fulkerson para encontrar dicho flujo, el cual proporciona una solución para la que $M^{(r)} = I$, así $detM^{(r)} \neq 0$. Por otro lado, si $detM^{(r)}$ no es nulo, por el Lema 5 existen valores para $(a, f, b^{(r)})$ sobre un cuerpo finito suficientemente grande, para que esto sea así. Luego como ya comentamos anteriormente $\widetilde{B}^{(r)} = B^{(r)}(M^{(r)})^{-1}$ cumple $A(I - F)^{-1}\widetilde{B}^{(r)} = I$, y $(a, f, \tilde{b}^{(r)})$ es una solución para el problema de codificación de red lineal, por lo que ha de existir al menos un flujo en la red ya que es resoluble. □

El resultado central sobre codificación de red lineal multidifusión que venimos persiguiendo establece que en una comunicación en la que podamos transmitir h mensajes entre el nodo emisor y cada receptor de manera individual, entonces usando codificación de red lineal también será posible la comunicación entre el emisor y todos los receptores de manera simultánea. El siguiente teorema reúne lo mostrado anteriormente, y representa una pieza de gran importancia en la resolubilidad del problema.

Teorema 7. *Consideramos un problema de codificación de red multidifusión donde se quieren transmitir h mensajes al conjunto R de receptores. Las siguientes afirmaciones son equivalentes:*

1. *El problema de codificación de red es resoluble.*
2. *El polinomio de transferencia $\prod_{r\in R} detM^{(r)}$ asociado al problema de comunicación es no nulo*
3. *Existe un sistema de flujos de tamaño h en la red*

Demostración. Probaremos que 2) $\Leftrightarrow$ 3) y 1) $\Leftrightarrow$ 3) para cerrar el ciclo de implicaciones.
2) $\Rightarrow$ 3): Si el polinomio de transferencia $\prod_{r\in R} detM^{(r)}$ es no nulo, en particular cada $detM^{(r)}$ con $r \in R$ será no nulo y por el Teorema 6 existirá un flujo de tamaño h para cada receptor r, $|R|$ en total, esto es, un sistema de flujos.
3) $\Rightarrow$ 2): Si existe un sistema de flujos $\mathcal{F} = \{F_1, \ldots, F_{|R|}\}$ basta aplicar de nuevo el Teorema 6, y para cada flujo F_i el determinante de la matriz de transferencia asociada $M^{(i)}$ será no nulo, por lo que el producto de todas ellas $\prod_{r\in R} detM^{(r)}$ tampoco se anulará.
1) $\Rightarrow$ 3): Es evidente, pues ya explicamos que si un problema es resoluble, ha de existir un flujo para cada receptor r, ya que sino faltaría información para llevar a cabo la decodificación.
3) $\Rightarrow$ 1): Acabamos de ver que si existe un sistema de flujos, el polinomio de transferencia asociado a la red no es idénticamente nulo. Ya observamos que el máximo grado que puede tener una variable en el polinomio de transferencia es $|R|$, luego por el Lema 5 existen valores $(a, f, b^{(r)})$ en un cuerpo finito suficientemente grande $\mathbb{F}_q(q = p^m \geq |R|)$ que no anulan dicho polinomio. Esto ofrece una solución $(a, f, \tilde{b}^{(r)})$, pues cada receptor r puede multiplicar el vector $z^{(r)}$ que le llega por $(M^{(r)})^{-1}$ para recuperar el vector mensaje original $x = (X_1, \ldots, X_h)$. □

Corolario 8. *El máximo número de mensajes que se pueden transferir a la vez, es igual al mínimo de entre los máximos flujos para cada receptor.*

Antes de continuar avanzando, es buen momento para aclarar una cuestión que dejábamos pendiente en la introducción: la posibilidad de que existan varios nodos emisores en la red. Hasta ahora en

los conceptos y explicaciones dadas solo se ha considerado un solo emisor en el grafo, alegando que para los problemas con más de uno existe un problema equivalente reduciendo los emisores a uno único. Veamos cómo transformar un problema en otro, sin que esto altere su resolubilidad.

Sea una red G en la que los mensajes X_i se generan en varios nodos emisores s_t, como por ejemplo en la Figura 1.4. Para obtener la red equivalente con un solo emisor, se añaden un nuevo nodo s en el que pasan a generarse todos los mensajes, y una nueva arista que conecte s y s_t por cada mensaje que se generase originalmente en s_t. Además es necesario ajustar los coeficientes de codificación correspondientes a las nuevas aristas añadidas, de tal manera que el mensaje X_i se transmita por la arista i. Esto lo convierte en el tipo de problema que se ha venido tratando hasta ahora.

El paso de un problema a otro preserva la resolubilidad, esto es si la red con varios emisores en la que se fija los coeficientes de codificación es resoluble, entonces al transformarla en una con un solo emisor también es resoluble, y viceversa. Esto se puede justificar en virtud del Teorema 7 encontrando un sistema de flujos. Si en la red con varios emisores existe un sistema de flujos, el sistema de flujos correspondiente en la red equivalente resultará de añadir a cada camino una de las nuevas aristas introducidas. El procedimiento inverso consiste en eliminar de cada camino en cada flujo del sistema de flujos, la arista que conecta el camino con el nodo emisor.

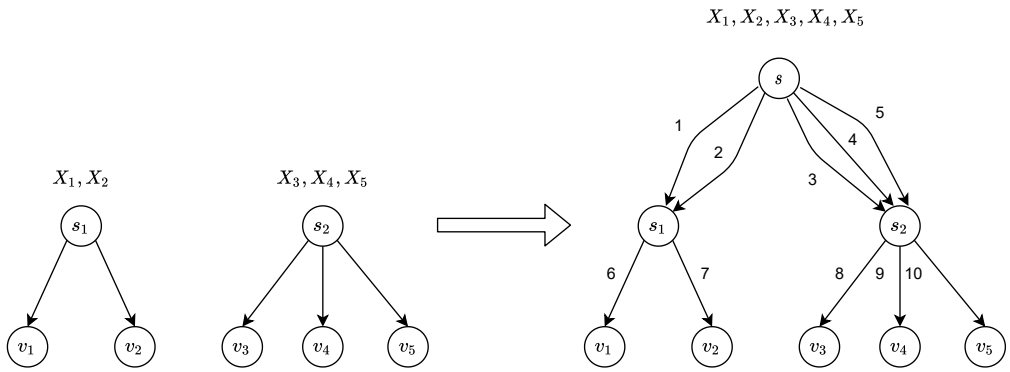


Figura 1.4: Transformación de una red con varios emisores en la equivalente con un solo emisor

Queda visto que si tenemos un problema de comunicación resoluble, entonces existen valores para los coeficientes de codificación y decodificación $(a, f, b^{(r)})$ para todo $r \in R$ que no anulan al polinomio de transferencia, en un cuerpo finito $\mathbb{F}_q$ con $q > |R|$. De hecho basta usar codificación de red lineal sobre dicho cuerpo para que la transmisión de información sea máxima (la dada por el flujo máximo) como se puede ver en [LSRYC].

1.2. Algoritmo de Jaggi-Sanders

A continuación describimos un conocido algoritmo determinista introducido en [JaSa] que resuelve el problema de network coding sobre un cuerpo finito $\mathbb{F}_q$ con $q \geq |R|$. Este lo consigue en tiempo polinomial encontrando coeficientes de codificación adecuados.

El algoritmo proporciona una solución si el problema es resoluble, o equivalentemente si existe un sistema de flujos en la red. Ya se comentó anteriormente que existen algoritmos suficientemente rápidos para encontrar flujos en una red, y por lo tanto sistemas de flujos.

Podemos dividir el algoritmo esencialmente en dos partes, la inicialización y la búsqueda de los coeficientes.

En la primera parte la red es modificada añadiendo un nuevo nodo emisor y h nuevas aristas, exactamente como se ha detallado líneas más arriba al transformar una red multiemisor en una

con un solo emisor. Se considera un sistema de flujos en la red $\mathcal{F} = (F_1, \ldots, F_{|R|})$ (si no existe, el problema no tiene solución) siendo F_i un flujo para el receptor r_i. Considerando la red ya modificada, se empiezan a fijar coeficientes de la siguiente manera. Si una arista e no aparece en $\mathcal{F}$ se fija $f_{i,e} = 0$ para todo $i \in in(e)$. Los coeficientes de codificación de las h nuevas aristas introducidas se determinan de tal manera que sus vectores de codificación global sean

$$(1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, \ldots, 1) \tag{1.9}$$

La segunda parte del algoritmo funciona como se explica a continuación. Para cada $r \in R$ inicializamos los conjuntos $C_1 = \cdots = C_{|R|}$ que contienen cada uno las h nuevas aristas y $B_1, \ldots, B_{|R|}$ que contienen sus correspondientes vectores de codificación global (es decir los vectores en (1.9)). Dichos conjuntos se irán actualizando siguiendo ciertas restricciones, que han de cumplirse en todo momento. Para $i = 1, \ldots, |R|$, se requiere que C_i contenga una arista de cada camino disjunto que compone el flujo F_r, mientras que el conjunto B_i de vectores de codificación global correspondientes a C_i ha de generar el espacio $\mathbb{F}_q^h$. Cabe resaltar que los conjuntos C_i y B_i son conjuntos ordenados, por lo que los escribiremos entre paréntesis.
La actualización se realiza como sigue: siguiendo el orden ancestral elegido para la red, desde el nodo emisor hacia los receptores, se visitan las aristas $j \in \mathcal{F}$. Considerando una arista j, nos interesan los receptores r_i tales que $j \in F_i$, y procedemos de la siguiente manera

- Añadimos j a C_i
- Eliminamos de C_i la arista predecesora a j en F_i. Esta es, la única arista $k \in C_i \cap in(j)$ para la que k y j son parte de un camino en el flujo F_i.

Si $l_1, \ldots, l_k$ son los valores de l para los que C_l se ha actualizado, y llamamos $i_1, \ldots, i_k$ a las aristas que se han eliminado de los conjuntos $C_{l_1}, \ldots, C_{l_k}$ el siguiente paso consiste en

- Asignar los coeficientes de codificación $f_{i_1,j}, \ldots, f_{i_k,j}$ de tal manera que todos los conjuntos de vectores de codificación global $B_{l_1}, \ldots, B_{l_k}$ asociados a $C_{l_1}, \ldots, C_{l_k}$ generen $\mathbb{F}_h^q$
- Si la arista j solo tiene un predecesor, digamos p, el único coeficiente a asignar es $f_{p,j}$, y basta fijar $f_{p,j} = 1$. Esto provoca que se tenga la igualdad $c_p = c_j$. El significado práctico es que el nodo que conecta p y j simplemente reenvía el paquete que le llega, sin realizar combinaciones.

Siguiendo los pasos recién explicados, tras haber recorrido todas las aristas, se cumplirá para cada $r \in R$, que C_r contiene las aristas entrantes a r, esto es, $C_r = in(r) \cap F_r$. Además, los vectores de codificación global de las aristas entrantes a cada receptor (es decir los conjuntos B_r) forman una base de $\mathbb{F}_q^h$, por lo que los coeficientes que se han ido eligiendo forman parte de una solución, y tenemos información suficiente para completarla resolviendo un sistema de ecuaciones lineales como veremos al comienzo de la siguiente sección.

Una condición necesaria para el funcionamiento del algoritmo que se suma a la de la existencia de un sistema de flujos, es que el tamaño del cuerpo $\mathbb{F}_q$ tiene que ser suficientemente grande, concretamente $q \geq |R|$. Si esto no se cumple, es posible que el algoritmo falle al intentar encontrar coeficientes de codificación válidos.

Una forma de asignar los coeficientes de codificación es hacerlo de forma aleatoria, e ir comprobando si los vectores de codificación global resultantes son adecuados. La probabilidad de que este método resulte exitoso no solo es positiva, sino suficientemente alta si se considera un cuerpo de tamaño no demasiado pequeño. El siguiente lema nos ayuda a saber cuántas elecciones no válidas puede haber para los coeficientes de codificación.

Lema 9. *Sea una base* $\{b_1, \ldots, b_h\}$ *de* $\mathbb{F}_q^h$ *y* $c \in \mathbb{F}_q^h$*, entonces existe una única elección de* $a \in \mathbb{F}_q$ *tal que*

$$c + ab_h \in Span_{\mathbb{F}_q}\{b_1, \ldots, b_{h-1}\} \tag{1.10}$$

Demostración. Expresando c en la base dada se tiene $c = c_1b_1 + \cdots + c_hb_h$, y se deduce que el único valor de a para que se cumpla (1.10) es $a = -c_h$. Así $c - c_hb_h = c_1b_1 + \cdots c_{h-1}b_{h-1} \in Span_{\mathbb{F}_q}\{b_1, \ldots, b_{h-1}\}$ □

Así, considerada una arista j, podemos aplicar el lema 9 para cada i_t y base B_{l_t} (justo antes de la actualización, es decir, antes de añadir j a C_{l_t} y eliminar i_t), siendo $b_h = c_{i_t}$ y

$$c = \sum_{l \in \{i_1, \ldots, i_k\} \setminus \{i_t\}} f_{l,j}c_l.$$

Por lo que en el paso en el que se buscan los coeficientes de codificación necesarios para calcular c_j, dados todos los coeficientes excepto $f_{i_t,j}$, hay únicamente una opción no válida para elegir dicho coeficiente para en un flujo concreto.

La probabilidad de éxito entonces en una actualización en el algoritmo, denotando $k' = |\{i_1, \ldots, i_k\}| = |(in(j) \cap \mathcal{F}|$ es al menos

$$\frac{q^{k'} - kq^{k'-1} + (k-1)}{q^{k'}} = 1 - \frac{k}{q} + \frac{k-1}{q^{k'}} \tag{1.11}$$

Esta cota procede de situarse en el peor de los casos que se explica a continuación. Hay k' coeficientes que asignar, luego $q^{k'}$ posibilidades; veamos cuántas de ellas no son válidas. Para cada combinación de $k' - 1$ coeficientes fijados, hay una sola mala elección para el coeficiente restante (para cada base B_{l_t} afectada por dicho coeficiente). Como hay $q^{k'-1}$ combinaciones posibles, y k bases de las que preocuparse al asignar coeficientes, entonces hay como mucho $kq^{k'-1}$ malas elecciones. En esta situación se cuenta la elección (obviamente no válida) $f_{i_1,j} = \cdots = f_{i_k,j} = 0$, un total de k veces, cuando dicha elección es común a todos los receptores, y basta contarla una vez, lo que reduce el número de malas elecciones a $kq^{k'-1} - k + 1$. Así, existen como mínimo $q^{k'} - kq^{k'-1} - k + 1$ asignaciones válidas.
La expresión en (1.11) es positiva para $q \geq k$, y como $k \leq |R|$, entonces para un problema de network coding resoluble es suficiente usar el cuerpo $\mathbb{F}_q$ si $q \geq |R|$.

Se muestra a continuación el pseudocódigo del algoritmo.

Entrada: Grafo dirigido acíclico $G = (V, E)$ con un orden ancestral definido, cuerpo finito $\mathbb{F}_q^h$ con $q \geq |R|$.
Salida: Coeficientes de codificación $a_{i,j}$ y $f_{i,j}$.
Inicialización:
Se modifica la red como se ha descrito en la explicación del algoritmo, siendo $\{e_1, \ldots, e_h\}$ las nuevas aristas. Se encuentra un sistema de flujos $\mathcal{F} = (F_1, \ldots, F_{|R|})$.
for $j \in E$ ***y*** $i = 1, \ldots, h$ **do**
  if $j \in out(e_i)$ *y* $e_{i,j} \notin \mathcal{F}$ **then**
    $a_{i,j} = 0$;
  end
end
for $(i, j) \in E \times E$ *con* $i \in in(j)$ *y* $(i, j) \notin \mathcal{F}$ **do**
  $f_{i,j} = 0$;
end
for $l = 1, \ldots, |R|$ **do**
  $C_l := (e_1, \ldots, e_h)$;
  $B_l := ((1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, \ldots, 1))$
end
Actualización:
for $e \in \mathcal{F}$ *siguiendo el orden ancestral* **do**
  Sean $(r_{l_1}, \ldots, r_{l_k})$ los receptores tales que e está en los flujos $F_{l_1} \ldots, F_{l_k}$;
  Sean $(p_1, \ldots, p_k)$ los predecesores correspondientes a e en $F_{l_1} \ldots, F_{l_k}$;
  for $i = 1, \ldots, k$ **do**
    $C_{l_i} := (C_{l_i} \backslash \{p_i\}) \cup \{e\}$;
  end
  Encontrar coeficientes de codificación $(f_{p_1,e}, \ldots, f_{p_k,e})$ para que B_{l_i} genere $\mathbb{F}_q^h$ para todo $i = 1, \ldots, k$;
  for $i = 1, \ldots, k$ **do**
    if p_i *es alguna de las aristas* e_s *para algún* $s \in \{1, \ldots, h\}$ **then**
      nombrar $a_{s,e}$ a $f_{p_i.e}$;
    end
  end
end
devolver los $a_{i,j}, f_{i,j}$ que se han asignado

Algoritmo 1: Algoritmo Jaggi-Sanders

Veamos cómo funcionaría el algoritmo en un ejemplo concreto.

Ejemplo 10. Sea la red representada en la Figura 1.1 en la que se envían a los receptores r_1, r_2 los mensajes $X_1, X_2, X_3 \in \mathbb{F}_2$. El primer paso consiste en modificar la red añadiendo un nuevo emisor que envíe los mensajes a través de las nuevas aristas $\{e_1, e_2, e_3\}$ (en realidad este paso es solo necesario para las redes que inicialmente tengan más de un emisor). Un sistema de flujos incluyendo las nuevas aristas es $\mathcal{F} = \{F_1, F_2\}$

$$F_1 = \{(e_1, 1, 8), (e_2, 2, 5, 9, 12), (e_3, 3, 7, 10, 14)\}$$
$$F_2 = \{(e_1, 1, 4, 9, 13), (e_2, 2, 6, 10, 15), (e_3, 3, 11)\}$$

Empezamos tomando los primeros coeficientes de codificación $a_{1,1} = a_{2,2} = a_{3,3} = 1$ y $a_{1,2} = a_{1,3} = a_{2,1} = a_{2,3} = a_{3,1} = a_{3,2} = 0$ (el resto de $a_{i,j}$ también nulos) que se corresponde con establecer como primeros vectores de codificación global $c_1 = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}, c_2 = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}, c_1 = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}$.

Si existiesen pares (i, j) de aristas tal que $(i, j) \notin \mathcal{F}$, se fijaría $f_{i,j} = 0$, en nuestro caso no los

hay. Además, como las aristas $4, 5, 6, 7, 8, 11, 12, 13, 14, 15$ solo tienen un predecesor, fijamos $f_{1,4} = f_{2,5} = f_{2,6} = f_{3,7} = f_{1,8} = f_{3,11} = f_{9,12} = f_{9,13} = f_{10,14} = f_{10,15} = 1$.
Los conjuntos que iremos actualizando son, al iniciar:

$$C_1 = (e_1, e_2, e_3) \qquad B_1 = ((1,0,0),(0,1,0),(0,0,1))$$
$$C_2 = (e_1, e_2, e_3) \qquad B_2 = ((1,0,0),(0,1,0),(0,0,1))$$

Ahora siguiendo el orden ancestral en la red, tras cada actualización los conjuntos son

$$C_1 = (e_1, e_2, e_3) \to (1, e_2, e_3) \to (1, 2, e_3) \to (1, 2, 3) \to (1, 5, 3) \to (1, 5, 7) \to (8, 5, 7) \to$$
$$\to (8, 9, 7) \to (8, 9, 10) \to (8, 12, 10) \to (8, 12, 14)$$
$$C_2 = (e_1, e_2, e_3) \to (1, e_2, e_3) \to (1, 2, e_3) \to (1, 2, 3) \to (4, 2, 3) \to (4, 6, 3) \to (9, 6, 3) \to$$
$$\to (9, 10, 3) \to (9, 10, 11) \to (13, 10, 11) \to (13, 15, 11)$$

Los coeficientes que se asignarían de forma aleatoria en el algoritmo son los correspondientes a las aristas 9 y 10, es decir $f_{4,9}, f_{5,9}, f_{6,10}$ y $f_{7,10}$. Las etapas clave son entonces las actualizaciones que involucran dichas aristas. Para la arista 9, $(8, 5, 7) \to (8, 9, 7)$ y $(4, 6, 3) \to (9, 6, 3)$, y para la arista 10, $(8, 9, 7) \to (8, 9, 10)$ y $(9, 6, 3) \to (9, 10, 3)$. En la actualización de la arista 9, los vectores de codificación global son $B_i = ((1,0,0),(0,1,0),(0,0,1))$ con $i = 1, 2$, y para conservar una base necesariamente se tiene que fijar $f_{4,9} = f_{5,9} = 1$. En la actualización de la arista 10 (que es de hecho la siguiente a la 9) los vectores de codificación global quedan $B_i = ((1,0,0),(1,1,0),(0,0,1))$ con $i = 1, 2$, obligándonos a asignar $f_{6,10} = f_{7,10} = 1$. Estas elecciones son las que forman parte de una solución al problema, pues consiguen que las aristas $in(r_1) = \{8, 12, 14\}$ y $in(r_2) = \{13, 15, 11\}$ lleven una base $((1,0,0),(1,1,0),(0,1,1))$ y $((1,1,0),(0,1,1),(0,0,1))$ respectivamente) de $\mathbb{F}_2^3$ a los receptores tal como muestra la Figura 1.5

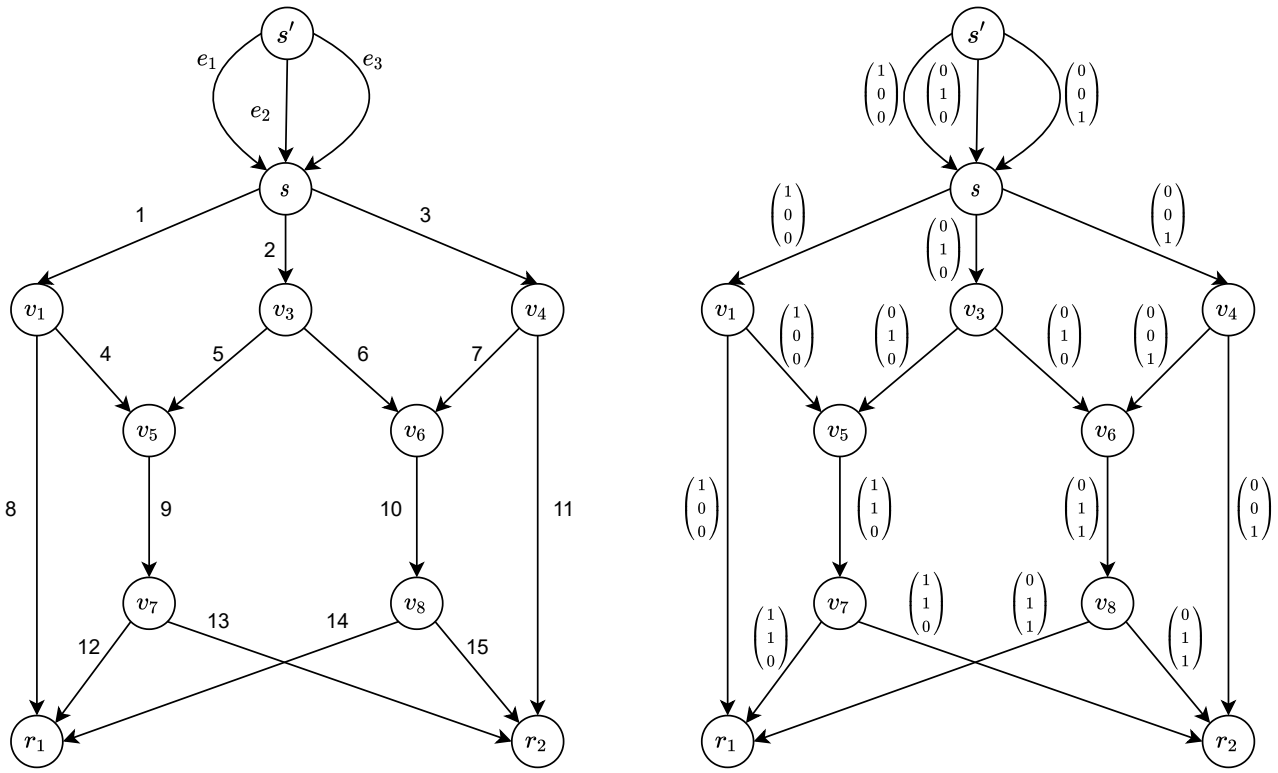


Figura 1.5

1.3. Random Network Coding

Consideremos un problema de codificación de red resoluble. Acabamos de comprobar que en un cuerpo finito suficientemente grande es posible encontrar coeficientes que resuelvan el problema. Lo que se propone con la codificación aleatoria, tal y como explican en [GeTh] y [Casp], es fijar los valores de un subconjunto (puede ser vacío) de coeficientes de codificación $a_{i,j}$'s y $f_{i,j}$'s de manera que no eviten que la comunicación sea exitosa (es decir, que exista alguna elección para el resto de coeficientes que completen una solución), y el resto de coeficientes asignarlos de manera aleatoria. Esto es conocido como codificación de red aleatoria o random network coding.

Aplicando este tipo de codificación cabe preguntarnos cuándo produce una solución al problema de comunicación y cómo de probable es que esto ocurra. Sobre ambas cuestiones escribimos a continuación.

Recuérdese cómo definíamos en (1.3) el vector de codificación global de una arista j, este es $c_j = (c_{1,j}, ..., c_{h,j})$ si $Y(j) = c_{1,j}X_1 + c_{2,j}X_2 + \cdots + c_{h,j}X_h$. Fijemos unos coeficientes de codificación en un grafo, resulta que si para cada receptor $r \in R$ los vectores de codificación global de las aristas en $in(r)$ forman una base del espacio $\mathbb{F}_q^h$, entonces los coeficientes de codificación elegidos son válidos para una solución del problema, y de la siguiente manera podemos encontrar los coeficientes de descodificación $b_{i,j}^{(r)}$ que completen la solución.
Para cada receptor $r \in R$ llamemos matriz global de codificación, denotada por $C^{(r)}$, a la matriz cuadrada cuyas columnas son los vectores de codificación global de las aristas $\{i_1^r, i_2^r, \ldots, i_h^r\} = in(r)$ entrantes en r. Así, denotando por $Y^{(r)} \in \mathbb{F}_q^{in(r)}$ al vector cuyas entradas son las variables que llegan a r, tenemos

$$Y^{(r)} = (Y(i_1^r), Y(i_2^r), \ldots, Y(i_h^r)) = (X_1, X_2, \ldots, X_h) \begin{bmatrix} c_{1,i_1^r} & c_{1,i_2^r} & \cdots & c_{1,i_h^r} \\ c_{2,i_1^r} & c_{2,i_2^r} & \cdots & c_{2,i_h^r} \\ \vdots & \vdots & \ddots & \vdots \\ c_{h,i_1^r} & c_{h,i_2^r} & \cdots & c_{h,i_h^r} \end{bmatrix} = (X_1, X_2, \ldots, X_h)C^{(r)}$$

Como suponemos que llega una base de $\mathbb{F}_q^h$ a r, la matriz $C^{(r)}$ es no singular y su inversa será la que determine los coeficientes de descodificación buscados, pues multiplicando por $(C^{(r)})^{-1}$ por la derecha se tiene

$$(X_1, X_2, \ldots, X_h) = (Y(i_1^r), Y(i_2^r), \ldots, Y(i_h^r)(C^{(r)})^{-1} = Y^{(r)}(C^{(r)})^{-1}$$

Sabemos que las únicas filas no nulas de la matriz de descodificación $B^{(r)}$ son las correspondientes a las aristas $\{i_1^r, i_2^r, \ldots, i_n^r\}$, luego basta tomarlas iguales a las filas de $(C^r)^{-1}$ para que la descodificación

$$(Z_1^r, Z_2^r, \ldots, Z_h^r) = (Y(1), Y(2), \ldots, Y(|E|)B^{(r)} = (X_1, X_2, \ldots, X_h)$$

sea exitosa. Así, conocidas las variables $Y(j)$ y los vectores de codificacicación de una red, podemos determinar si es posible o no resolver el problema.

Veamos una pequeña muestra de lo recién comentado, utilizando el Ejemplo 1 de la red mariposa.

Ejemplo 11. Volviendo la vista al ejemplo de la red mariposa, nos encontrábamos con dos receptores r_1 y r_2 a los que se quería hacer llegar un mensaje $x = (X_1, X_2)$. Observando la red, se tiene $in(r_1) = \{5, 8\}$ y $in(r_2) = \{7, 9\}$. Ya calculamos las variables $Y(j)$ y los vectores de codificación global, los que nos interesan son

$$Y(5) = X_1 \qquad Y(8) = X_1 + X_2 \qquad Y(7) = X_2 \qquad Y(9) = X_1 + X_2$$
$$c_5 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad c_8 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad c_7 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \qquad c_9 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Así, construyendo las matrices $C^{(r_i)}$ se cumple

$$(Y(5), Y(8)) = (X_1, X_2)C^{(r_1)} = (X_1, X_2)\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$
$$(Y(7), Y(9)) = (X_1, X_2)C^{(r_2)} = (X_1, X_2)\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$$

Luego basta calcular las inversas (en $\mathbb{F}_2$) de $C^{(r_1)}$ y $C^{(r_2)}$ para despejar y obtener x de las ecuaciones anteriores, y a partir de sus filas, construir las matrices de descodificación $B^{(r_1)}$ y $B^{(r_2)}$

$$(C^{(r_1)})^{(-1)} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \Rightarrow B^{(r_1)} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \qquad (C^{(r_2)})^{(-1)} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \Rightarrow B^{(r_2)} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 1 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}$$

Si no sabemos cómo funciona la red, y solo nos son conocidas las variables $Y(j)$, podemos averiguar los vectores de codificación global de cada arista, siguiendo el proceso siguiente. Se envían sucesivamente como vector mensaje los vectores de la base canónica de $\mathbb{F}_q^h$

$$x \in \{(1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, 0, \ldots, 1)\}$$

De esta forma, para cada arista j la variable $Y(j)$ irá tomando los valores de las componentes de su vector de codificación global, pues al enviar cada vector mensaje se tendrá

$$\begin{aligned} Y(j) &= c_{1,j} \cdot 1 + c_{2,j} \cdot 0 + \cdots + c_{h,j} \cdot 0 = c_{1,j} \\ Y(j) &= c_{1,j} \cdot 0 + c_{2,j} \cdot 1 + \cdots + c_{h,j} \cdot 0 = c_{2,j} \\ &\vdots \\ Y(j) &= c_{1,j} \cdot 0 + c_{2,j} \cdot 0 + \cdots + c_{h,j} \cdot 1 = c_{h,j} \end{aligned}$$

y tras enviar los h vectores mensaje, ya conoceremos todas las entradas del vector de codificación $c_j = (c_{1,j}, c_{2,j}, \ldots, c_{h,j})$.

De igual manera, este método nos puede servir para averiguar la matriz de transferencia $M^{(r)}$ para cada receptor. Sabemos que la información que proporciona un receptor r como salida tras transmitir un mensaje x por la red, viene dada por el vector $z^{(r)} = xM^{(r)}$. Luego enviar por la red los vectores de la base canónica de $\mathbb{F}_q^h$, proporcionará sucesivamente las filas de la matriz $M^{(r)}$. Explícitamente

$$\begin{aligned} z^{(r)} &= (1, 0, \ldots, 0)M^{(r)} = m^1 \\ z^{(r)} &= (0, 1, \ldots, 0)M^{(r)} = m^2 \\ &\vdots \\ z^{(r)} &= (0, 0, \ldots, 1)M^{(r)} = m^h \end{aligned}$$

siendo m^i la i-ésima fila de $M^{(r)}$. Por lo que al finalizar los envíos, se podrá reconstruir la matriz de transferencia, al conocer sus filas. Como ya se comentó, conocer dicha matriz hace inmediata la descodificación, pues basta multiplicar por su inversa el mensaje codificado recibido para obtener el vector mensaje original $x = z^{(r)}(M^{(r)})^{(-1)}$.

El método explicado que consiste en introducir como vectores mensajes la base canónica de $\mathbb{F}_q^h$ de hecho es útil para afrontar el problema de comunicación siguiente. Si las matrices de transferencia están fijadas (esto es, coeficientes de codificación fijos) y son conocidas por los receptores, el problema de codificación de red lineal se dice que es *coherente*; en cualquier otro caso el problema se define como *no coherente*. Un ejemplo de este último, consiste en considerar un problema de random network coding, en el que además de elegirse coeficientes desconocidos al azar, estos no son fijos, sino que cambian cada vez. Como se ve a continuación, el método explicado anteriormente propone una forma de resolverlo:
Se propone enviar h vectores mensaje $p_1, \ldots, p_h$ a través de una red de la cual no se tiene información. El envío de cada vector mensaje produce en cada receptor $r \in R$ la salida $p'_i = p_i M^{(r)}$, y tras enviarlos todos se pueden recoger dichas salidas en la matriz $PM^{(r)}$ siendo los vectores mensaje p_i las filas de P. Ahora bien, no podemos recuperar los mensajes originales al no poseer información sobre las matrices $M^{(r)}$, las cuales ya no serán las mismas cuando se cambien los coeficientes de codificación la siguiente vez. Para solventar este problema, se introducen como cabecera de los p_i los vectores e_i de la base canónica de $\mathbb{F}_q^h$. De esta forma se envía

$$\left[\frac{e_1}{p_1}\right] = \left[\frac{1 \quad 0 \quad \cdots \quad 0}{p_{1,1} \quad p_{1,2} \quad \cdots \quad p_{1,h}}\right], \ldots, \left[\frac{e_h}{p_h}\right] = \left[\frac{0 \quad 0 \quad \cdots \quad 1}{p_{h,1} \quad p_{h,2} \quad \cdots \quad p_{h,h}}\right]$$

que recopilado en una matriz corresponde a $\left[\begin{smallmatrix} I_h \\ P \end{smallmatrix}\right]$, y cada receptor $r \in R$ recibe $\left[\begin{smallmatrix} M^{(r)} \\ PM^{(r)} \end{smallmatrix}\right]$. Esto es

$$\left[\begin{array}{cccc} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ \hline p_{1,1} & p_{1,2} & \cdots & p_{1,h} \\ \vdots & \vdots & \ddots & \vdots \\ p_{h,1} & p_{h,2} & \cdots & p_{h,h} \end{array}\right] M^{(r)} = \left[\frac{I_h}{P}\right] M^{(r)} = \left[\frac{M^{(r)}}{PM^{(r)}}\right]$$

por lo que los receptores conocerán las matrices $PM^{(r)}$ y $M^{(r)}$, y será posible recuperar los mensajes originales realizando la operación

$$PM^{(r)}(M^{(r)})^{-1} = P$$

Sigamos pensando en un problema resoluble sobre el que fijamos ciertos coeficientes que formen parte de una solución, y el resto se eligen de manera aleatoria. Podemos obtener cotas sobre la probabilidad con que la codificación aleatoria produce una solución. De la siguiente proposición deduciremos una de esas cotas.

Proposición 12. *Sea $f(x_1, x_2, \ldots, x_n)$ un polinomio no nulo con coeficientes en un cuerpo finito $\mathbb{F}_q$ con $q > d$, donde d es el mayor grado de f en x_i con $1 \leq i \leq n$. Sean $a_1, a_2, \ldots, a_n$ elementos de $\mathbb{F}_q$ elegidos siguiendo una distribución uniforme de probabilidad. Entonces*

$$P\{f(a_1, a_2, \ldots, a_n) \neq 0\} \geq \left(1 - \frac{d}{q}\right)^n$$

De hecho $P\{f(a_1, a_2, \ldots, a_n) \neq 0\} \to 1$ cuando hacemos tender q a infinito.

Demostración. Sea $f(x_1, \ldots, x_n)$ cumpliendo las condiciones del teorema, expresémoslo como un polinomio en x_n con coeficientes en el anillo $\mathbb{F}_q[x_1, \ldots, x_{n-1}]$, esto es

$$f(x_1, \ldots, x_n) = h(x_1, \ldots, x_{n-1})x_n^k + \cdots$$

donde k es el grado de f en x_n, y el polinomio no nulo $h(x_1, \ldots, x_{n-1}) \in \mathbb{F}_q[x_1, \ldots, x_{n-1}]$ el coeficiente para la variable x_n. Razonaremos por inducción sobre n. Para $n = 1$, f es un polinomio de grado k en una sola variable, por lo que tiene como mucho k raíces en $\mathbb{F}_q$, luego hay al menos $q-k$ elementos en $\mathbb{F}_q$ que no anulan a f y por la regla de Laplace $P\{f \neq 0\} \geq \left(\frac{q-k}{q}\right) = \left(1 - \frac{k}{q}\right) \geq \left(1 - \frac{d}{q}\right)$ y se cumple la proposición. Supongamos pues que la proposición se cumple para $n-1$ $n \geq 1$ y probémosla para n.
Se eligen aleatoriamente $a_1, \ldots, a_{n-1}$ en $\mathbb{F}_q$. Por hipótesis de inducción $P\{(h(a_1, \ldots, a_{n-1}) \neq 0\} \geq \left(1 - \frac{d}{q}\right)^{n-1}$. Si $h(a_1, \ldots, a_{n-1}) \neq 0$, entonces $f(a_1, a_2, \ldots, a_{n-1}, x_n)$ es un polinomio de una sola variable y estamos en el caso de la proposición para $n = 1$ luego
$P\{f(a_1, \ldots, a_n) \neq 0 | h(a_1, \ldots, a_{n-1}) \neq 0\} \geq \left(\frac{q-k}{q}\right) = \left(1 - \frac{k}{q}\right) \geq \left(1 - \frac{d}{q}\right)$.
Así, si denotamos por A el suceso $f(a_1, \ldots, a_n) \neq 0$ y por B el suceso $h(a_1, \ldots, a_{n-1}) \neq 0$, tenemos

$$\begin{aligned}
P\{f(a_1, a_2, \ldots, a_n) \neq 0\} &= P(A) = P(A \cap B) + P(A \cap B^c) \geq P(A \cap B) = P(B)P(A|B) && (1.12)\\
&= P\{h(a_1, \ldots, a_{n-1}) \neq 0\} P\{f(a_1, \ldots, a_n) \neq 0 | \\
&\qquad\qquad h(a_1, \ldots, a_{n-1}) \neq 0\} && (1.13)\\
&\geq \left(1 - \frac{d}{q}\right)^{n-1} P\{f(a_1, a_2, \ldots, a_n) \neq 0 | h(a_1, \ldots, a_{n-1}) \neq 0\} && (1.14)\\
&\geq \left(1 - \frac{d}{q}\right)^{n-1} \left(1 - \frac{d}{q}\right) && (1.15)\\
&= \left(1 - \frac{d}{q}\right)^{n} && (1.16)
\end{aligned}$$

Habiéndonos servido de la hipótesis de inducción para llegar a (1.14), y de la proposición para $n = 1$ para pasar a (1.15).

Por último, es evidente que si fijado el polinomio aumentamos q (esto es, el tamaño del cuerpo) la cota para la probabilidad obtenida tiende a 1. □

Podemos entonces aplicar el resultado anterior en el contexto de una red con coeficientes en un cuerpo $\mathbb{F}_q$ de la siguiente forma. Como ya señalamos en una observación previa, $|R|$ es el máximo grado del polinomio de transferencia con respecto a cualquier variable, por lo que si elegimos η coeficientes de forma aleatoria, en virtud de la Proposición 12 la probabilidad de que la comunicación sea exitosa es mayor o igual que

$$\left(1 - \frac{|R|}{q}\right)^{\eta}$$

En [GeTh], haciendo uso de bases de Gröbner, se consiguen cotas más ajustadas, reduciendo el exponente η. Por ejemplo, se demuestra que la probabilidad de que la comunicación funcione es mayor o igual que $\left(1 - \frac{|R|}{q}\right)^{\eta'}$ donde η' es el número de aristas j para las que algún coeficiente $a_{i,j}$ o $f_{i,j}$ ha sido elegido al azar. Esto ya supone una mejora significativa pues η' está acotada por $|E|$. Otra propuesta que aumenta aún más la probabilidad, consiste en tomar η' como el máximo número de aristas j que componen un flujo, para las que algún coeficiente $a_{i,j}$ o $f_{i,j}$ ha sido elegido al azar.

Capítulo 2

Códigos con la Métrica de Rango

Una parte fundamental a la hora de explicar la comunicación segura que queremos tratar en este trabajo, la componen los códigos para la métrica de rango. En este capítulo introducimos la métrica de rango y los códigos matriciales sujetos a esta métrica, así como su relación con los códigos en bloque usuales. Presentamos los polinomios linealizados sobre una extensión de cuerpos, que permiten ver la similitud entre los códigos Reed-Solomon y los códigos MRD, una clase de códigos para la métrica de rango que serán una de las claves para el desarrollo posterior del trabajo. Para el desarrollo del capítulo han sido de utilidad los documentos [GoRav], [Gabi] y [Sil].

En lo que sigue se establecen la notación y bases sobre los códigos lineales para la métrica de rango o rank metric codes. Estos son códigos lineales en bloque usados como códigos correctores de errores: transmiten la información en paquetes de la misma longitud, añadiendo una redundancia que será la que permita detectar y/o corregir la posible información contaminada en el mensaje. La diferencia con los códigos correctores clásicos es la métrica sobre la que se miden los errores, en lugar de usar la conocida distancia de Hamming, se utiliza la llamada distancia de rango.
Recordemos algunas nociones sobre los códigos lineales en bloque, para poner en contexto la posterior construcción de los códigos para la métrica de rango.

Definición 13. *Un código en bloque lineal q-ario $\mathcal{C}$ de longitud n es un subespacio vectorial de $\mathbb{F}_q^n$. Sus elementos, vectores de tamaño n, se denominan palabras o palabras código. Un código lineal tiene un total de q^k palabras distintas*

Para medir el error que se pueda cometer en el envío de palabras, y lo cercanas que están unas palabras de otras, se introduce el concepto de distancia de Hamming.

Definición 14. *Dados dos elementos $x = (X_1, \ldots, X_n), y = (Y_1, \ldots, Y_n) \in \mathbb{F}_q^n$ se define la distancia de Hamming entre x e y como el número de coordenadas en el que se diferencian, esto es,*

$$d(x, y) = \#\{i : 1 \leq i \leq n, X_i \neq Y_i\}$$

A partir de este concepto, llamamos *distancia mínima* del código $\mathcal{C}$ al valor

$$d = d(\mathcal{C}) = \text{mín}\{d(x, y) : x, y \in \mathcal{C}, x \neq y\}$$

Dada una palabra código $x \in \mathcal{C}$ se define su peso $w(x)$ como el número de coordenadas no nulas de x, y se define el peso mínimo $w(\mathcal{C})$ del código, como el mínimo de los pesos de sus elementos, excluyendo el vector 0. En un código lineal, el peso mínimo es igual a la distancia mínima.

Un código lineal $\mathcal{C} \subseteq \mathbb{F}_q^n$ suele ser representado por sus parámetros $[n, k, d]_q$ (o $[n, k]_q$) siendo n su longitud (es decir, la longitud de cada una de sus palabras), k la dimensión como espacio vectorial y d su distancia mínima.

Es sabido que se cumplen una variedad de cotas entre los parámetros de cualquier código lineal. Una de gran relevancia es la cota de Singleton.

Teorema 15 (Cota de Singleton). *Sea $\mathcal{C} \subseteq \mathbb{F}_q^n$ un $[n,k,d]_q$ código lineal, entonces*

$$k + d \leq n + 1 \tag{2.1}$$

Demostración. De cada una de las q^k palabras distintas del código, se suprimen las $d-1$ últimas coordenadas, obteniendo las q^k palabras de la forma $(x_1, \ldots, x_{n-d+1})$, que siguen siendo distintas (pues si hubiese dos iguales, la distancia entre ambas sería menor que d, contradiciendo el hecho de que la distancia mínima es d). Luego esas q^k palabras están entre las q^{n-d+1} formas posibles de construir un vector de longitud $n-d+1$ con elementos en $\mathbb{F}_q$. Así, $q^k \leq q^{n-d+1}$ y se deduce la cota buscada. □

Un código que cumpla con igualdad (2.1) se dice que es de Máxima Distancia de Separación, o MDS.

Se repasa ahora la definición de dos matrices que describen un código lineal.

Definición 16. *Se denomina matriz generadora de un código $[n,k,d]_q$ a una matriz G de tamaño $k \times n$ con entradas en $\mathbb{F}_q$, cuyas filas formen una base del código (lo que implica que el rango de G es k)*

La forma de codificar un mensaje $x \in \mathbb{F}_q^k$ pasa por el uso de una matriz generadora del código, realizándose la operación $xG = y$ para producir el mensaje codificado $y \in \mathbb{F}_q^n$ con la información redundante que posibilita la corrección de errores.
Otra forma de describir un código lineal $\mathcal{C}$, viene determinada por la llamada *matriz de control*.

Definición 17. *Llamaremos matriz de control (o matriz de paridad) de un código, a una matriz H de tamaño $(n-k) \times n$, rango $n-k$, con entradas en $\mathbb{F}_q$ para la que se cumple que $x \in \mathcal{C} \Leftrightarrow Hx^t = 0, \forall x \in \mathcal{C}$. Es decir,*

$$\mathcal{C} = \{x \in \mathbb{F}_q^n : Hx^t = 0\}$$

La relación entre ambas matrices es inmediata, basta fijarse en su construcción para darse cuenta de que, dadas G y H matrices generadora y de control de un código $\mathcal{C}$, se cumple $GH^t = HG^t = 0$. Por supuesto, las matrices tanto de control como generadora de un código lineal no son únicas, al no serlo las bases en las que se puede expresar un espacio vectorial. Ahora bien, cada matriz generadora G proporciona una codificación, pues $\mathcal{C} = \{xG : x \in \mathbb{F}_q^k\}$.

En el siguiente ejemplo se muestran las matrices descritas para un código Reed-Solomon (RS), una clase de códigos MDS que pueden tener una descripción similar a los códigos de Gabidulin que se verán más adelante.

Ejemplo 18. Sea el cuerpo finito $\mathbb{F}_{13}$ y α un elemento primitivo, en este caso $\alpha = 2$ es válido pues $ord(2) = 12$. El código RS tiene longitud $n = q - 1 = 12$, y pongamos que corrige $t = 2$ errores, luego como $t = \lfloor \frac{d-1}{2} \rfloor$, elegimos el código de distancia mínima $d = 5$. Al ser un código MDS se tiene que cumplir $n - k = d - 1 = 4$ y se deducen los parámetros $[12, 8, 5]_{13}$. Fijado $x = \{1, \alpha, \alpha^2, \ldots, \alpha^{q-2}\} = \{1, 2^1, \ldots, 2^{11}\}$ y el espacio vectorial de los polinomios sobre $\mathbb{F}_{13}$ de grado menor que $k = 8$, denotado por P_8; el código puede verse como la evaluación de los polinomios de P_8 en las componentes de x. Así, codificar un mensaje es equivalente a verlo como un polinomio de

P_8 y evaluarlo en x. Una matriz generadora del código es entonces

$$G_{k\times n} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \alpha & \alpha^2 & \cdots & \alpha^{11} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha^7 & (\alpha^2)^7 & \cdots & (\alpha^{11})^7 \end{bmatrix}$$

Y una matriz de control del código se puede construir como se muestra en [JuTo] de la forma

$$H_{n-k\times n} = \begin{bmatrix} 1 & \alpha & \alpha^2 & \cdots & \alpha^{11} \\ 1 & \alpha^2 & (\alpha^2)^2 & \cdots & (\alpha^{11})^2 \\ 1 & \alpha^3 & (\alpha^2)^3 & \cdots & (\alpha^{11})^3 \\ 1 & \alpha^4 & (\alpha^2)^4 & \cdots & (\alpha^{11})^4 \end{bmatrix}$$

2.1. Distancia de rango

Se denota por $\mathbb{F}_q^{n\times m}$ al conjunto de todas las matrices $n \times m$ sobre el cuerpo $\mathbb{F}_q$, y escribimos $\mathbb{F}_q^n = \mathbb{F}_q^{1\times n}$. Esto es, $v \in \mathbb{F}_q^n$ es un vector fila, salvo que se aclare otra cosa en el contexto. Dada una matriz A, el espacio generado por las filas (o columnas) de A será $\langle A\rangle$.

Se considera el cuerpo finito $\mathbb{F}_{q^m}$ visto como la extensión de grado m del cuerpo finito $\mathbb{F}_q$. De hecho $\mathbb{F}_{q^m}$ es un espacio vectorial de dimensión m sobre $\mathbb{F}_q$, y podemos considerar una base $\mathcal{B} = \{\alpha_1, \ldots, \alpha_m\}$ de $\mathbb{F}_{q^m}$ sobre $\mathbb{F}_q$. Así, cada elemento $a \in \mathbb{F}_{q^m}$ tiene una representación única $(a_1, \ldots, a_m)$ en la base $\mathcal{B}$, y puede escribirse $a = \sum_{i=1}^m a_i\alpha_i$ con $a_i \in \mathbb{F}_q$.
Entonces, está bien definida la biyección $[\cdot]_{\mathcal{B}} : \mathbb{F}_{q^m} \to \mathbb{F}_q^m$ que asocia a cada elemento $a \in \mathbb{F}_q^m$ el vector fila con sus coordenadas en la base $\mathcal{B}$

$$[a]_{\mathcal{B}} = (a_1, \ldots, a_m)$$

Si ahora pensamos en el espacio vectorial n-dimensional $\mathbb{F}_{q^m}^n$ sobre el cuerpo $\mathbb{F}_{q^m}$, la anterior biyección podemos extenderla a $[\cdot]_{\mathcal{B}} : \mathbb{F}_{q^m}^n \to \mathbb{F}_q^{n\times m}$. Dado $x = (X_1, \ldots, X_n) \in \mathbb{F}_{q^m}^n$, si $(a_{i,1}, a_{i,2}, \ldots, a_{i,m})$ son las coordenadas de X_i en la base $\mathcal{B}$, entonces $[\cdot]_{\mathcal{B}}$ aplica x en la matriz cuyas filas son las coordenadas de las entradas del vector x. Así

$$[x]_{\mathcal{B}} = \begin{bmatrix} [X_1]_{\mathcal{B}} \\ \vdots \\ [X_n]_{\mathcal{B}} \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,m} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,m} \end{bmatrix} \tag{2.2}$$

Esta biyección es de gran importancia, pues es la forma de relacionar un vector $x \in \mathbb{F}_{q^m}^n$ con su correspondiente matriz $[x]_{\mathcal{B}}$, que nos permitirá entender un código para la métrica de rango bien como un código en bloque usual o bien como un código matricial.

Ejemplo 19. Consideremos la extensión de cuerpos $\mathbb{F}_2 \hookrightarrow \mathbb{F}_{2^4} = \mathbb{F}_{16}$ de grado 4, siendo $p(x) = x^4 + x + 1 \in \mathbb{F}_2[x]$ el polinomio primitivo para generarla. Sea $\alpha \in \mathbb{F}_{16}$ raíz de $p(x)$, una base de $\mathbb{F}_{16}$ como espacio vectorial sobre $\mathbb{F}_2$ es $\mathcal{B} = \{1, \alpha, \alpha^2, \alpha^3\}$. De hecho, al ser α elemento primitivo, los elementos de $\mathbb{F}_{16}$ se pueden escribir como $\{0, 1, \alpha, \alpha^2, \ldots, \alpha^{14}\}$. Haciendo las cuentas usando $\alpha^4 + \alpha + 1 = 0$, los elementos de $\mathbb{F}_{16}$ expresados en $\mathcal{B}$ son

$0 = 0$	$1 = 1$	$\alpha = \alpha$	$\alpha^2 = \alpha^2$
$\alpha^3 = \alpha^3$	$\alpha^4 = \alpha + 1$	$\alpha^5 = \alpha + \alpha^2$	$\alpha^6 = \alpha^3 + \alpha^2$
$\alpha^7 = 1 + \alpha + \alpha^3$	$\alpha^8 = 1 + \alpha^2$	$\alpha^9 = \alpha + \alpha^3$	$\alpha^{10} = 1 + \alpha + \alpha^2$
$\alpha^{11} = \alpha + \alpha^2 + \alpha^3$	$\alpha^{12} = 1 + \alpha + \alpha^2 + \alpha^3$	$\alpha^{13} = 1 + \alpha^2 + \alpha^3$	$\alpha^{14} = 1 + \alpha^3$

Luego dado un caso concreto $x = (\alpha^2, \alpha^7, \alpha^{14}, \alpha^9) \in \mathbb{F}_{16}^4$, su imagen por la biyección $[\cdot]_{\mathcal{B}}$ es

$$[x]_{\mathcal{B}} = \begin{bmatrix} [\alpha^2]_{\mathcal{B}} \\ [\alpha^7]_{\mathcal{B}} \\ [\alpha^{14}]_{\mathcal{B}} \\ [\alpha^9]_{\mathcal{B}} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

Considerando un vector n-dimensional $x = (x_1, x_2, \ldots, x_n) \in \mathbb{F}_{q^m}^n$, nos es de utilidad adaptar el concepto de rango para un vector.

Definición 20. *Sea $x = (X_1, X_2, \ldots, X_n) \in \mathbb{F}_{q^m}^n$, se define rango de x, denotado por $rg(x)$, al máximo número de coordenadas de x que son linealmente independientes sobre $\mathbb{F}_q$.*

Nota. Esta definición es en realidad equivalente a decir, que el rango de x es el rango de la matriz $[x]_{\mathcal{B}}$, fijada una base $\mathcal{B}$ de $\mathbb{F}_{q^m}$ sobre $\mathbb{F}_q$.

Conociendo ya el concepto usual de rango de una matriz, y el reciente rango de un vector, estamos en condiciones de introducir la distancia que se utilizará al tratar con los códigos con métrica de rango.

Definición 21. *Sean las matrices $A, B \in \mathbb{F}_q^{n\times m}$, se define la distancia de rango como el rango de la matriz diferencia $A - B$, es decir*

$$\begin{aligned} d_R : \mathbb{F}_q^{n\times m} \times \mathbb{F}_q^{n\times m} &\longrightarrow \mathbb{N} \\ (A, B) &\longmapsto rg(A - B) \end{aligned}$$

Comprobar las condiciones necesarias para ser una distancia es inmediato teniendo en cuenta las propiedades básicas de las matrices. En cuanto a la desigualdad triangular, basta reescribir las matrices de la forma adecuada y recordar la propiedad $rg(A + B) \leq rg(A) + rg(B)$. Recordemos que el rango de una matriz $A \in \mathbb{F}_q^{n\times m}$ es el número máximo de filas (o columnas) linealmente independientes y es como mucho mín$\{n, m\}$.

En las condiciones expuestas, al igual que se ha hablado de rango de un vector, tiene sentido adaptar la definición de distancia de rango para vectores de $\mathbb{F}_{q^m}^n$. Y esta será la distancia de rango entre sus imágenes por $[\cdot]_{\mathcal{B}}$.

Definición 22. *Sean $x, y \in \mathbb{F}_{q^m}^n$. La distancia de rango entre x e y se define como $d_R(x, y) = d_R([x]_{\mathcal{B}}, [y]_{\mathcal{B}})$*

Así con esta explicación, se introducen los códigos para la métrica de rango, que pueden presentarse desde dos perspectivas, la matricial y la vectorial. La siguiente definición muestra la perspectiva matricial.

Definición 23. *Un código para la métrica de rango o código matricial de longitud n es un conjunto no vacío de matrices $n \times m$ sobre $\mathbb{F}_q$, en particular un subespacio lineal (en $\mathbb{F}_q$) $\mathcal{C} \subset \mathbb{F}_q^{n\times m}$, cuyos elementos llamaremos palabras código.*

La distancia que usaremos para estos códigos, es la distancia de rango explicada en la Definición 21.
Con la reciente forma introducida en (2.2) de relacionar unívocamente un vector en $\mathbb{F}_{q^m}^n$ con una matriz en $\mathbb{F}_q^{n\times m}$ a través de $[\cdot]_{\mathcal{B}}$, se puede entender un código matricial como un código lineal

en bloque usual $\mathcal{C}$ sobre el cuerpo $\mathbb{F}_{q^m}$. Y esto es lo que presenta la perspectiva vectorial de un código con la métrica de rango. Es decir, cada palabra $x \in \mathbb{F}_{q^m}^n$ de un código $\mathcal{C} \subseteq \mathbb{F}_{q^m}^n$ puede verse aplicándola $[\cdot]_{\mathcal{B}}$, como una palabra $[x]_{\mathcal{B}}$ de un código matricial $\mathcal{C}' \subseteq \mathbb{F}_q^{n\times m}$, y viceversa (recordemos que $[\cdot]_{\mathcal{B}}$ es una biyección).
En la versión vectorial, donde las palabras código son vectores de $\mathbb{F}_{q^m}$, la distancia de rango no es otra que la presentada en la Definición 22.

Nota. Dado un código clásico $\mathcal{C} \subseteq \mathbb{F}_{q^m}^n$, al código $\mathcal{C}' \subseteq \mathbb{F}_q^{n\times m}$ cuyas palabras son las imágenes por $[\cdot]_{\mathcal{B}}$ de las palabras de $\mathcal{C}$, lo denotaremos por $[\mathcal{C}]_{\mathcal{B}}$.

De forma similar a cómo se define la distancia mínima de Hamming para códigos clásicos, la distancia de rango mínima de un código matricial $\mathcal{C} \subset \mathbb{F}_q^{n\times m}$ es la distancia mínima entre todos los pares de palabras código distintas de $\mathcal{C}$, y se denota por $d_R(\mathcal{C})$. Esto es

$$d_R(\mathcal{C}) = \text{mín}\{d_R(x, y) : x, y \in \mathcal{C}, x \neq y\}$$

Considerando la métrica de rango, si definimos el peso de una palabra como su rango, al igual que con la distancia de Hamming la distancia mínima de rango de un código coincide con el peso mínimo.

Una forma de descodificar basándonos en esta distancia es la usual: recibida una palabra $y \in \mathbb{F}_q^{n\times m}$, si $y \in \mathcal{C}$ se interpreta que la palabra se ha transmitido correctamente. Si $y \notin \mathcal{C}$, se considera que ha habido algún error en la comunicación, y se devuelve una palabra código $z \in \mathcal{C}$ tal que la distancia de rango entre y y z sea mínima. Si esta distancia mínima se cumple para varias palabras, esta forma de descodificar falla. Ahora bien, si para alguna palabra $x \in \mathcal{C}$ se tiene $d_R(y, x) < d_R(\mathcal{C})/2$, entonces dicha distancia es la mínima alcanzable y solo la cumple x, por lo que será la palabra que se devuelva.

Ejemplo 24. Sea el cuerpo finito $\mathbb{F}_{2^3}$ generado por el polinomio primitivo $x^3 + x + 1 \in \mathbb{F}_2[x]$ y sea $\alpha \in \mathbb{F}_8$ raíz de dicho polinomio. Entonces, se sabe que $\mathcal{B} = \{1, \alpha, \alpha^2\}$ es una base de $\mathbb{F}_8$ sobre $\mathbb{F}_2$. Podemos construir el código $\mathcal{C} = \langle(1, \alpha)\rangle \subseteq \mathbb{F}_8^2$ de dimensión $k = 1$ generado sobre $\mathbb{F}_8$.
Buscando la versión matricial de $\mathcal{C}$ haciendo uso de los cálculos $\alpha(1, \alpha) = (\alpha, \alpha^2)$ y $\alpha^2(1, \alpha) = (\alpha, \alpha + 1)$, tenemos

$$[(1, \alpha)]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, [(\alpha, \alpha^2)]_{\mathcal{B}} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, [(\alpha^2, \alpha + 1)]_{\mathcal{B}} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}. \tag{2.3}$$

Y describimos el código $\mathcal{C}' = [\mathcal{C}]_{\mathcal{B}}$, cuya dimensión sobre $\mathbb{F}_2$ es $mk = 3$, como sigue

$$\mathcal{C}' = \left\langle \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \right\rangle \subseteq \mathbb{F}_2^{2\times 3} \tag{2.4}$$

2.2. Códigos de Máxima Distancia de Rango o MRD

Al igual que ocurre en los códigos en bloque con la distancia de Hamming, los códigos matriciales también satisfacen cotas sobre su tamaño que involucran la distancia de rango.

La distancia mínima de rango d_R de un código puede ser acotada. Para empezar, una cota burda de d_R para un código en bloque sobre $\mathbb{F}_{q^m}$, es m (recordemos que d_R es el rango de una matriz $n \times m$, luego $d_R \leq \text{mín}\{m, n\}$).
Se muestra en [Gabi], que dadas dos normas n_1 y n_2 sobre un espacio vectorial n-dimensional sobre $\mathbb{F}_{q^m}$, si $n_1(x) \leq n_2(x), \forall x$, entonces las distancias mínimas d_1 y d_2 sobre un mismo código que

inducen dichas normas verifican $d_1 \leq d_2$. Haciendo uso de esto, gracias a que el peso de Hamming de un vector, y el rango de un vector tal como ya se ha definido son normas que cumplen dicha propiedad, las distancias mínimas que inducen d_H y d_R (distancia mínima de Hamming y distancia mínima de rango) verifican $d_R \leq d_H$.
Entonces considerado un código en bloque $[n, k]$ sobre $\mathbb{F}_{q^m}$, gracias a la cota de Singleton descrita en el Teorema 15 vemos que la distancia mínima de rango satisface

$$d_R \leq n - k + 1 \tag{2.5}$$

Continuando con la representación vectorial, podemos determinar el tamaño de un código. Consideramos un código lineal $\mathcal{C}$ sobre $\mathbb{F}_{q^m}$ de parámetros $[n, k, d_R]$, esto es un subespacio del espacio $\mathbb{F}_{q^m}^n$. Se puede describir el código en base a una matriz generadora G con entradas en $\mathbb{F}_{q^m}$, que tendrá rango k. Como se sabe, cualquier palabra código es una combinación lineal (en $\mathbb{F}_{q^m}$) de las k filas de G, luego el número de palabras del código es

$$|\mathcal{C}| = q^{mk}$$

El tamaño de un código para la métrica de rango viene acotado por la cota de Singleton para la métrica de rango.

Teorema 25 (**Cota de Singleton**). *Sea $\mathcal{C} \subseteq \mathbb{F}_q^{n\times m}$ un código lineal con distancia de rango mínima d_R. Entonces se verifica*

$$\begin{aligned} |\mathcal{C}| &\leq q^{\min\{n(m-d_R+1),m(n-d_R+1)\}} \\ &= q^{\max\{n,m\}(\min\{n,m\}-d_R+1)} \end{aligned} \tag{2.6}$$

Demostración. La demostración es similar a la de la cota de Singleton para la distancia de Hamming. Supongamos, sin pérdida de generalidad, que $n \leq m$. De cada palabra código de $\mathcal{C}$ suprimimos las últimas $d_R - 1$ filas, obteniendo $|\mathcal{C}|$ matrices de tamaño $(n - d_R + 1) \times m$, todas ellas distintas pues de haber dos matrices iguales la distancia entre las palabras código de las que provienen sería como mucho $d_R - 1$, contradiciendo el hecho de que la distancia mínima es d_R. Como hay $q^{m(n-d_R+1)}$ formas de construir una matriz $(n - d_R + 1) \times m$ sobre $\mathbb{F}_q$, se deduce $|\mathcal{C}| \leq q^{m(n-d_R+1)}$.
Si $m < n$ el razonamiento es el mismo, suprimiendo columnas en vez de filas. □

Los códigos que cumplen con igualdad la cota (2.6) son llamados códigos de máxima distancia de rango o MRD, y como veremos al final del capítulo existen para cualquier elección de los parámetros n, m, q y $d_R \leq \min\{m, n\}$. De hecho, cuando $m \geq n$ la cota de Singleton para la métricas de Hamming y de rango coinciden, y todo código MRD es también MDS.

Nos centramos en el caso particular en que $n \leq m$. Una familia importante de códigos MRD, son los llamados códigos de Gabidulin, introducidos por primera vez de la mano del autor que les da nombre y Delsarte en [Gabi].

Definición 26 (**Códigos de Gabidulin**). *Sean $g_1, g_2, \ldots, g_n \in \mathbb{F}_{q^m}$ elementos linealmente independientes sobre $\mathbb{F}_q$. Entonces, se denomina código de Gabidulin generado por $g_1, g_2, \ldots, g_n$ al código definido por la matriz generadora*

$$G = \begin{bmatrix} g_1 & g_2 & \cdots & g_n \\ g_1^{[1]} & g_2^{[1]} & \cdots & g_n^{[1]} \\ \vdots & \vdots & \ddots & \vdots \\ g_1^{[k-1]} & g_2^{[k-1]} & \cdots & g_n^{[k-1]} \end{bmatrix} = \begin{bmatrix} g_1 & g_2 & \cdots & g_n \\ g_1^{q^1} & g_2^{q^1} & \cdots & g_n^{q^1} \\ \vdots & \vdots & \ddots & \vdots \\ g_1^{q^{k-1}} & g_2^{q^{k-1}} & \cdots & g_n^{q^{k-1}} \end{bmatrix}$$

siendo $[i] = q^i$. Su dimensión es k, y al ser MRD tiene distancia mínima de rango $d_R = n - k + 1$.

Para demostrar que G define un código MRD es necesario introducir polinomios linealizados, lo cual se hace en la próxima subsección.

También se pueden definir a partir de su matriz de paridad de control como se describe a continuación. Si $h_1, h_2, \ldots, h_n \in \mathbb{F}_{q^m}$ son elementos linealmente independientes sobre $\mathbb{F}_q$, entonces la matriz de control de la forma

$$H = \begin{bmatrix} h_1 & h_2 & \cdots & h_n \\ h_1^{[1]} & h_2^{[1]} & \cdots & h_n^{[1]} \\ \vdots & \vdots & \ddots & \vdots \\ h_1^{[n-k-1]} & h_2^{[n-k-1]} & \cdots & h_n^{[n-k-1]} \end{bmatrix}$$

define un código de Gabidulin de longitud n, dimensión k y distancia mínima de rango $d_R = n-k+1$.

Aunque no vayan a ser objeto de nuestro estudio, cabe nombrar otras dos familias de códigos MRD que guardan estrecha relación con los códigos de Gabidulin.
Unos son los códigos de Gabidulin generalizados, que se caracterizan como sigue

Definición 27. *Sean $g_1, g_2, \ldots, g_n \in \mathbb{F}_{q^m}$ elementos linealmente independientes sobre $\mathbb{F}_q$, y sea $s \in \mathbb{N}$ coprimo con m (m.c.d(s,m)=1). Entonces, se denomina código de Gabidulin generalizado $\mathcal{C} \subseteq \mathbb{F}_{q^m}^n$ al código con matriz generadora*

$$G = \begin{bmatrix} g_1 & g_2 & \cdots & g_n \\ g_1^{[s]} & g_2^{[s]} & \cdots & g_n^{[s]} \\ \vdots & \vdots & \ddots & \vdots \\ g_1^{[s(k-1)]} & g_2^{[s(k-1)]} & \cdots & g_n^{[s(k-1)]} \end{bmatrix}$$

Como se puede apreciar, los códigos de Gabidulin son un caso particular de los códigos de Gabidulin generalizados (el caso $s = 1$), pero fueron estos últimos los más recientes en aparecer, gracias a E.Gabidulin y A.Kshevetskiy, generalizando los primeros.
La otra familia de códigos MRD requiere que $n = m$. Se parte de considerar un código $\mathcal{C}$ de máxima distancia de rango con parámetros $(n, k, d_R = n-k+1)$ sobre $\mathbb{F}_{q^m}$, y realizar l productos cartesianos de sí mismo. Así, el código definido por $\mathcal{C}^l = \mathcal{C} \times \ldots \times \mathcal{C}$, es un código de longitud $n' = nl$, dimensión $k' = kl$ y distancia mínima de rango $d'_R = d_R$. Este código así definido, se sabe que es MRD si y solo si $n = m$. Además cabe fijarse en que, al contrario que los códigos de Gabidulin, la longitud de estos códigos MRD es mayor que m, esta es $n' = nl = ml \geq m$.

Una herramienta que relaciona la definición de los códigos Reed Solomon con la definición de códigos de Gabidulin se proporciona en la siguiente sección.

2.2.1. Polinomios linealizados

Una clase de polinomios que juegan un papel importante en el estudio de los códigos para la métrica de rango son los siguientes.

Definición 28. *Un polinomio linealizado (o q-polinomio) sobre $\mathbb{F}_{q^m}$ es un polinomio de la forma*

$$f(x) = \sum_{i=0}^{n} f_i x^{[i]}$$

donde $f_i \in \mathbb{F}_{q^m}$ y $x^{[i]} = x^{q^i}$. Llamamos q-grado de $f(x)$ al mayor entero i tal que $f_i \neq 0$. El espacio vectorial de polinomios linealizados sobre $\mathbb{F}_{q^m}$ de q-grado menor o igual que s lo denotamos por $Lin_q(m, s)$, y tiene dimensión $dim_{\mathbb{F}_{q^m}}(Lin_q(m, s)) = s + 1$.

En general relajaremos la notación escribiendo $f(x) = g(x)$ si $f(x) \equiv g(x)$, es decir si $f(x) - g(x) \equiv 0$ es el polinomio nulo.

Exponemos algunas de sus propiedades en esta sección.

Los polinomios linealizados reciben su nombre debido a la siguiente propiedad. Dado un polinomio linealizado f, se verifica que para todo $a_1, a_2 \in \mathbb{F}_q$ y $\beta_1, \beta_2 \in \mathbb{F}_{q^m}$

$$f(a_1\beta_1 + a_2\beta_2) = a_1 f(\beta_1) + a_2 f(\beta_2)$$

Es decir, la evaluación de un polinomio linealizado es una transformación $\mathbb{F}_q$-lineal de $\mathbb{F}_{q^m}$ en sí mismo. En particular, el conjunto de raíces en $\mathbb{F}_{q^m}$ de un polinomio linealizado f es un subespacio de $\mathbb{F}_{q^m}$, ya que es el núcleo de la transformación recién descrita, y lo denotamos por $V(f) \subseteq \mathbb{F}_{q^m}$. Se tiene por el Teorema Fundamental del Álgebra que la dimensión de $V(f)$ sobre $\mathbb{F}_{q^m}$ es menor o igual que el q-grado de f, siendo estrictamente menor si se repite alguna raíz.

Sean dos polinomios linealizados f y g con q-grados n y k respectivamente. Se define la composición, o multiplicación simbólica de f y g, como el polinomio

$$f(x) \otimes g(x) = f(g(x))$$

Esta operación no es conmutativa en general. El polinomio $P(x) = f(x) \otimes g(x)$ es un polinomio linealizado con q-grado $t = n + k$, cuyos coeficientes se pueden calcular de la forma

$$P_l = \sum_{i=\text{máx}\{0,l-k\}}^{\text{mín}\{l,n\}} f_i g_{l-i}^{[i]} = \sum_{j=\text{máx}\{0,l-n\}}^{\text{mín}\{l,k\}} f_{l-j} g_j^{[l-j]}, \quad l = 0, \ldots, t.$$

En particular, si $n \leq k$ entonces

$$P_l = \sum_{i=0}^{n} f_i g_{l-i}^{[i]}, \quad n \leq l \leq k$$

mientras que si $k \leq n$, entonces

$$P_l = \sum_{j=0}^{k} f_{l-j} g_j^{[l-j]}, \quad k \leq l \leq n$$

Se sabe que el conjunto de polinomios linealizados sobre $\mathbb{F}_{q^m}$ con las operaciones de suma $(+)$ y multiplicación simbólica $(\otimes)$ forman un anillo no conmutativo con elemento identidad, que además no tiene divisores de cero.

Se define a continuación el q-inverso de un polinomio linealizado

$$f(x) = \sum_{i=0}^{n} f_i x^{[i]}$$

como el polinomio linealizado

$$\bar{f}(x) = \sum_{i=0}^{n} \bar{f}_i x^{[i]}$$

donde $\bar{f}_i = f_{n-i}^{[i-n]}$ con $i = 0, \ldots, n$. Si n no se especifica, se toma como el q-grado de $f(x)$.

Una de las propiedades más destacables de los polinomios linealizados es la de proporcionar aplicaciones para núcleos previamente especificados, es decir, interpolar. Dado un subconjunto $S \subseteq \mathbb{F}_{q^m}$,

existe un único polinomio linealizado mónico $M_S(x) = \sum_{i=0}^{n} M_i x^{[i]}$ de menor q-grado, cuyo espacio de raíces contiene a S. Este polinomio recibe el nombre de q-polinomio mínimo de S.
Además, $M_S(x)$ también puede definirse como

$$M_S(x) = \prod_{\alpha \in \langle S \rangle} (x - \alpha)$$

por lo que el q-grado de $M_S(x)$ es igual a la $\dim\langle S\rangle$, e igual a $\dim(ker(M_S(x))$ vista como aplicación lineal.

Por otra parte, si $f(x)$ es un polinomio linealizado cuyo espacio de raíces contiene a S, entonces se tiene

$$f(x) = Q(x) \otimes M_S(x)$$

para algún polinomio linealizado $Q(x)$. Esto implica que $M_{S\cup\{\alpha\}}(x) = M_{M_S(\alpha)}(x) \otimes M_S(x)$ para cualquier α. De hecho, el polinomio $M_S(x)$ puede ser calculado de manera recursiva de la siguiente forma. Sea $\{s_1 \ldots, s_t\}$ una base de $\langle S\rangle$. Iniciamos $M_{s_1}(x) = x - x^{[1]} s_1 / s_1^{[1]}$, y para $i = 2, \ldots, t$ denotamos $z_i = M_{\{s_1,\ldots,s_{i-1}\}}(s_i)$ y calculamos $M_{\{s_1,\ldots,s_i\}}(x) = M_{z_i}(x) \otimes M_{\{s_1,\ldots,s_{i-1}\}}(x)$.

Habiendo visto cómo se construyen los códigos de Gabidulin, cabe notar que estos pueden describirse en términos de polinomios linealizados, tal como los códigos Reed-Solomon se pueden describir en términos de polinomios convencionales como se presentó en el ejemplo 18. Dados elementos linealmente independientes $g_1, \ldots, g_n \in \mathbb{F}_{q^m}$, recordamos cómo es la matriz generadora de un código de Gabidulin

$$G = \begin{bmatrix} g_1 & g_2 & \cdots & g_n \\ g_1^{[1]} & g_2^{[1]} & \cdots & g_n^{[1]} \\ \vdots & \vdots & \ddots & \vdots \\ g_1^{[k-1]} & g_2^{[k-1]} & \cdots & g_n^{[k-1]} \end{bmatrix}$$

Así, un código de Gabidulin es el conjunto de palabras $c \in \mathbb{F}_{q^m}^n$

$$c = (f(g_1), f(g_2), \ldots, f(g_n))$$

donde $f(x)$ es algún polinomio linealizado de q-grado menor que k. Es decir, las palabras de un código Gabidulin son las evaluaciones de polinomios linealizados en los elementos linealmente independientes que definen la matriz generadora del código. Así cada palabra $c \in \mathcal{C}$ se forma calculando

$$c = (f(g_1), f(g_2), \ldots, f(g_n)) = (f_0, f_1, \ldots, f_{k-1})G$$

para todo polinomio linealizado $f(x) = \sum_{i=0}^{k-1} f_i x^{[i]} \in Lin_q(m, k-1)$.

Por último, mostramos un resultado que se dejaba pendiente al introducir los códigos de Gabidulin: la matriz que genera esta clase de códigos efectivamente define un código MRD.

Lema 29. *Sea $n > 0$ y sean $f(x), g(x) \in Lin_q(m, n-1)$. Si $\alpha_1, \ldots, \alpha_n \in \mathbb{F}_{q^m}$ son elementos linealmente independientes tal que $f(\alpha_i) = g(\alpha_i)$ para $i = 1, \ldots, n$, entonces $f(x) = g(x)$.*

Demostración. Sea $h(x) = f(x) - g(x)$, entonces $\alpha_1, \ldots, \alpha_n$ son raíces de $h(x)$, y por tanto también lo serán sus q^n combinaciones lineales, por lo que $h(x)$ tiene q^n raíces distintas, mientras que el q-grado de $h(x)$ es n. Es sabido que si un polinomio tiene más raíces que su grado, entonces éste solo puede ser el polinomio nulo, con lo que se concluye que $h(x) = 0$. □

Teorema 30. *Para todos $1 \leq d \leq n$, existe un código lineal $\mathcal{C} \subseteq \mathbb{F}_{q^m}^n$ para la métrica de rango tal que $d_R(\mathcal{C}) = d$ y $dim_{\mathbb{F}_{q^m}}(\mathcal{C}) = k = n - d + 1$, es decir un código MRD.*

Demostración. Sea $E = \{\alpha_1, \ldots, \alpha_n\} \subseteq \mathbb{F}_{q^m}$ un conjunto de elementos linealmente independientes, viendo $\mathbb{F}_{q^m}$ como espacio vectorial de dimensión m sobre $\mathbb{F}_q$. Dichos elementos existen ya que se asume $n \leq m$. Se define la aplicación $\mathbb{F}_{q^m}$-lineal

$$\begin{aligned} ev_E : Lin_q(m, n-d) &\longrightarrow \mathbb{F}_{q^m}^n \\ f &\longmapsto (f(\alpha_1), \ldots, f(\alpha_n)) \end{aligned}$$

Demostraremos que $\mathcal{C} = ev_E(Lin_q(m, n-d)) \subseteq \mathbb{F}_{q^m}^n$ es un código para la métrica de rango con las propiedades deseadas.
En primer lugar $\mathcal{C}$ es $\mathbb{F}_{q^m}$-lineal. Sea ahora $f \in Lin_q(m, n-d)$ un polinomio linealizado no nulo, y sea $W \subseteq \mathbb{F}_{q^m}$ el espacio generado sobre $\mathbb{F}_q$ por las evaluaciones $f(\alpha_1), \ldots, f(\alpha_n)$. El polinomio f induce una aplicación evaluación $\mathbb{F}_q$-lineal $f : \langle \alpha_1, \ldots, \alpha_n \rangle_{\mathbb{F}_q} \to \mathbb{F}_{q^m}$. El conjunto $Im(f)$ es de hecho W, y por el teorema del rango-nulidad se tiene $\dim_{\mathbb{F}_q}(W) = n - \dim_{\mathbb{F}_q} V(f)$. Como $\dim_{\mathbb{F}_q} V(f)$ es menor o igual que el q-grado de f se tiene $\dim_{\mathbb{F}_q}(W) \geq n - (n-d) = d$. Por tanto $dim_{\mathbb{F}_q}(W) = dim(\langle f(\alpha_1), \ldots, f(\alpha_n) \rangle) \geq d$ implica que $rg(f(\alpha_1), \ldots, f(\alpha_n)) \geq d$, y esto ocurre para todo $f \in Lin_q(m, n-d)$ por lo que obtenemos $d_R(\mathcal{C}) \geq d$. En particular, como $d \geq 1$, la aplicación ev_E es inyectiva en virtud del lema 29, y la dimensión de $\mathcal{C}$ es $k = dim_{\mathbb{F}_{q^m}}(Lin_q(m, n-d)) = n - d + 1$. Por último de la cota de Singleton (2.6) se obtiene $d_R(\mathcal{C}) = d$. □

Capítulo 3

Entropía e Información Mutua

En este capítulo sentamos las herramientas matemáticas necesarias para poder medir cuánta información obtiene un intruso que observa cierta cantidad de conexiones en una red de comunicación. En él introducimos algunos conceptos básicos útiles sobre teoría de la información: se definen las cantidades denominadas *entropía* e *información mutua* como funcionales de distribuciones de probabilidad, y las relaciones entre ellas. La noción de entropía y sus propiedades reflejan cómo debería ser de manera intuitiva una medida de información. Esto lo extendemos con la llamada información mutua, que es una medida de la información que contiene una variable aleatoria sobre otra. La bibliografía utilizada para este capítulo son los primeros capítulos de los libros [CoJo] y [LoVei].

3.1. Entropía

Introducimos primero el concepto de entropía como medida de la incertidumbre sobre una variable aleatoria. Sea X una variable aleatoria discreta de rango finito $\mathcal{X}$ con función de masa de probabilidad $p(x) = Pr\{X = x\}, x \in \mathcal{X}$ (lo escribiremos como $X \sim p(x)$).

Definición 31. *La entropía $H(X)$ de una variable aleatoria discreta X viene definida por*

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) log(p(x))$$

Nota. Para valores x de probabilidad nula, dado que $\lim_{p(x)\to 0} p(x) log\, p(x) = 0$, adoptamos el convenio $p(x) log\, p(x) = 0$

El valor numérico de la entropía depende de la base del logaritmo, y si no se especifica lo suponemos en base 2. Si la base del logaritmo es b, entonces se denota la entropía como $H_b(X)$ y su unidad de medida se denomina unidad de base b. Si la base queda especificada previamente, en este trabajo se escribirá $H(X)$ independientemente de cual sea. Para ciertos valores de la base b, las unidades reciben un nombre especial; si $b = 2$ se denominan bits. Por ejemplo, la entropía del lanzamiento de una moneda es 1 bit. La relación de la entropía tomada en distintas bases no es más que un cambio de escala como veremos a continuación.

Cabe notar que la entropía es un funcional de la distribución de X, pues no depende de los valores concretos que pueda tomar X, sino solo de sus probabilidades. Por ello si $p = (p_1, p_2, \ldots, p_n)$ es un vector de probabilidades, la entropía también podemos denotarla por $H(p)$.

En la siguiente nota damos una visión que se usará repetidamente.

Nota. Si denotamos la esperanza matemática o valor esperado por E, consideramos $X \sim p(x)$ y una función g tal que $g(X)$ es una variable aleatoria. Entonces el valor esperado de $g(X)$ lo escribiremos de la forma

$$Eg(X) = \sum_{x \in X} g(x)p(x)$$

De esta manera si tomamos $g(x) = log\frac{1}{p(x)} (= -log\, p(x))$, la entropía de X puede ser interpretada como la esperanza de la variable aleatoria $g(X)$. Esto es

$$H(X) = E\, log\frac{1}{p(x)} = -E\, log\, p(x)$$

Algunas consecuencias directas de la definición de entropía son las siguientes

Lema 32. *Sea X variable aleatoria discreta con función de probabilidad $p(x)$. Se cumple*

1. $H(X) \geq 0$
2. $H_b(X) = (log_b a)H_a(X)$

Demostración. 1. Como $0 \leq p(x) \leq 1$ entonces $log(\frac{1}{p(x)}) \geq 0$ y se deduce lo buscado.
2. Basta conocer el cambio de base del logaritmo $log_b p = (log_b a) log_a p$ □

Por la primera propiedad, vemos que la cota inferior de la entropía es el cero, y únicamente se alcanza cuando solo hay un suceso de probabilidad no nula, es decir, cuando no hay aleatoriedad. En cuanto a la cota superior, se puede ver en [LoVei] que es una función creciente del número de sucesos de probabilidad no nula. Tal cota solo se alcanza para una distribución de probabilidad uniforme, es decir, una distribución uniforme representa la incertidumbre máxima (es aquella en la que la probabilidad de acertar lo que va a ocurrir es mínima).
Por la segunda propiedad, la entropía puede entonces ser cambiada de una base a otra, simplemente multiplicando por un escalar.

Ponemos a continuación un ejemplo del uso de la entropía.

Ejemplo 33. Sea X la variable aleatoria discreta que toma los valores

$$X = \begin{cases} a & \text{con probabilidad } 1/2 \\ b & \text{con probabilidad } 1/4 \\ c & \text{con probabilidad } 1/8 \\ d & \text{con probabilidad } 1/8 \end{cases}$$

Entonces la entropía de X es

$$H(X) = -\frac{1}{2}log\frac{1}{2} - \frac{1}{4}log\frac{1}{4} - \frac{1}{8}log\frac{1}{8} - \frac{1}{8}log\frac{1}{8} = \frac{7}{4} \text{ bits.}$$

Como decíamos, la entropía mide cuánta incertidumbre se tiene sobre una variable aleatoria X. Supongamos que queremos determinar el valor de X con el mínimo número posible de preguntas de sí o no (preguntas binarias). Teniendo en cuenta la distribución de probabilidad, como la opción más probable entre las cuatro es "a", la pregunta más eficiente que podemos hacer en primer lugar es si $X = a$. Esto divide las posibilidades en dos. Si la respuesta es "sí", hemos terminado, si es "no", como el suceso más probable entre los restantes es $X = b$, la segunda pregunta que haríamos es "¿Es $X = b$?". Esto de nuevo se divide en dos casos, si la respuesta es negativa, la tercera (y última) pregunta puede ser "¿Es $X = c$?". Así, el número esperado de preguntas binarias para

determinar X (que es de hecho el mínimo esperado) es $7/4 = 1{,}75$. Nosotros no lo demostraremos, pero se puede ver en [CoJo] que el mínimo número esperado de preguntas binarias requeridas para determinar X está entre $H(X)$ y $H(X)+1$.

Se muestra otro ejemplo del uso de la entropía que se repetirá en numerosas ocasiones.

Ejemplo 34. Sea X una variable aleatoria discreta que toma valores en $\mathbb{F}_q^l$ siguiendo una distribución uniforme, es decir se tiene

$$p(X = x) = \frac{1}{|\mathbb{F}_q^l|} = \frac{1}{q^l}, \forall x \in \mathbb{F}_q^l$$

Entonces la entropía de X medida en unidades de base q es

$$H_q(X) = -\sum_{x \in \mathbb{F}_q^l} p(x) log_q\, p(x) = -log_q\, \frac{1}{q^l} = l$$

3.1.1. Entropía conjunta y entropía condicional

Extendemos a continuación la definición de entropía para dos variables aleatorias. Sean X e Y variables aleatorias discretas de rango discreto y finito $\mathcal{X}$ e $\mathcal{Y}$, y funciones de probabilidad $p(x)$ y $p(y)$ respectivamente. Consideramos conjuntamente ambas variables, y tenemos una nueva variable aleatoria bidimensional (X, Y) con rango finito y discreto $\mathcal{X} \times \mathcal{Y}$. La entropía de la variable (X, Y) viene dada por la definición siguiente.

Definición 35. *La entropía conjunta $H(X, Y)$ del par de variables aleatorias discretas (X, Y) con función de probabilidad conjunta $p(x, y)$ se define como*

$$H(X, Y) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) log(p(x, y))$$

que puede ser expresado como

$$H(X, Y) = -E\, log\, p(X, Y)$$

Por supuesto la anterior definición puede extenderse para un número arbitrario de variables aleatorias discretas.
Una propiedad que usaremos más adelante es la siguiente.

Teorema 36. *Sean $X \sim p(x)$ e $Y \sim p(y)$, y $p(x, y)$ su función de probabilidad conjunta. Se verifica*

$$H(X, Y) \leq H(X) + H(Y)$$

La igualdad se da si y solo si las variables son independientes.

Demostración. Conociendo las probabilidades marginales $p(x) = p(X = x) = \sum_{y \in \mathcal{Y}} p(x, y)$ y $p(y) = p(Y = y) = \sum_{x \in \mathcal{X}} p(x, y)$ se cumple

$$H(X) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) log\, p(x)$$
$$H(Y) = -\sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} p(x, y) log\, p(y)$$

Por tanto

$$\begin{aligned} H(X)+H(Y) &= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)log(p(x)p(y)) \\ &\geq -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)log\, p(x,y) \\ &= H(X,Y) \end{aligned}$$

La igualdad se da si y solo si $p(x,y) = p(x)p(y), \forall x \in \mathcal{X}, y \in \mathcal{Y}$, es decir, si X e Y son independientes. □

Dadas las variables aleatorias X e Y, si representamos por $p(y|x) = p(Y = x|X = x)$ los valores de la distribución de probabilidades conociendo X de la variable aleatoria Y, y denotamos como $H(Y|X = x)$ la entropía de dicha distribución. Entonces dicha entropía será

$$H(Y|X=x) = -\sum_{y\in\mathcal{Y}} p(y|x)log\, p(y|x)$$

Así para cada valor x de la variable X se tiene una entropía $H(Y|X = x)$. Por lo que tal entropía puede verse como una función de la variable X, y su valor esperado es lo que denominamos entropía condicional como recoge la siguiente definición.

Definición 37. *Sea $(X,Y) \sim p(x,y)$, la entropía condicional $H(Y|X)$ se define como*

$$\begin{aligned} H(Y|X) &= \sum_{x\in X} p(x)H(Y|X=x) \\ &= -\sum_{x\in X} p(x)\sum_{y\in Y} p(y|x)log\, p(y|x) \\ &= -\sum_{x\in X}\sum_{y\in Y} p(x,y)log\, p(y|x) \\ &= -E\, log\, p(Y|X). \end{aligned}$$

Las definiciones de entropía conjunta y entropía condicional se relacionan por el siguiente teorema conocido como regla de la cadena.

Teorema 38 (Regla de la cadena).

$$H(X,Y) = H(X) + H(Y|X)$$

Demostración. Usando las definiciones ya introducidas, y que $p(x,y) = p(x)p(y|x)$ tenemos

$$\begin{aligned} H(X,Y) &= -\sum_{x\in X}\sum_{y\in Y} p(x,y)log(p(x,y)) \\ &= -\sum_{x\in X}\sum_{y\in Y} p(x,y)log(p(x)p(y|x)) \\ &= -\sum_{x\in X}\sum_{y\in Y} p(x,y)log\, p(x) - \sum_{x\in X}\sum_{y\in Y} p(x,y)log\, p(y|x) \\ &= -\sum_{x\in X} p(x)log\, p(x) - \sum_{x\in X}\sum_{y\in Y} p(x,y)log\, p(y|x) \\ &= H(X) + H(Y|X). \end{aligned}$$

De forma equivalente, si tomamos logaritmos en la expresión $p(X,Y) = p(X)p(Y|X)$ y después el valor esperado se obtiene el teorema

$$\begin{aligned} log\, p(X,Y) &= log\, p(X) + log\, p(Y|X) \\ E\, log\, p(X,Y) &= E\, log\, p(X) + E\, log\, p(Y|X) \\ H(X,Y) &= H(X) + H(Y|X) \end{aligned}$$

□

Corolario 39.

$$H(X,Y|Z) = H(X|Z) + H(Y|X,Z)$$

Análogamente tenemos el resultado para más de dos variables recogido en el siguiente teorema.

Teorema 40. *Sean las variables aleatorias discretas $X_1, X_2, \ldots, X_n$ con función de probabilidad $p(x_1, x_2, \ldots, x_n)$, se verifica*

$$H(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} H(X_i|X_{i-1}, \ldots, X_1)$$

Demostración. Aplicando repetidamente la regla de la cadena para dos variables, y el corolario 39 tenemos

$$\begin{aligned} H(X_1, X_2) &= H(X_1) + H(X_2|X_1) \\ H(X_1, X_2, X_3) &= H(X_1) + H(X_2, X_3|X_1) \\ &= H(X_1) + H(X_2|X_1) + H(X_3|X_2, X_1) \\ &\vdots \\ H(X_1, X_2, \ldots, X_n) &= H(X_1) + H(X_2|X_1) + \cdots + H(X_n|X_{n-1}, \ldots, X_1) \\ &= \sum_{i=1}^{n} H(X_i|X_{i-1}, \ldots, X_1) \end{aligned}$$

□

Escribimos a continuación un ejemplo en el que ver la entropía conjunta y condicionada.

Ejemplo 41. Sea el par de variables aleatorias (X,Y) cuya distribución conjunta viene recogida en la siguiente tabla

	X				
Y		1	2	3	4	
	1	1/8	1/16	1/32	1/32	1/4
	2	1/16	1/8	1/32	1/32	1/4
	3	1/16	1/16	1/16	1/16	1/4
	4	1/4	0	0	0	1/4
		1/2	1/4	1/8	1/8	1

Al tomar X e Y valores en el conjunto $\{1,2,3,4\}$ de tamaño 4, tiene sentido considerar la entropía H_4 cuya unidad de medida (unidad de base 4) la llamaremos simplemente paquete. Sin embargo hacemos los cálculos en bits por simplicidad y después expresamos los resultados en paquetes

multiplicando por el factor $log_4 2 = 2$. Viendo las distribuciones marginales de X $(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8})$ y de Y $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$, calculamos $H(X)$ y $H(Y)$

$$H(X) = -\sum_{x=1}^{4} p(X=x) log\, p(X=x) = -\frac{1}{2}log\frac{1}{2} - \frac{1}{4}log\frac{1}{4} - \frac{1}{8}log\frac{1}{8} - \frac{1}{8}log\frac{1}{8} = \frac{7}{4} \text{ bits} = \frac{7}{2} \text{ paquetes}$$

$$H(Y) = -\sum_{y=1}^{4} p(Y=y) log\, p(Y=y) = -\frac{1}{4}log\frac{1}{4} - \frac{1}{4}log\frac{1}{4} - \frac{1}{4}log\frac{1}{4} - \frac{1}{4}log\frac{1}{4} = 2 \text{ bits} = 4 \text{ paquetes}$$

Usando la notación $H(p_1, p_2, p_3, p_4) = -\sum_{i=1}^{4} p_n log\, p_n$, calculemos la entropía condicional $H(Y|X)$.

$$\begin{aligned} H(Y|X) &= \sum_{x=1}^{4} p(X=x) H(Y|X=x) \\ &= \frac{1}{2}H(\frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{2}) + \frac{1}{4}H(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}, 0) + \frac{1}{8}H(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}, 0) + \frac{1}{8}H(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}, 0) \\ &= \frac{1}{2}\cdot\frac{7}{4} + \frac{1}{4}\frac{3}{2} + \frac{1}{8}\cdot\frac{3}{2} + \frac{1}{8}\cdot\frac{3}{2} \\ &= \frac{13}{8} \text{ bits} = \frac{13}{4} \text{ paquetes} \end{aligned}$$

De igual manera calculamos $H(X|Y) = \frac{11}{8}$ bits $= \frac{11}{4}$ paquetes, y usando la igualdad $H(X,Y) = H(X) + H(Y|X)$ obtenemos la entropía conjunta $H(X,Y) = \frac{27}{8}$ bits $= \frac{27}{4}$ paquetes.

Nota. Cabe darse cuenta de que $H(Y|X) \neq H(X|Y)$, sin embargo si se da la igualdad $H(X) - H(X|Y) = H(Y) - H(Y|X)$.

3.2. Información mutua

La entropía $H(X)$ de una variable aleatoria X es una medida de la incertidumbre a priori del experimento aleatorio que representa dicha variable, y vimos que $H(X|Y)$ es una medida de la incertidumbre tras conocer otra información. Así, la diferencia $H(X) - H(X|Y)$ será la reducción de incertidumbre debida al conocimiento del resultado de la variable Y. Por eso se dice que dicha diferencia es la información que la variable Y nos aporta sobre la variable X. A esta diferencia es lo que se conoce como información mutua.

Definición 42. *Sean las variables aleatorias X e Y, se define información mutua $I(X;Y)$ como*

$$I(X;Y) = H(X) - H(X|Y)$$

Proposición 43. *Dadas las variables aleatorias discretas X e Y, se verifica*

1. *$I(X;Y) \geq 0$ y $I(X;Y) = 0 \Leftrightarrow X$ e Y son independientes.*
2. *$I(X;X) = H(X)$*
3. *$I(X;Y) = I(Y;X)$*

Demostración. 1. Por la regla de la cadena sabemos $H(X|Y) = H(X,Y) - H(Y)$, y gracias al teorema 36, $H(X,Y) \leq H(X) + H(Y)$, luego $H(X|Y) = H(X,Y) - H(Y) \leq H(X) + H(Y) - H(Y)$ y se concluye $H(X|Y) \leq H(X)$.

2. Por definición $I(X;X) = H(X) - H(X|X)$ y $H(X|X) = 0$ se deduce de su definición.
3. Resulta que

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) = H(X) + H(Y) - (H(Y) + H(X|Y)) \\ &= H(X) + H(Y) - (H(X) + H(Y|X)) \\ &= H(Y) - H(Y|X) \\ &= I(Y;X) \end{aligned}$$

de donde en la primera igualdad se suma y resta $H(Y)$ y en la segunda se usa la regla de la cadena. □

La proposición anterior nos dice que X nos da tanta información de Y como la que Y nos da sobre X, es por esto que esa información se denomina información mutua.
Otros resultados que se derivan de las propiedades demostradas hasta ahora son las siguientes.

Proposición 44. *Dadas las variables aleatorias discretas X e Y, se verifica*

- $I(X;Y) \leq min\{H(X), H(Y)\}$
- $I(X;Y) = H(X) + H(Y) - H(X,Y)$

Demostración. 1. Se deduce de la simetría de I

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \leq H(X) \\ I(X;Y) &= H(Y) - H(Y|X) \leq H(Y) \end{aligned}$$

2. Se tiene usando la regla de la cadena

$$I(X;Y) = H(X) - H(X|Y) = H(X) + H(Y) - H(X,Y)$$

□

Ejemplo 45. En el ejemplo 41, la información mutua entre las variables X e Y es

$$I(X;Y) = H(X) - H(X|Y) = \frac{7}{4} - \frac{11}{8} = \frac{3}{8} \text{ bits.}$$

3.2.1. Información mutua condicional

Conocer una tercera variable Z, puede hacer variar la distribución marginal o conjunta de dos variables aleatorias dadas X e Y. Definimos ahora la información mutua condicional como la reducción de la incertidumbre de X a causa del conocimiento de Y, cuando Z viene dado.

Definición 46. *La información mutua condicional de las variables aleatorias X e Y, dada la variable Z es*

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z)$$

La información mutua conserva todas las propiedades de la información mutua. En particular, satisface una regla de la cadena.

Teorema 47. *Sean las variables aleatorias discretas X, Y, Z, se cumple*

$$I(X,Y;Z) = I(X;Z) + I(Y;Z|X)$$

Demostración.

$$\begin{aligned} I(X,Y;Z) &= H(X,Y) - H(X,Y|Z) \\ &= H(X) + H(Y|X) - (H(X|Z) + H(Y|X,Z)) \\ &= (H(X) - H(X|Z)) + (H(Y|X) - H(Y|X,Z)) \\ &= I(X;Z) + I(Y;Z|X) \end{aligned}$$

Habiendo usado la regla de la cadena y el corolario 39 en la segunda igualdad. □

Por supuesto se puede extender la regla para cualquier número de variables como mostramos a continuación.

Teorema 48. *Sean las variables aleatorias discretas $X_1, X_2, \ldots, X_n, Y$, se verifica*

$$I(X_1, X_2, \ldots, X_n; Y) = \sum_{i=1}^{n} I(X_i; Y | X_{i-1}, X_{i-2}, \ldots, X_1)$$

Demostración.

$$\begin{aligned} I(X_1, X_2, \ldots, X_n; Y) &= H(X_1, X_2, \ldots, X_n) - H(X_1, X_2, \ldots, X_n|Y) \\ &= \sum_{i=1}^{n} H(X_i|X_{i-1}, \ldots, X_1) - \sum_{i=1}^{n} H(X_i|X_{i-1}, \ldots, X_1, Y) \\ &= \sum_{i=1}^{n} I(X_i; Y|X_1, X_2, \ldots, X_{i-1}) \end{aligned}$$

□

Capítulo 4

Seguridad en Network Coding

En este capítulo nos centraremos en los problemas de seguridad que surgen al utilizar network coding. Nuestro interés principal es estudiar cómo de segura es una comunicación en la que se aplica network coding lineal, contra un espía en la red. Necesitamos crear un esquema de codificación que maximice la incertidumbre del espía sobre los mensajes enviados, consiguiendo que llegue a los receptores la información enviada originalmente. Además, este esquema será universal, es decir, el código usado para codificar los mensajes, y el código de red (la elección de las combinaciones entre paquetes que se hacen a lo largo de la red) son independientes entre sí. La descripción que damos a continuación sobre una red susceptible de ser espiada, tiene sus orígenes en el canal espiado de Ozarow y Wyner [OzWy], o a partir de ahora, Wiretap Channel tipo II. En [RoSol] muestran que el problema de seguridad en network coding puede verse como una generalización a una red del Wiretap Channel II, sin embargo es en [SiKsch] donde se consigue construir un esquema universal realizando algunas modificaciones, como por ejemplo sustituyendo los códigos MDS que utiliza [RoSol] para la codificación, por códigos MRD.

4.1. Wiretap Channel tipo II

Este modelo considera una comunicación punto a punto, es decir un emisor y un receptor, unidos por un canal o arista. Sobre este canal hay un espía o intruso, que es capaz de observar cierta información. Se busca que la comunicación sea segura teniendo en cuenta que no se quiere usar una clave criptográfica.
El emisor genera un mensaje $S = (S_1, S_2, \ldots, S_k) \in \mathbb{F}_{q^m}^k$ que quiere enviar de forma segura, y se codifica en una palabra $X = (X_1, X_2, \ldots, X_n) \in \mathbb{F}_{q^m}^n$. Esta palabra se transmite por el canal, y el espía observa $\mu \leq n$ símbolos de X, representados por la palabra $W = (X_i,\, i \in \mathcal{I})$ donde $\mathcal{I} \subseteq \{1, \ldots, n\}$ tiene cardinal μ. Para este modelo muestran en [OzWy] que el número máximo de símbolos que se pueden transmitir de manera segura es $n-\mu$. El objetivo es hacer llegar el mensaje al receptor sin que el espía obtenga ninguna información de S, sin importar qué μ símbolos intercepte. Este concepto de comunicación segura, se representa por las siguientes dos condiciones

$$H(S|X) = 0 \tag{4.1}$$

$$I(S;W) = 0, \qquad \forall \mathcal{I} : |\mathcal{I}| = \mu \tag{4.2}$$

Donde el logaritmo se toma en base q^m, por lo que la información se miden unidades q^m-arias o paquetes. La igualdad (4.1) se traduce en que dado X, no se tiene incertidumbre sobre S, pues la decodificación ha de ser única. Es decir, S está completamente determinado por el conocimiento de X. Por (4.2) entendemos que el hecho de conocer W no disminuye la incertidumbre sobre S, es

decir, conocer W no aporta al espía información extra sobre S. Un esquema de comunicación que cumpla la primera condición, se dice que es una comunicación de *error-cero*, y si cumple la segunda condición que es de *secreto perfecto*. El objetivo es codificar S en X cumpliendo ambas condiciones.

4.1.1. Esquema de codificación por cogrupo

Se presenta ahora el esquema de codificación por cogrupo (coset coding scheme) de Ozarow y Wyner usado para este modelo, y sobre el que trabajaremos en varias ocasiones. Para ello, repasamos el concepto de cogrupo y algunas propiedades importantes.

Definición 49. *Dado $\mathcal{C}$ un código lineal $[n,k]_q$ con matriz de control $H_{n-k\times n}$ y un elemento $x \in \mathbb{F}_q^n$, se define el síndrome de x como*

$$s(x) = Hx^t \in \mathbb{F}_q^{n-k}$$

De la anterior definición se deduce que x pertenece al código $\mathcal{C}$ si y solo si su síndrome es el vector nulo.

Definición 50. *Dado $\mathcal{C}$ un código lineal $[n,k]_q$, y un elemento $a \in \mathbb{F}_q^n$, llamamos cogrupo que contiene al elemento a al conjunto*

$$a + \mathcal{C} = \{a + c \mid c \in \mathcal{C}\}$$

Dado un código $\mathcal{C} \subseteq \mathbb{F}_q^n$ de dimensión k, hay q^{n-k} cogrupos de $\mathcal{C}$ en $\mathbb{F}_q^n$ (siendo $\mathcal{C}$ uno de ellos). Los cogrupos forman una partición de $\mathbb{F}_q^n$, es decir son disjuntos y su unión compone todo el espacio. Además, cada cogrupo contiene el mismo número de vectores, este es q^k.

Proposición 51. *Dado un código lineal $\mathcal{C}$, dos elementos de $\mathbb{F}_q^n$ pertenecen al mismo cogrupo si y solo si tienen el mismo síndrome.*

Demostración. $\Rightarrow$) Sea $a \in \mathbb{F}_q$ y sean $x, y \in (a + \mathcal{C})$. Esto es, $x = a + c$ para cierto $c \in \mathcal{C}$, y $y = a + c'$ con $c' \in \mathcal{C}$. Sea H matriz de control del código, sabiendo que $s(z) = Hz^t = 0, \forall z \in \mathcal{C}$ y que H puede verse como una aplicación lineal,veamos que x e y tienen el mismo síndrome: $s(x) = Hx^t = H(a+c)^t = Ha^t = Ha^t + Hc'^t = H(a+c')^t = s(y)$.
$\Leftarrow$) Como x e y tienen el mismo síndrome, $Hx^t = Hy^t$ y se tiene $H(x-y)^t = 0$. Esto significa que $x - y \in \mathcal{C}$. Luego como $x = x + 0$ e $y = x + (y - x)$, se tiene que x e y están en el mismo cogrupo $(x + \mathcal{C})$. □

Si fijamos $k = n - \mu$, y sea $\mathcal{C}$ un $[n, \mu]$ código lineal MDS con matriz de control $H \in \mathbb{F}_{q^m}^{k\times n}$. La codificación se lleva a cabo eligiendo aleatoriamente (siguiendo una distribución uniforme) $X \in \mathbb{F}_{q^m}^n$ tal que $S = HX$ (considerando S y X como columnas). Es decir, el mensaje S lo vemos como un síndrome (el tamaño es adecuado, es un vector de $k = n - \mu$ entradas); este síndrome gracias a la proposición 51 determina un cogrupo, este es el cogrupo cuyos elementos tienen síndrome S. De este cogrupo se elige aleatoriamente un elemento X, que será la palabra que codifica a S y se transmite por la red.
En cuanto a la decodificación, recibida la palabra X se obtiene el mensaje querido calculando el síndrome $S = HX$. Como se comentó anteriormente, si espían μ símbolos, se pueden transmitir de forma segura como mucho $k = n - \mu$ símbolos. Es decir, si queremos comunicar con seguridad k símbolos con un código MDS $[n, n-k]$, necesitamos que no se espíen más de $\mu = n - k$ símbolos.

Vamos a demostrar que el esquema descrito es seguro, es decir cumple (4.1) y (4.2).
Por un lado, es trivial que $H(S|X) = 0$, pues conocido X, el mensaje S está completamente

determinado por $S = HX$.
Para ver que $I(S;W) = 0$, nos servimos de la simetría de I, el teorema 47 y las definiciones vistas en el capítulo 3. Por el teorema 47 se tiene $I(S;W) = I(S,X;W) - I(X;W|S)$. Ahora $I(S,X;W) = I(X,S;W)$ y volviendo a aplicar el mismo teorema se verifica $I(X,S;W) = I(X;W) + I(S;W|X)$. Uniendo esto tenemos

$$\begin{aligned} I(S;W) &= I(S,X;W) - I(X;W|S) = I(X;W) + I(S;W|X) - I(X;W|S) && (4.3)\\ &= H(W) - H(W|X) + H(W|X) - H(W|S,X) - H(X|S) + H(X|W,S) && (4.4)\\ &= H(W) - H(X|S) + H(X|W,S) && (4.5)\\ &= H(W) - \mu + H(X|W,S) && (4.6)\\ &\leq H(X|W,S) = 0 && (4.7) \end{aligned}$$

A (4.5) se llega sabiendo $H(W|S,X) = 0$ pues W es función de X: la elección de μ símbolos de X (de hecho, más adelante escribiremos $W = BX$ para cierta matriz B). El paso a (4.6) se sigue pues $H(X|S) = -\sum_s \sum_x p(x,s) log(p(x|s)) = -log\, p(x|s) = log((q^m)^\mu) = \mu$ ya que se elige X siguiendo una distribución uniforme sobre los elementos del cogrupo determinado por S, que tiene tamaño $|\mathbb{F}_q^m|^\mu = (q^m)^\mu$. La desigualdad (4.7) se cumple ya que $H(W) \leq \mu$; por último se verifica $H(X|W,S) = 0$ pues en un código (n,μ) MDS, una palabra X queda determinada si conocemos el cogrupo al que pertenece (S), y μ de sus símbolos.

Se muestra un ejemplo de comunicación por un canal espiado usando la codificación por cogrupo.

Ejemplo 52. Consideremos los parámetros $n = 2$, $k = 1$ y $\mu = 1$. El código MDS sobre $\mathbb{F}_2$ que usamos para la codificación por cogrupo es $\mathcal{C} = \{(X_1, X_1) : X_1 \in \mathbb{F}_2\}$ definido por las matrices generadora y de control

$$G = H = \begin{bmatrix} 1 & 1 \end{bmatrix}$$

Queremos transmitir un bit $S \in \{0,1\}$ y vemos dicho mensaje como un síndrome, que determina dos cogrupos en $\mathbb{F}_2^2$. Así, si $S = 0$ se codificará y enviará por el canal en la palabra $(0,0)$ o $(1,1)$ con igual probabilidad, y si $S = 1$, la palabra será $(1,0)$ o $(0,1)$ de forma equiprobable.

Mensaje S	0	1
Elementos (X_1, X_2) del cogrupo	$\{(0,0),(1,1)\}$	$\{(1,0),(0,1)\}$

Conocer X_1 o X_2 (es decir $\mu = 1$) no proporciona información sobre S, sin embargo conocer ambos (X_1, X_2) (es decir $\mu = 2$) es suficiente para determinar $S = X_1 + X_2$.

4.2. Wiretap Network

El modelo que acabamos de presentar en el que solo se considera una comunicación por una arista, puede ser extendido a un escenario de network coding lineal introduciendo un espía en una red como las que se describen en el capítulo 1. En este caso, el espía es capaz de observar $\mu < n$ aristas de la red, recogidas en el conjunto $\mathcal{I}$. De forma análoga a como ocurre en el modelo del canal espiado, la tasa máxima de transmisión segura es de $n - \mu$ paquetes, lo cual se demostrará más adelante. En el primer capítulo, considerábamos un mensaje a enviar como un vector $X = (X_1, \ldots, X_n) \in \mathbb{F}_q^n$ cuyas componentes, elementos de $\mathbb{F}_q$, se combinaban y transmitían por las aristas de la red. A partir de ahora se quiere tomar como cuerpo finito la extensión $\mathbb{F}_{q^m}$ de grado m sobre $\mathbb{F}_q$, y pensar en cada componente $X_i \in \mathbb{F}_{q^m}$ del mensaje como un paquete $[X_i]_\mathcal{B} \in \mathbb{F}_q^m$. Así, vemos un mensaje indistintamente como un vector $X \in \mathbb{F}_{q^m}^n$ o como una matriz $[X]_\mathcal{B} \in \mathbb{F}_q^{n \times m}$ cuyas filas son los paquetes a transmitir.

Una red en la que usemos network coding lineal que permita transmitir con éxito n paquetes de longitud m a todos sus receptores diremos que es una red lineal multidifusión o código de red lineal $(n, m)_q$.

Definición 53. *Un esquema de comunicación consiste en un código de red $(n, m)_q$ que determina las operaciones de network coding en la red, y un código en bloque lineal $[n, \mu]$ que codifica el mensaje a enviar por el emisor.*

Se resume en la figura 4.1 en qué consiste un esquema de comunicación. Si se quiere enviar un mensaje $S \in \mathbb{F}_{q^m}^k$, el emisor lo codifica en $X \in \mathbb{F}_{q^m}^n$ usando una codificación por cogrupo, y se transmite por la red aplicando network coding. Así, si un intruso observa los paquetes en μ aristas, estos se pueden representar por $W = BX \in \mathbb{F}_{q^m}^{\mu}$, donde $B \in \mathbb{F}_{q^m}^{\mu \times n}$ es la matriz cuyas filas son los vectores de codificación global de las aristas en $\mathcal{I}$.

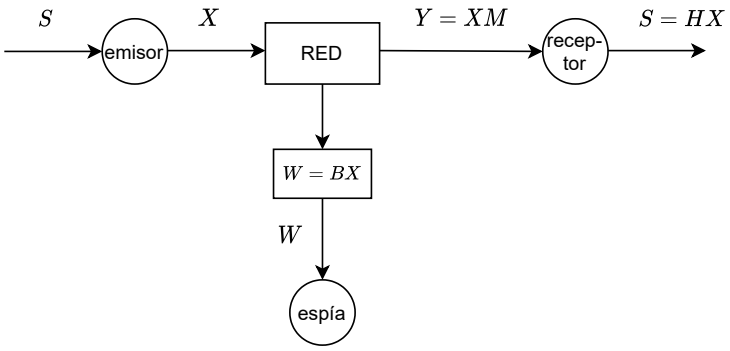


Figura 4.1: Esquema de comunicación

Escribimos las condiciones de seguridad para una comunicación en red, análogamente a como se hizo para un canal

$$H(S|Y) = 0 \tag{4.8}$$
$$I(S; W) = 0, \ \ \forall \mathcal{I} : |\mathcal{I}| = \mu \tag{4.9}$$

Con vistas al esquema que queremos construir, en lo que sigue se reescribe la condición (4.8), y se trabaja sobre la condición (4.9), proponiendo una formulación equivalente, y escribiendo en términos más concretos de la red y el código, una condición suficiente para que esta se cumpla.

Por un lado, puesto que consideramos un código de red válido para la transmisión de un mensaje, es decir, X puede ser recuperado al recibir Y, la condición $H(S|Y) = 0$ puede ser sustituida por $H(S|X) = 0$. Lo que esto significa es que, para que Y determine completamente S (es decir se pueda descodificar unívocamente), es suficiente que lo haga X, ya que la red recupera X unívocamente al llegar Y $(H(X|Y) = 0)$.

Por otro lado, hacer filas de B linealmente dependientes de otras no aumenta la información que proporciona W sobre S $(I(S; W))$, por lo que basta considerar las matrices B de rango máximo en la condición (4.9). Así, sustituimos dicha condición por la equivalente dada por

$$I(S; W) = 0, \ \ \forall \mathcal{I} : |\mathcal{I}| = \mu, \ rg(B) = \mu \tag{4.10}$$

Damos ahora un resultado que proporciona una condición suficiente para satisfacer (4.10), y que nos será de utilidad en resultados posteriores.

Proposición 54. *Sea $\mathbb{F}$ un cuerpo finito, una matriz $H \in \mathbb{F}^{k \times n}$, y $B \in \mathbb{F}^{\mu \times n}$. Denotemos $\mathcal{S} = \{Hx : x \in \mathbb{F}^n\}$ y $\mathcal{X}_s = \{x \in \mathbb{F}^n : s = Hx\}$. Por último, sean $S \in \mathcal{S}$, $X \in \mathbb{F}^n$ y $W = BX$ variables aleatorias. Entonces*

1. *Dado $S = s$, si X sigue una distribución uniforme sobre $\mathcal{X}_s$, entonces*

$$rg\left(\begin{bmatrix} H \\ B \end{bmatrix}\right) = rg(H) + rg(B) \Rightarrow I(S;W) = 0$$

2. *Si S sigue una distribución uniforme sobre $\mathcal{S}$, entonces*

$$I(S;W) = 0 \Rightarrow rg\left(\begin{bmatrix} H \\ B \end{bmatrix}\right) = rg(H) + rg(B)$$

Demostración. 1. Sea $\mathcal{W} = \{Bx : x \in \mathbb{F}^n\}$ y se define

$$\mathcal{X}_{s,w} = \{x \in \mathbb{F}^n : \begin{bmatrix} s \\ w \end{bmatrix} = \begin{bmatrix} H \\ B \end{bmatrix} x\}$$

Se prueban primero las siguientes desigualdades

a) $H(W) \leq log_{|\mathbb{F}|}|\mathcal{W}| = rg(B)$

b) $H(X|S) = log_{|\mathbb{F}|}|\mathcal{X}_S| = dim(ker(H)) = n - rg(H)$

c) $H(X|S,W) \leq log_{|\mathbb{F}|}|\mathcal{X}_{S,W}| = n - rg\left(\begin{bmatrix} H \\ B \end{bmatrix}\right)$

Para probar *a*), recordemos que la entropía es máxima cuando se considera una distribución de probabilidad uniforme. Así

$$H(W) = -\sum_{w \in \mathcal{W}} p(w) log_{|\mathbb{F}|} p(w) \leq log_{|\mathbb{F}|}|\mathcal{W}|$$

y si vemos B como una aplicación lineal, se tiene que $|\mathcal{W}| = |\mathbb{F}|^{dim(Im(B))} = |\mathbb{F}|^{rg(B)}$; con lo que se concluye

$$H(W) \leq log_{|\mathbb{F}|}|\mathbb{F}|^{rg(B)} = rg(B).$$

Para probar *b*) notemos que los conjuntos $\mathcal{X}_s$ tienen el mismo tamaño para todo $s \in \mathcal{S}$, el razonamiento es similar al seguido en la proposición 51. Esto es porque podemos ver $ker(H)$ como un cogrupo en $\mathbb{F}^n$, y cada $s \in \mathcal{S}$ como un "síndrome", por lo que los conjuntos $\mathcal{X}_s$ son también cogrupos. Así, hay $|\mathbb{F}|^{n-dim(Ker(H))}$ conjuntos, cada uno de tamaño $|\mathcal{X}_s| = |\mathbb{F}|^{dim(Ker(H))} = |\mathbb{F}|^{n-rg(H)}$. Teniendo esto en cuenta, y que por hipótesis X es uniforme en $\mathcal{X}_s$, se consigue lo buscado como sigue

$$\begin{aligned} H(X|S) &= -\sum_{s \in S}\sum_{x \in X} p(s,x) log_{|\mathbb{F}|} p(x|s) = -\sum_{s \in S}\sum_{x \in \mathcal{X}_s} p(s,x) log_{|\mathbb{F}|} p(x|s) \\ &= \sum_{s \in S}\sum_{x \in \mathcal{X}_s} p(s,x) log_{|\mathbb{F}|}|\mathcal{X}_s| = log_{|\mathbb{F}|}|\mathcal{X}_s| = log_{|\mathbb{F}|}|\mathbb{F}|^{n-rg(H)} = n - rg(H). \end{aligned}$$

Por último, la prueba de *c*) es similar a la de *b*), teniendo en cuenta que X solo es uniforme en $\mathcal{X}_s$, y no necesariamente en $\mathcal{X}_{s,w}$, de ahí la desigualdad.

Conociendo *a*), *b*), *c*), y la igualdad $I(S;W) = H(W) - H(X|S) + H(X|W,S)$ vista en (4.5) se tiene

$$I(S;W) \leq rg(H) + rg(B) - rg\left(\begin{bmatrix} H \\ B \end{bmatrix}\right) \tag{4.11}$$

de donde se deduce el resultado buscado.
2. Si A es una matriz, denotemos por $\langle A\rangle$ el espacio generado por sus filas. Obviamente $rg(A) = dim(\langle A\rangle)$. Primero observemos que se tiene

$$dim(\langle H\rangle \cap \langle B\rangle) = rg(H) + rg(B) - rg\left(\begin{bmatrix} H \\ B \end{bmatrix}\right)$$

Esto es debido a que se cumple

$$\left\langle \begin{bmatrix} H \\ B \end{bmatrix} \right\rangle = \langle H\rangle + \langle B\rangle$$

Por lo tanto,

$$\begin{aligned} rg\left(\begin{bmatrix} H \\ B \end{bmatrix}\right) &= dim(\langle H\rangle + \langle B\rangle) \\ &= rg(H) + rg(B) - dim(\langle H\rangle \cap \langle B\rangle) \end{aligned}$$

Supongamos ahora que $dim(\langle H\rangle \cap \langle B\rangle) = t > 0$. Entonces existen matrices de rango máximo $T_1 \in \mathbb{F}^{t\times\mu}$ y $T_2 \in \mathbb{F}^{t\times k}$ tales que $T_1B = T_2H$ y $rg(T_2H) = t$. Esto implica que

$$T_1W = T_1BX = T_2HX = T_2S$$

Como S es uniforme, se tiene que $I(S;W) \geq H(T_2S) = t > 0$ y se deduce el resultado querido. □

Gracias al primer punto del anterior resultado, una comunicación con un código con matriz de control $H^{n-\mu\times n}$ en la que se espían μ aristas, verifica la condición de seguridad (4.10) si se cumple

$$rg\left(\begin{bmatrix} H \\ B \end{bmatrix}\right) = n, \ \ \forall \mathcal{I} : |\mathcal{I}| = \mu, \ rg(B) = \mu \tag{4.12}$$

O de forma equivalente, $\langle H\rangle \cap \langle B\rangle = \emptyset$. Es decir, que ninguna combinación lineal de μ o menos vectores de codificación global espiados pertenezcan al espacio $\langle H\rangle$.

Se muestra a continuación un resultado que anticipábamos al inicio de la sección: una red puede transmitir un máximo de $k = n-\mu$ paquetes de forma segura utilizando la codificación por cogrupo basada en un código $[n, \mu]$ MDS sobre $\mathbb{F}_q$. Para ello debe cumplirse la condición (4.12) entre la matriz de control del código y la red.

Teorema 55. *Sea $\mathcal{I} \subset E$ con $|\mathcal{I}| = \mu < n$ el conjunto de aristas observadas en una red que transmite n paquetes, y se denota por $W \in \mathbb{F}_q^\mu$ los paquetes espiados. Sea $B \in \mathbb{F}_q^{\mu\times n}$ cuyas filas son los vectores de codificación global de las aristas espiadas, y sean $S \in \mathbb{F}_q^k$ y $X \in \mathbb{F}_q^n$ las variables aleatorias que representan el mensaje original y codificado respectivamente a enviar. Entonces se pueden transmitir de forma segura $k \leq n - \mu$ paquetes.*

Demostración. El requisito de seguridad significa que $I(S;W) = H(S) - H(S|W) = 0$ para todo $\mathcal{I} \subset E$, y $H(S|X) = 0$. Consideremos el valor $H(W,S,X) = H(W,X,S)$, desarrollándolo de acuerdo al Teorema 40 se tiene

$$H(W) + H(S|W) + H(X|S,W) = H(W) + H(X|W) + H(S|X,W) \tag{4.13}$$
$$\Rightarrow H(X|S,W) = H(X|W) - H(S) \tag{4.14}$$
$$\Rightarrow 0 \leq n - rg(B) - k \tag{4.15}$$

Donde se llega a (4.14) sabiendo que $H(S|W) = H(S)$ y $H(S|X,W) = 0$ por el requisito de seguridad. Para pasar a (4.15) se usa que $H(S) = k$ y que $H(X|W) = n-rg(B)$, cuya demostración

es idéntica a la realizada en la Proposición 54 1.b).
Puesto que siempre existe una elección de aristas tal que $rg(B) = \mu$, la tasa de transmisión segura está acotada como

$$k \leq n - \mu$$

□

Ejemplo 56. Extendamos el esquema del ejemplo 52 de un canal a una red, en concreto a la red mariposa. Se tienen los parámetros $n = 2$, $k = 1$, $\mu = 1$, y se quiere enviar un bit $S \in \{0, 1\}$ siguiendo una codificación por cogrupo basada en la matriz $H = \begin{bmatrix} 1 & 1 \end{bmatrix}$, igual que en el ejemplo nombrado. Recordemos que si de la palabra codificada (X_1, X_2) el espía conoce ambos paquetes, el mensaje queda completamente determinado por $S = X_1 + X_2$.

Si se usa el código de red usual sobre $\mathbb{F}_2$ de la figura 4.2(a) y el espía pincha una arista ($\mu = 1$), basta que sea una de las aristas (v_3, v_4), (v_4, r_1) o (v_4, r_2) para conocer el mensaje. De hecho, es el propio código de red el que rompe la seguridad, "deshaciendo" en v_3 la codificación por cogrupo.

Sin embargo, si se usa el mismo código de red sobre $\mathbb{F}_3 = \{0, 1, 2\}$ pero cambiando la combinación de paquetes que hace v_3 por $X_1 + 2X_2$ como en la figura 4.2(b), obteniendo $\begin{bmatrix} 1 & 2 \end{bmatrix}$ como vector de codificación global de la arista (v_3, v_4), el espía no obtiene información sea cual sea la arista que observe. Aunque esto proporcione seguridad, no es lo que perseguimos, pues hemos necesitado cambiar el código de red y aumentar el tamaño del cuerpo en el que se trabaja.

Así, de acuerdo con (4.12) la comunicación basada en el código MDS con $H = \begin{bmatrix} 1 & 1 \end{bmatrix}$ será segura mientras el vector de codificación global de cualquier arista sea linealmente independiente de $\begin{bmatrix} 1 & 1 \end{bmatrix}$.

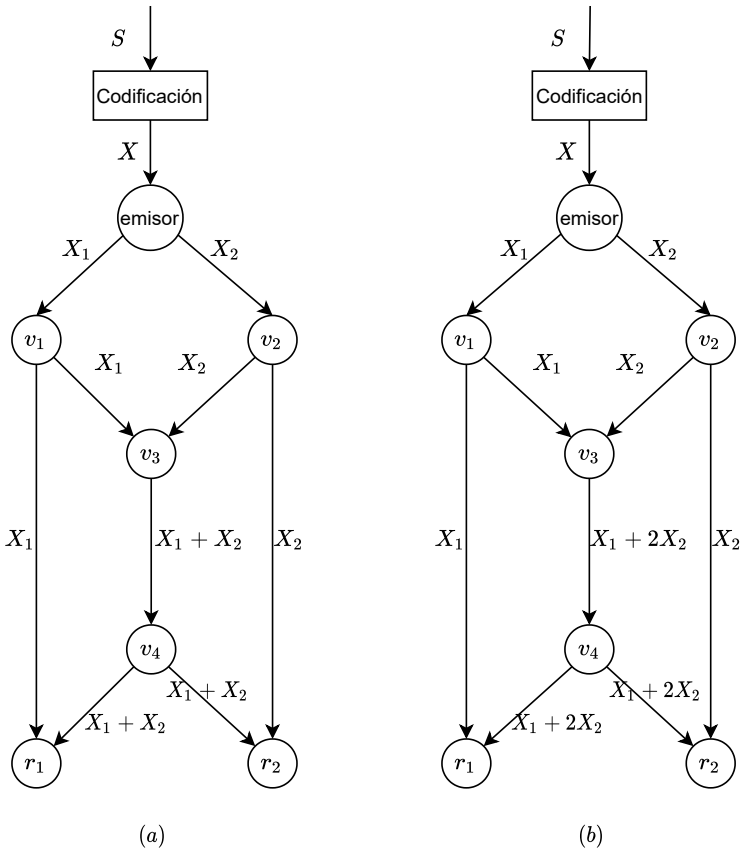


Figura 4.2: Red mariposa con un código de red seguro (a) e inseguro (b) con respecto a la codificación por cogrupo basada en $H = \begin{bmatrix} 1 & 1 \end{bmatrix}$

Como vemos, hay una dependencia entre el código MDS y el diseño de la red que puede comprometer la seguridad de la comunicación. Por ejemplo, no se podría aplicar random network coding, pues la red ha de ser cuidadosamente elegida teniendo en cuenta H para cumplir las restricciones.

Es por esto que en el modelo que describimos a continuación, sustituimos los códigos MDS por los MRD, en busca de un esquema universal.

4.2.1. Seguridad universal vía códigos MRD

El objetivo de lograr una comunicación segura que salve la problemática recién expuesta, motiva la siguiente definición.

Definición 57. *Un esquema seguro de comunicación se dice que es universal, si provee seguridad y fiabilidad (contra errores) para cualquier elección de código de red (network coding) que permita la comunicación. Es decir, debe verificar*

$$H(S|X) = 0 \tag{4.16}$$
$$I(S;W) = 0,\ \forall B \tag{4.17}$$

Donde la matriz B que representa las aristas espiadas, es irrelevante mientras estas sean como mucho μ.
Un esquema universal es entonces, un código lineal que pueda ser diseñado o elegido independientemente, y aplicado al inicio de cualquier código de red (es decir una red en la que se aplica network coding). Así, cualquier esquema universal es compatible con el random netwok coding, pues no necesitamos conocer la red para construir o elegir el código.
Con vistas en la condición (4.12), para cumplir la definición de esquema de seguridad universal basta que se satisfaga lo siguiente

$$H(S|X) = 0 \tag{4.18}$$
$$rg\left(\begin{bmatrix} H \\ B \end{bmatrix}\right) = n,\ \ \forall B:\ rg(B) = \mu \tag{4.19}$$

Las ideas clave del esquema universal que proponemos son las siguientes:

- Se toma como cuerpo la extensión de grado m de $\mathbb{F}_q$, esta es $\mathbb{F}_{q^m}$. Además, el grado de la extensión m tiene que ser al menos n (número de paquetes a enviar). Es decir, cada paquete del mensaje tiene que ser de tamaño $m \geq n$.

- La codificación por cogrupo utilizada se basa en la matriz de control de un código MRD sobre $\mathbb{F}_{q^m}$, en lugar de un MDS. Esto es compatible con aplicar network coding lineal sobre el cuerpo $\mathbb{F}_q$, ya que la codificación y descodificación solo se aplica en el emisor y los receptores, y las operaciones en la red están bien definidas al tratarse de una extensión de cuerpos ($\mathbb{F}_q \subseteq \mathbb{F}_{q^m}$).

- Considerar el código MRD sobre $\mathbb{F}_{q^m}$ y el código de red sobre $\mathbb{F}_q$, implica que la matriz de control H está definida sobre $\mathbb{F}_{q^m}$ mientras que la matriz B sigue teniendo sus entradas en $\mathbb{F}_q$. Esto hace que haya muchas más posibilidades para H en comparación con B, por lo que es más concebible que exista un H cumpliendo (4.19), y una eliminación de la dependencia entre el código y la red.

El resultado principal de este capítulo es consecuencia del siguiente teorema.

Teorema 58. *Sea $\mathcal{C}$ un código lineal $[n, \mu]$ sobre $\mathbb{F}_{q^m}$ con matriz de control $H \in \mathbb{F}_{q^m}^{n-\mu \times n}$. Entonces $\mathcal{C}$ es un código MRD (es decir, $d_R(\mathcal{C}) = n - \mu + 1$) con $m \geq n$ si y solo si la matriz*

$$M = \begin{bmatrix} H \\ B \end{bmatrix} \tag{4.20}$$

es no singular para todo $B \in \mathbb{F}_q^{\mu \times n}$ con $rg(B) = \mu$. Es decir $rg(M) = rg(H) + rg(B) = n$.

Demostración. Vamos a demostrar que $\mathcal{C}$ no es MRD (es decir, $d_R(\mathcal{C}) \leq n - \mu$) si y solo si existe alguna matriz $B \in \mathbb{F}_q^{\mu \times n}$ de rango máximo tal que M es singular.
$\Rightarrow$) Supongamos que $d_R(\mathcal{C}) \leq n - \mu$. Entonces existe alguna palabra $x \in \mathcal{C}$ no nula tal que $rg(x) \leq n - \mu$. Esto implica que existe alguna matriz $B \in \mathbb{F}_q^{\mu \times n}$ de rango máximo tal que $Bx = 0$. Como $x \in \mathcal{C}$ también se tiene $Hx = 0$. Entonces, $Mx = 0$ y M es singular.
$\Leftarrow$) Supongamos ahora que M es singular para algún $B \in \mathbb{F}_q^{\mu \times n}$ de rango máximo. Entonces existe $x \in \mathbb{F}_{q^m}^n$ no nulo tal que $Mx = 0$. Esto implica en particular que $Hx = 0$, es decir, $x \in \mathcal{C}$, y también $Bx = 0$, por lo que $rg(x) \leq n - \mu$. Por lo tanto, $d_R(\mathcal{C}) \leq n - \mu$. □

Estamos en condiciones ahora de presentar el resultado principal del capítulo.

Teorema 59. *Se considera un código de red lineal $(n, m)_q$ sujeto a μ observaciones. Sea $\mathcal{C}$ un código lineal $[n, \mu]$ sobre $\mathbb{F}_{q^m}$ con matriz de control $H \in \mathbb{F}_{q^m}^{n-\mu \times n}$. Un esquema de codificación por cogrupo basado en H que quiera transmitir el máximo de $k = n - \mu$ paquetes, es de seguridad universal si y solo si el código definido por H es MRD con $m \geq n$*

Demostración. $\Rightarrow$) Primero supongamos que el esquema es universal, y veamos que ello implica que el código $\mathcal{C}$ es MRD. Gracias a la proposición 54, como un esquema universal cumple en particular $I(S; W) = 0$, se tiene

$$rg = \begin{bmatrix} H \\ B \end{bmatrix} = rg(H) + rg(B) = n$$

y con esto aplicando directamente el teorema 58 se concluye que $\mathcal{C}$ es MRD.
$\Leftarrow$) Vemos ahora que si el código es MRD, el esquema cumple las condiciones de seguridad universal (4.18) y (4.19). Si $\mathcal{C}$ es MRD, el teorema 58 proporciona la condición (4.19)

$$rg = \begin{bmatrix} H \\ B \end{bmatrix} = rg(H) + rg(B) = n \quad \forall B : \ rg(B) = \mu$$

Falta ver que $H(S|X) = 0$, pero esto es directo al considerar un esquema de codificación por cogrupo, que dada una palabra la descodifica de manera única. □

En lo que sigue, procedemos a probar que el esquema propuesto es óptimo con respecto a la longitud de paquetes, es decir, no existe un esquema de seguridad universal con $m < n$. O lo que es lo mismo, el esquema minimiza la longitud de paquete requerida de entre todos los esquemas universales. Para ello, necesitaremos los siguientes dos resultados.

Lema 60. *Sean $S \in \mathcal{S}$, $W \in \mathcal{W}$ y $X \in \mathcal{X}$ variables aleatorias discretas con $S = f(X)$ y $W = g(X)$. Se supone S uniforme y $|\{x \in \mathcal{X} : g(x) = w\}| = |\mathcal{S}|$, $\forall w \in \mathcal{W}$. Entonces $I(S; W) = 0$ implica $H(X|S, W) = 0$.*

La demostración del anterior lema se sirve de la no negatividad de la entropía relativa o divergencia de Kullback-Leibler, y se puede consultar en [Sil]

Lema 61. *Sea $\mathcal{C} \subseteq \mathbb{F}_{q^m}^n$ un código para la métrica de rango. Dado $B \in \mathbb{F}_q^{\mu \times n}$, sea $g_B : \mathcal{C} \to \mathbb{F}_{q^m}^{\mu}$ definido por $x \mapsto Bx$. Entonces g_B es inyectiva para todo B de máximo rango si y solo si $d_R(\mathcal{C}) \geq n - \mu + 1$.*

Demostración. Vamos a demostrar que $d_R(\mathcal{C}) \leq n - \mu \Leftrightarrow$ para algún B de máximo rango g_B no es inyectiva.
$\Rightarrow$) Si $d_R(\mathcal{C}) \leq n - \mu$, entonces existen $x, y \in \mathcal{C}$ distintos tal que $rg(y - x) \leq n - \mu$. Esto implica que existe algún $B \in \mathbb{F}_q^{\mu \times n}$ de rango máximo tal que $B(y - x) = 0$, es decir $Bx = By$. Por lo que g_B no es inyectiva para esta elección de B.
$\Leftarrow$) Supongamos que g_B no es inyectiva para algún $B \in \mathbb{F}_q^{\mu \times n}$ de rango máximo. Entonces existen $x, y \in \mathcal{C}$ distintos tal que $B(y - x) = 0$. Esto implica que $rg(y - x) \leq n - \mu$. Por tanto, $d_R(\mathcal{C}) \leq n - \mu$. □

Ahora estamos preparados para presentar el resultado que buscábamos.

Teorema 62. *Sea un código de red lineal $(n, m)_q$ sujeto a μ observaciones. Una comunicación con seguridad universal del máximo número de paquetes alcanzable $k = n - \mu$ es posible solo si $m \geq n$.*

Demostración. Suponemos que se tiene una comunicación de seguridad universal que alcanza la transmisión de paquetes máxima, por lo que $H(S|X) = 0$ y $I(S; W) = 0$. Por un lado $H(S|X) = 0$, por lo que podemos asumir $S = f(X)$.
Transmitir de manera segura el máximo de paquetes (k) implica que $H(S) = k = n - \mu$. Esto se consigue con una distribución uniforme, por lo que S es uniforme sobre $\mathbb{F}_{q^m}^{\mu}$. Además, para cualquier B de rango máximo (μ), siguiendo el razonamiento usado en otras ocasiones se tiene

$$|\{x \in \mathbb{F}_{q^m}^n : Bx = w\}| = |\mathbb{F}_{q^m}|^{n - rg(B)} = |\mathbb{F}_{q^m}|^{n-\mu} = |\mathbb{F}_{q^m}|^k$$

para todo $w \in \mathbb{F}_{q^m}^{\mu}$. Las condiciones del lema 60 se cumplen y como $I(S; W) = 0$, se tiene $H(X|S, W) = 0$.
Denotemos ahora para cada $s \in \mathbb{F}_{q^m}^k$ el conjunto $\mathcal{X}_s = \{x \in \mathbb{F}_{q^m}^n : f(x) = s\}$. La condición $H(X|S, W) = 0$ significa que X queda unívocamente determinado conocidos $W = BX$ y S, es decir que $X \in \mathcal{X}_S$. Esto en particular se cumple para todo B de rango máximo y podemos pensar en B como una aplicación inyectiva de $\mathcal{X}_s$ en $\mathbb{F}_{q^m}^{\mu}$, ya que si $Bx = By$ con $x \neq y$, necesariamente x e y están en distintos $\mathcal{X}_s$ (pues sino, conocer B y S no determina unívocamente X). Por lo tanto, si vemos cada $\mathcal{X}_s$ como un código para la métrica de rango estamos en condiciones de aplicar el lema 61 y llegamos a que $d_R(\mathcal{X}_s) \geq n - \mu + 1$.
Por otro lado, los conjuntos $\mathcal{X}_s$ forman una $|\mathbb{F}_{q^m}^k|$-partición de $\mathbb{F}_{q^m}^n$, un conjunto por cada $s \in \mathbb{F}_{q^m}^k$. Así el tamaño medio de cada uno de estos conjuntos es $|\mathbb{F}_{q^m}^n|/|\mathbb{F}_{q^m}^k| = q^{m(n-k)} = q^{m\mu}$, es decir al menos uno de ellos tiene cardinal mayor o igual a $q^{m\mu}$ (pues si todos tuviesen tamaño menor que $q^{m\mu}$, no formarían todo $\mathbb{F}_{q^m}^n$).
Repasando lo obtenido hasta ahora, se tiene al menos un código para la métrica de rango $\mathcal{X}_s$ para el que se cumple $d_R(\mathcal{X}_s) \geq n - \mu + 1$ y también $|\mathcal{X}_s| \geq q^{m\mu}$; concluyamos que en virtud de la cota de Singleton (2.6) esto solo es posible si $m \geq n$. Se tiene que verificar

$$q^{m\mu} \leq |\mathcal{X}_s| \leq q^{max\{m,n\}(min\{m,n\} - d_R + 1)}$$

Si $m \geq n$, estudiando solo los exponentes se necesita que $m\mu \leq m(n - d_R + 1)$ lo cual se cumple si $d_R = n - \mu + 1$.
Si $n > m$, se tiene que cumplir $m\mu \leq n(m - d + 1) \leq n(m - n + \mu - 1 + 1)$ o lo que es lo mismo, $n^2 + n(-\mu - m) + m\mu = (n - m)(n - \mu) \leq 0$. Pero esto es absurdo, ya que $n > m$ y $n > \mu$. Luego se concluye que necesariamente $m \geq n$. □

Como muestra el anterior teorema, si $m < n$ no existen esquemas de seguridad universal. Para $m \geq n$, no solo existen, sino que alcanzan la transmisión máxima de paquetes. La existencia de estos esquemas universales, hacen que no sea necesario aumentar el tamaño del cuerpo sobre el que se opera con network coding, lo único que se requiere es que se transmitan paquetes de longitud suficiente ($m \geq n$), lo cual no es muy complejo en la práctica.

A continuación se ilustra lo explicado en la sección con un ejemplo de comunicación multidifusión a través de la red de la figura 1.1 estudiada en el capítulo 1. Sin embargo, es importante recordar que la red por la que se realice la comunicación es irrelevante mientras que el código de red permita la transmisión de todos los paquetes, es decir que los receptores sean capaces de recuperar el mensaje enviado por el emisor.

Ejemplo 63. Consideramos una comunicación con parámetros $m = n = 3$, $\mu = 2$ y $k = n - \mu = 1$. Sea la extensión de cuerpos $\mathbb{F}_2 \hookrightarrow \mathbb{F}_{2^3}$ generada por una raíz $\alpha \in \mathbb{F}_{2^3}$ del polinomio primitivo $f(x) = x^3 + x + 1$ sobre $\mathbb{F}_2$. Los elementos 1, α y α^2 son linealmente independientes y de acuerdo al capítulo 2, un posible código $[n, \mu]$ MRD sobre $\mathbb{F}_{q^m}$ viene dado por la matriz de control

$$H = \begin{bmatrix} 1 & \alpha & \alpha^2 \end{bmatrix}$$

Así, el mensaje a enviar es $S \in \mathbb{F}_{2^3}$ y se codifica en $X = \begin{bmatrix} X_1 & X_2 & X_3 \end{bmatrix} \in \mathbb{F}_{2^3}$ cumpliéndose $S = HX$. Para elegir X, dado S podemos escoger $X_2, X_3 \in \mathbb{F}_{2^3}$ de forma aleatoria, y X_1 quedará determinado al tener que cumplirse

$$S = HX = X_1 + \alpha X_2 + \alpha^2 X_3.$$

Por lo que $X_1 = S + \alpha X_2 + \alpha^2 X_3$. Supongamos que el espía en la red pincha 2 aristas, las mostradas en la figura 4.3, obteniendo $W = BX$, con

$$B = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

Se tiene entonces

$$W = B \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} S + \alpha X_2 + \alpha^2 X_3 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} S + (1+\alpha) X_2 + \alpha^2 X_3 \\ X_2 + X_3 \end{bmatrix}$$

y separando las variables tenemos

$$W = \begin{bmatrix} 1 \\ 0 \end{bmatrix} S + \begin{bmatrix} 1+\alpha & \alpha^2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} X_2 \\ X_3 \end{bmatrix}$$

o también

$$W = \begin{bmatrix} 1 & 1+\alpha & \alpha^2 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} S \\ X_2 \\ X_3 \end{bmatrix}$$

Esto es un sistema lineal de 2 ecuaciones con 3 incógnitas sobre $\mathbb{F}_{2^3}$, por lo que dado S hay una única solución para (X_2, X_3) para cada valor de W. Es decir $P(W|S) = 1/|\mathbb{F}_{2^3}|^2 = 1/8^2$, $\forall S, W$, por lo que W y S son independientes. Por tanto $P(W|S) = P(W) = 1/8^2$, $\forall S, W$ así que gracias a la independencia comentada y la regla de la cadena para la entropía se tiene $H(W|S) = H(W)$ y se concluye $I(W;S) = H(W) - H(W|S) = 0$. Lo que significa que el espía no obtiene ninguna información sobre S.

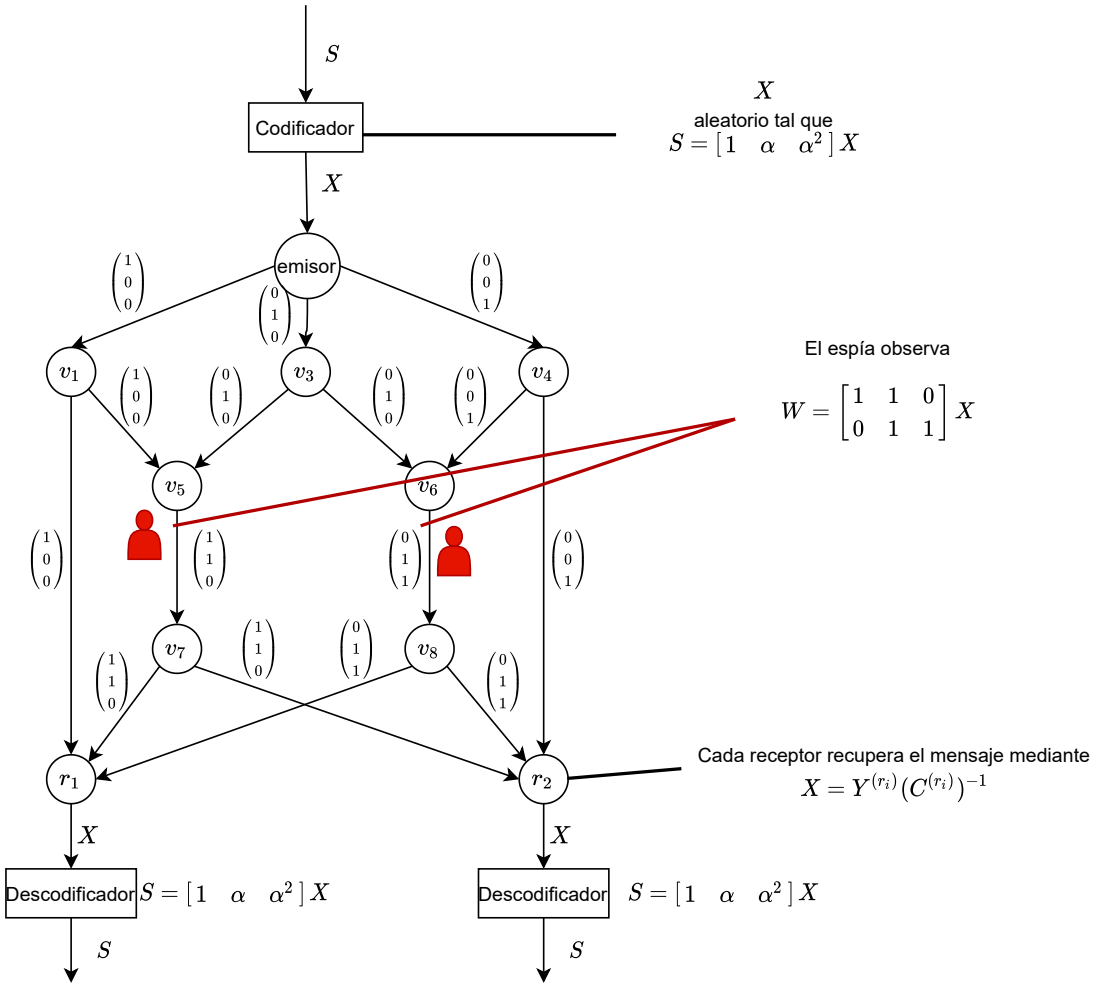


Figura 4.3: Ejemplo de comunicación espiada

Es obvio que si el intruso espía menos aristas ($\mu = 1$), sigue sin obtener información, sea cual sea la que espíe. Pongamos que pinchan cualquiera de las aristas cuyo vector de codificación global es $\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$, es decir

$$B = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$$

Siguiendo el razonamiento anterior

$$W = B \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} S + \alpha X_2 + \alpha^2 X_3 \\ X_2 \\ X_3 \end{bmatrix} = S + \alpha X_2 + \alpha^2 X_3$$

Por lo que dado S, tenemos una ecuación en $\mathbb{F}_{2^3}$, en la que hay 8 pares de valores (X_2, X_3) que son solución para cada valor de W. Así $P(W|S) = 1/8\ \forall S, W$ y S y W son independientes, deduciéndose como antes $I(W;S) = 0$.

Sin embargo, aunque el intruso espíe más del límite de aristas establecido como seguro, este es $\mu = 2$, puede que siga sin obtener información. Supongamos que se pinchan $\mu = 3$ aristas, cualesquiera que tengan como vectores de codificación global las filas de

$$B = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

Entonces se verifica

$$W = B \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} S + \alpha X_2 + \alpha^2 X_3 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} X_3 \\ X_2 \\ X_2 + X_3 \end{bmatrix}$$

O también

$$W = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} X_2 \\ X_3 \end{bmatrix}$$

Obteniendo un sistema de 3 ecuaciones con 2 variables, en el que una ecuación es redundante. De hecho, estamos en una situación similar al primer caso del ejemplo: hay exactamente una solución (X_2, X_3) para cada valor W, y recordemos que tanto X_2 como X_3 son elegidos aleatoriamente en $\mathbb{F}_{2^3}$, por lo que $P(W) = P(W|S) = 1/8^2 \ \forall S, W$ y se concluye de nuevo con $I(W; S) = 0$.

En realidad esto es lo que dicta la teoría vista en este capítulo: se garantiza una comunicación segura mientras no se espíen más de $\mu = 2$ aristas cuyos vectores de codificación global sean linealmente independientes, es decir se ha de cumplir $rg(B) \leq 2$. En el caso recién expuesto se cumple $rg(B) = 2$, aunque se hayan espiado $\mu > 2$ aristas.

Ahora veamos qué puede ocurrir si se rompe esta condición de seguridad, es decir si el intruso observa un número de aristas mayor que 2 cuyos vectores de codificación global son linealmente independientes. Así, imaginemos la situación en que espían $\mu = 3$ aristas tales que

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

En ese caso se tiene

$$W = B \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} S + \alpha X_2 + \alpha^2 X_3 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} S + \alpha X_2 + \alpha^2 X_3 \\ X_2 \\ X_3 \end{bmatrix}$$

Obteniendo el intruso el mensaje codificado completo $W = X$. Más detalladamente, reescribiéndolo como en los casos anteriores se observa

$$W = \begin{bmatrix} W_1 \\ W_2 \\ W_3 \end{bmatrix} = \begin{bmatrix} 1 & \alpha & \alpha^2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} S \\ X_2 \\ X_3 \end{bmatrix} \tag{4.21}$$

Por lo que el mensaje S podría revelarse como $S = W_1 + \alpha W_2 + \alpha^2 W_3$.
Si se quiere describir en términos de entropía, tenemos que (4.21) es un sistema lineal de 3 ecuaciones en 3 variables, que para cada valor de W existe una única solución para $(S, X_2.X_3)$. Como X determina unívocamente S, al tener que el espía observa $W = X$ se cumple que $H(S|W) = H(S|X) = 0$. Por otro lado se tiene que $H(S) = k = 1$. Se concluye entonces que $I(S; W) = H(S) - H(S|W) = 1$, por lo que el espía obtiene toda la información.

En resumen, lo importante como se ha visto durante el capítulo a partir del Teorema 55, es que en una comunicación de seguridad universal, un espía que pinche $\mu \leq n - k$ aristas con vectores de codificación global linealmente independientes nunca obtendrá información del mensaje.

4.2.2. Sobre una red sujeta a ruido

Aunque no sea el objeto de nuestro trabajo, conviene al menos mostrar el papel que juegan los errores en una comunicación por red. De hecho, el esquema explicado en este capítulo no solo proporciona seguridad a la red, sino que también provee de fiabilidad en este sentido, tal y como presentan [Sil] y [SiKsch].

Al igual que el network coding tiene sus ventajas, también puede conllevar inconvenientes. ¿Qué ocurre si hay paquetes corruptos en la red, o un intruso los inyecta?. El hecho de combinar información en los nodos de la red, hace que un paquete erróneo pueda contaminar el resto de paquetes que dependen de este, propagando los errores como ejemplifica la figura 4.4. Este problema de propagación puede sobrepasar la capacidad correctora de los códigos para la métrica de Hamming, por ello son necesarios los códigos correctores con la métrica de rango.

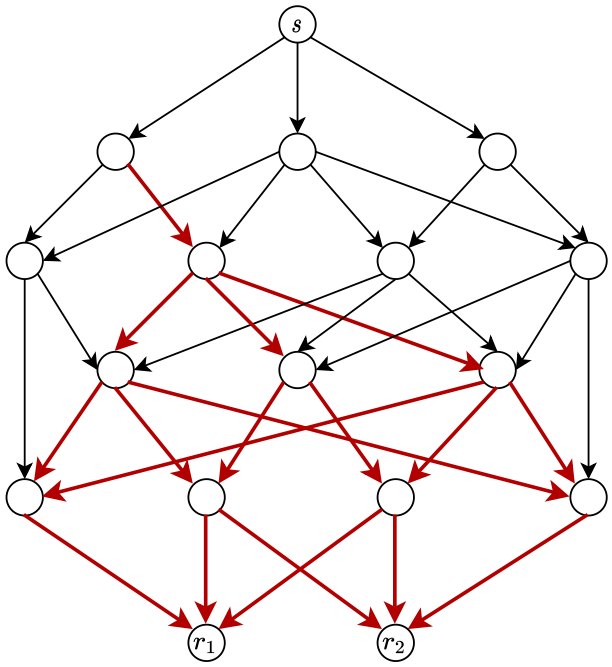


Figura 4.4: Ejemplo de red contaminada por un paquete erróneo

Diferenciamos a continuación los borrones o paquetes perdidos de los errores o paquetes erróneos. Se recuerda del capítulo 1 que para una comunicación exitosa de n paquetes, los vectores de codificación global entrantes a cada receptor $r \in R$ tienen que ser una base de $\mathbb{F}_q^n$, es decir $rg(C^{(r)}) = n$. Así, se define la deficiencia de rango de un código de red como

$$\rho = n - \min_{r \in R} rg(C^{(r)})$$

En el contexto de network coding esto es análogo a la pérdida de paquetes, por lo que una deficiencia de rango ρ también es entendido como ρ borrones.
Por otro lado, supongamos que por una arista i circula un paquete $Y'(i) \in \mathbb{F}_{q^m}^n$ que no es el esperable, es decir en el que se han introducido errores. Entonces se define el paquete error de la arista i como $k(i) = Y'(i) - Y(i)$ siendo $Y(i)$ el paquete sin errores que debiera estar transmitiéndose. Por la linealidad de la red, el paquete transmitido por una arista $j \in out(i)$ será $Y'(j) = Y(j) + f_{i,j}(Y'(i) - Y(i))$ para algún $f_{i,j} \in \mathbb{F}_q$. Se puede escribir entonces

$$Y^{(r)} = XC^{(r)} + KF$$

siendo $Y^{(r)} \in \mathbb{F}_{q^m}^n$ los paquetes recibidos por el receptor r, $X \in \mathbb{F}_{q^m}^n$ los paquetes originales enviados, $C^{(r)} \in \mathbb{F}_q^{n \times |E|}$ la matriz de codificación global de r, $K \in \mathbb{F}_{q^m}^{|E|}$ los paquetes error inyectados en la

red, y $F \in \mathbb{F}_q^{|E|\times|E|}$ la matriz cuyas entradas $f_{i,j}$ son las transformaciones aplicadas a los errores a lo largo de la red. El número de filas no nulas de E se denota por $wt(E)$, que es el número total de paquetes erróneos.

Se puede plantear una comunicación por una red con deficiencia de rango ρ y sujeta a t errores. Una codificación para la que exista una descodificación que cumpla la propiedad de error-cero descrita al inicio del capítulo se dice que tiene una capacidad correctora de t errores y ρ borrones, y esto según [SiKsch] ocurre si el código usado es MRD y se cumple $d_R(\mathcal{C}) > 2t + \rho$.
Si consideramos un problema completo de una comunicación por una red con deficiencia de rango ρ, sujeta a t errores y μ observaciones, también en [SiKsch] establecen que el esquema de comunicación tiene capacidad correctora de t errores y ρ borrones y es de seguridad universal bajo μ observaciones si el código usado es MRD con $m \geq n$ y $k \leq n - 2t - \rho - \mu$.

Es decir, un esquema universal como el descrito en este capítulo, bajo los parámetros ρ, t, μ, puede alcanzar la tasa (de hecho óptima bajo estos supuestos) de $n - 2t - \rho - \mu$ paquetes.

Bibliografía

[Ahls] Ahlswede, Rudolf, et al. "Network information flow." IEEE Transactions on information theory 46.4 (2000): 1204-1216.

[GeTh] Geil, Olav, and Casper Thomsen. "Aspects of random network coding." Algebraic Geometry Modeling in Information Theory. 2013. 47-81.

[Ros] Rossen, Kenneth. "Discrete Mathematics and its Applications." McGraw Hill (2003).

[LSRYC] Li, S-YR, Raymond W. Yeung, and Ning Cai. "Linear network coding." IEEE transactions on information theory 49.2 (2003): 371-381.

[Yeung] Yeung, Raymond W. Information theory and network coding. Springer Science & Business Media, 2008.

[JaSa] Jaggi, Sidharth, et al. "Polynomial time algorithms for multicast network code construction." IEEE Transactions on Information Theory 51.6 (2005): 1973-1982.

[Gabi] Gabidulin, Ernest Mukhamedovich. "Theory of codes with maximum rank distance." Problemy Peredachi Informatsii 21.1 (1985): 3-16.

[LoVei] López García, Candido Antonio, and Manuel Fernández Veiga. Teoría de la Información y Codificación. Enxeñaría telemática, 2002.

[CoJo] Cover, Thomas M., and Joy A. Thomas. "Entropy, relative entropy and mutual information." Elements of information theory 2 (1991): 1-55.

[OzWy] Ozarow, Lawrence H., and Aaron D. Wyner. "Wire-tap channel II." AT&T Bell Laboratories technical journal 63.10 (1984): 2135-2157.

[RoSol] El Rouayheb, Salim Y., and Emina Soljanin. "On wiretap networks II." 2007 IEEE International Symposium on Information Theory. IEEE, 2007.

[SiKsch] Silva, Danilo, and Frank R. Kschischang. "Universal secure network coding via rank-metric codes." IEEE Transactions on Information Theory 57.2 (2011): 1124-1135.

[Sil] Silva, Danilo. Error control for network coding. Ph. D. dissertation: University of Toronto, 2009.

[JuTo] Justesen, Jørn, and Tom Høholdt. A course in error-correcting codes. Vol. 1. European Mathematical Society, 2004.

[Casp] Thomsen, Casper. Random linear network coding, zeros of multivariate polynomials and affine variety codes. Ph. D. Thesis: Aalborg University, 2011.

[HoLun] Ho, Tracey, and Desmond Lun. Network coding: an introduction. Cambridge University Press, 2008.

[GoRav] Gorla, Elisa, and Alberto Ravagnani. "Codes endowed with the rank metric." Network Coding and Subspace Designs. Springer, Cham, 2018. 3-23.