

完全二部グラフを用いた RAID6 での Data と Parity の配置

東邦大学 理学部 情報科学科
足立 智子, 杉浦 智吉

Toho University, Department of Information Sciences
Tomoko Adachi, Tomoyoshi Sugiura

1. はじめに

RAID 6 とは, 複数のディスクを組み合わせた RAID システムにおいて, 二重にパリティを生成し, ディスク故障を二個まで許容するシステムである. 一方, グラフ理論においては, マッチングや 1-因子分解に関して多くの研究がなされている. 本稿では, グラフの特性を用いた RAID 6 での二個のディスク故障を許容するモデルについて論じる.

ディスクでのブロックは, グラフの一組の辺 (u, v) に対応し, タグ $\{u, v\}$ が割り当てられる. タグが v を含んでいるすべてのブロックにおいて, ビットの排他的論理和から得られたパリティは, 頂点 v に対応するディスクでのブロックに格納される. パリティブロックは, 自己ループ (v, v) によって表され, タグ $\{v, v\}$ が割り当てられる.

2. ディスク故障から復旧する仕組み

アレイのディスクが一個故障した場合に, この故障したディスクのデータブロック, パリティブロックを, 一連のステップを繰り返すことにより復旧を行う. 故障したディスクの中の復旧されていないブロック — このブロックはタグ $\{u, v\}$ を持つ — は, 残っているブロックのうち u を含むタグを持つすべてのブロックが故障していない場合, または, 前の一連のステップで復旧が終わっている場合に, 一般性を失わずに, この一連のステップにより復旧できる.

この一連のステップの中では, 辺 (u, v) がグラフ G において頂点 u と結合する唯一の辺である場合も起こりうる. この場合, 故障したディスクの中で u に対応するブロックは, 現在行われているステップの中で, 復旧されないままである. 現在行われているステップが 1 番目であるなら, 辺 (u, v) は故障したディスクの中のブロック — u に結合するブロック — に対応する唯一の辺である.

タグ $\{u, v\}$ を持つブロックを復旧した後, 同様にして, v を含むタグ $\{v, w\}$ を持つブロックを復旧することができる. つまり, 残っているブロックのうち v を含むタグを持つすべてのブロックが, 故障していない場合, または, 前の一連のステップで復旧が終わっている場合, 故障したディスクでのブロックに対応する高々二本の辺は, v と結合できる.

同様にして, この繰り返し — 残っているすべてのデータとタグが w を含んでいるパリティブロックは完全または前の繰り返しで復旧 — により, w を含むタグを持っているデータブロックだけを復旧することができる.

このようにして, 故障したディスク上の連続したデータブロックは, 辺が閉路とならないなら復旧することができる.

3. $N-1$ 個のディスクでのデータとパリティの配置

$2N$ 個の頂点を持つ完全二部グラフを用いて、 $N-1$ 個のディスクでのディスク故障を許容するアレイにおいて、どのようにデータとパリティの配置をモデル化すればよいのかを論じる。この配置のモデル化には近似1-因子分解を利用する。ここで、 N は、素数 P または $2P-1$ である。 N が素数 P である場合については第4節で、 N が $2P-1$ (P は素数)である場合については第5節で、データとパリティの配置のモデル化を示す。

本節では、ディスクアレイのモデル化に関してその考え方の基幹となる近似1-因子分解に関する定理を示す。

定理 3.1 P を素数とする、 $2P$ 個の頂点を持つ完全二部グラフの各頂点に自己ループを付け加えたグラフを $Q_{p,p}$ とする。このとき、 $Q_{p,p}$ は、 P 個の1-因子に分解でき、任意の二つの1-因子の和集合の辺がハミルトン閉路を形成するように1-因子分解できる。

[証明] $Q_{p,p} = (V, E)$, $V = V_1 \cup V_2$, $V_1 = \{0, 1, \dots, P-1\}$, $V_2 = \{P, P+1, \dots, 2P-1\}$ とする。 $Q_{p,p}$ の各辺 (u, v) に、 $u + v \pmod{P}$ とラベル付けする。次の(a), (b)を示せば定理が証明できる。

(a) 共通のラベルの辺と $Q_{p,p}$ の頂点を持つグラフは $Q_{p,p}$ の1-因子である。この命題を示すために、まず、共通のラベルを持つ P 本の独立した辺が存在することを指摘する。これは、共通のラベルを持つ2本の辺は隣接しないことからわかる。もし、隣接する2本の辺 (u, v) と (v, w) が共通のラベル e を持つと仮定すると、

$$\begin{aligned} e &\equiv u + v \pmod{P} \quad \text{かつ} \quad e \equiv v + w \pmod{P} \\ u + v &\equiv v + w \pmod{P} \\ \therefore u &\equiv w \pmod{P} \end{aligned}$$

となる。頂点 u と w は、共に頂点 v に隣接しているので、同じ部分集合に属する。したがって $u = w$ となるが、これは仮定に矛盾している。さて、もしラベル e_1 を持つ辺が P 本より少ないならば、 $e_1 \neq e_2$, $0 \leq e_1, e_2 \leq P-1$ となるラベル e_2 を持つ辺が P 本より多くなる。これは、 $Q_{p,p}$ には P^2 本の辺と P 個のラベルがあることよりわかる。このとき、ラベル e_2 を持つ2本の辺は隣接しなければならない。これは矛盾している。したがって、共通のラベルを持つ P 本の独立した辺が存在する。この事実より、共通のラベルを持つ辺集合によって誘導された辺誘導部分グラフは $Q_{p,p}$ の1-因子であることがわかる。

(b) 任意の二つの1-因子の和集合の辺がハミルトン閉路を形成する。この命題を背理法で証明する。ラベル e_1, e_2 を持つ二つの1-因子の辺を交互に進むことにより閉路 $v_1, v_2, v_3, \dots, v_t, v_1$ ($t < 2P$) が形成できると仮定する。一般性を失わずに、辺 $(v_1, v_2), (v_3, v_4), \dots, (v_{t-1}, v_t)$ はラベル e_1 を持ち、 $(v_2, v_3), (v_4, v_5), \dots, (v_t, v_1)$ はラベル e_2 を持つと仮定できる。このとき t は偶数であることに注意して、 $t=2k$, $k < P$ とおくと、

$$\begin{aligned} e_1 &\equiv (v_1 + v_2) \pmod{P}, & e_2 &\equiv (v_2 + v_3) \pmod{P}, \\ e_1 &\equiv (v_3 + v_4) \pmod{P}, & e_2 &\equiv (v_4 + v_5) \pmod{P}, \\ &\dots & &\dots \\ e_1 &\equiv (v_{t-1} + v_t) \pmod{P}, & e_2 &\equiv (v_t + v_1) \pmod{P} \end{aligned}$$

となるので、 $t(e_1 - e_2)/2 \equiv 0 \pmod{P}$ が容易にわかる。ゆえに $k(e_1 - e_2) \equiv 0 \pmod{P}$ 。 P と k は互いに素であり、 $e_1, e_2 < P$ なので、これは矛盾する。□

4. $P-1$ 個のディスクでのデータとパリティの配置

N が素数 P である場合に、 $2N(=2P)$ 個の頂点の完全二部グラフを用い、 $N-1(=P-1)$ 個のディスクにおけるデータとパリティの配置のモデル化を示す。

ディスクアレイのモデル化は、素数 P の値を固定して図示しても一般性を失わないので、本節の前半では、 $P=5$ に固定した例や図を用いて、一般の素数 P の場合のディスクアレイの構成法を説明する。本節の後半では、このディスクアレイで二個のディスクが故障した際の復旧の手順を説明する。

完全二部グラフ $K_{N,N}$ の各頂点に自己ループを加えたグラフ $Q_{N,N}$ を考える。このグラフ $Q_{N,N}$ は、 $2N$ 個の頂点と N^2+2N 本の辺（うち $2N$ 本は自己ループ）を持つ。グラフ $Q_{N,N}$ の $2N$ 本の自己ループをパリティブロックに対応させ、自己ループ以外の N^2 本の辺をデータブロックに対応させる。グラフ $Q_{N,N}$ に対応させて、 N^2 個のデータブロック $\{u, v\}$ ($N \leq u \leq 2N-1$, $0 \leq v \leq N-1$) を N 個のディスク Disk0, Disk1, Disk2, ..., Disk $N-1$ に配置する。このデータブロックの配置には、 $2N$ 個のパリティブロック $\{(0, 0), \{1, 1\}, \{2, 2\}, \dots, \{2N-1, 2N-1\}\}$ は含まれていない。 $N-1$ 個のディスクに配置するため、ディスクの個数を1個減らす必要があるが、その方法については後述する。また、パリティブロックを配置する領域を作るためのデータブロックの取り除き方についても後述する。

素数 $P=5$ の場合、完全二部グラフ $K_{5,5}$ の各頂点に自己ループを加えたグラフ $Q_{5,5}$ は、10 個の頂点と 35 本の辺（うち 10 本は自己ループ）を持つ。このグラフ $Q_{5,5}$ の自己ループを除く 25 本の辺をデータブロック $\{u, v\}$ ($5 \leq u \leq 9$, $0 \leq v \leq 4$) に対応させて、5 個のディスク Disk0, Disk1, Disk2, Disk3, Disk4 に配置する。このデータブロックの配置には損失したデータを復旧する際に用いる 10 個のパリティブロックは含まれていない。

ここで、データブロックが配置されるディスクの定め方を説明する。グラフ $Q_{N,N}$ におけるデータブロックの配置のディスク番号を M ($M = 0, 1, 2, 3, \dots, N-1$) とすると、データブロック $\{u, v\}$ が配置されるディスクは、

$$u + v \equiv M \pmod{N}$$

を満たすディスクである。例示している図は $N=P=5$ の場合であり、データブロック $\{9, 4\}$ は、

$$9 + 4 \equiv 3 \pmod{5}$$

であるため、Disk3 に配置される。他のデータブロックについても同様に配置されるディスク番号が定まる。

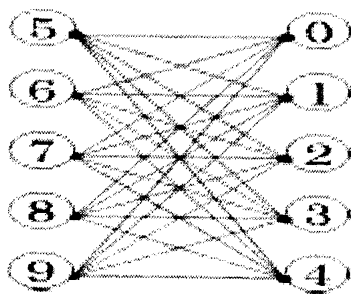


図 4.1 グラフ $Q_{5,5}$

Disk0	Disk1	Disk2	Disk3	Disk4
$\{5, 0\}$	$\{5, 1\}$	$\{5, 2\}$	$\{5, 3\}$	$\{5, 4\}$
$\{6, 4\}$	$\{6, 0\}$	$\{6, 1\}$	$\{6, 2\}$	$\{6, 3\}$
$\{7, 3\}$	$\{7, 4\}$	$\{7, 0\}$	$\{7, 1\}$	$\{7, 2\}$
$\{8, 2\}$	$\{8, 3\}$	$\{8, 4\}$	$\{8, 0\}$	$\{8, 1\}$
$\{9, 1\}$	$\{9, 2\}$	$\{9, 3\}$	$\{9, 4\}$	$\{9, 0\}$

図 4.2 グラフ $Q_{5,5}$ に対応するデータブロックの配置

今、グラフ $Q_{N,N}$ に対応するディスクアレイでは、 N^2 個のデータブロックが N 個のディスクに配置されている。配置したいディスクは $N-1$ 個であるため、一つのディスク Disk0 を、そのディスクに配置されている N 個のデータブロックごとに取り除く。グラフ $Q_{N,N}$ から Disk0 を取り除

くことは、グラフの頂点0と頂点Nを、これらの頂点に接続している $2N-1$ 本の辺ごとに取り除くことである。Disk0 に配置されたデータブロック $\{N, 0\}$ は、頂点0, N に接続しているが、すでに取り除かれているので、実際は $2N-2$ 本の辺を取り除く。さらに、 $N-1$ 個のデータブロック $\{v+N, v\}$ ($1 \leq v \leq N-1$) を取り除く。グラフ $Q_{N,N}$ に対応する N^2 個のデータブロックの配置から、 $N + (2N-2) + (N-1) = 4N-3$ 個のデータブロックを取り除くことにより、パリティブロックを配置するブロック(領域)が確保される。このようにして、 $N-1$ 個のディスクの $(N-1)^2 = N^2 + 2N + 1$ 個の領域に、 $N^2 - (4N-3) = N^2 - 4N + 3$ 個のデータブロックと、 $2N-2$ 個のパリティブロック $\{\{1, 1\}, \{2, 2\}, \{3, 3\}, \dots, \{N-1, N-1\}, \{N+1, N+1\}, \{N+2, N+2\}, \{N+3, N+3\}, \dots, \{2N-1, 2N-1\}\}$ が配置できる。

例示している図は $N=P=5$ の場合であるので、この図から Disk0 に対応する5個のデータブロック $\{\{5, 0\}, \{6, 4\}, \{7, 3\}, \{8, 2\}, \{9, 1\}\}$ を取り除くことになる。グラフ $Q_{5,5}$ から Disk0 のデータブロックに対応する辺(青破線で表示)を取り除いたグラフと、対応するデータブロックの配置を下図に示す。

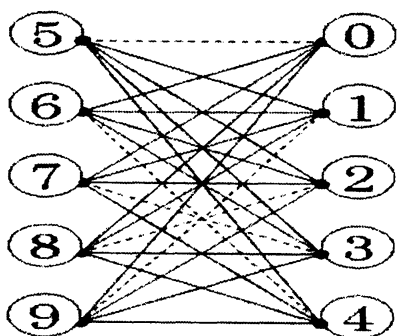


図 4.3 Disk0 に対応する辺(青破線)を除いたグラフ

Disk0	$\{5, 1\}$	$\{5, 2\}$	$\{5, 3\}$	$\{5, 4\}$
	$\{6, 0\}$	$\{6, 1\}$	$\{6, 2\}$	$\{6, 3\}$
	$\{7, 4\}$	$\{7, 0\}$	$\{7, 1\}$	$\{7, 2\}$
	$\{8, 3\}$	$\{8, 4\}$	$\{8, 0\}$	$\{8, 1\}$
	$\{9, 2\}$	$\{9, 3\}$	$\{9, 4\}$	$\{9, 0\}$

図 4.4 Disk0 を除いたデータの配置

グラフ $Q_{5,5}$ から、Disk0 に配置されたデータブロックに対応する5本の辺は取り除かれている。頂点0と頂点5に接続している辺(赤破線)を取り除くと、図4.5の実線のグラフになる。頂点0と頂点5に接続している辺は9本あるが、辺 $\{5, 0\}$ に対応したデータブロックはDisk0 に配置され、すでに取り除かれているので、赤破線の辺は8本になる。図4.4にはDisk0に対応する頂点5が存在しているので、データブロック $\{\{5, 1\}, \{5, 2\}, \{5, 3\}, \{5, 4\}\}$ を取り除くと、図4.6になる。さらに、図4.6から、同様にして、頂点0に接続する辺に対応するデータブロック $\{\{6, 0\}, \{7, 0\}, \{8, 0\}, \{9, 0\}\}$ を取り除くと、図4.7になる。

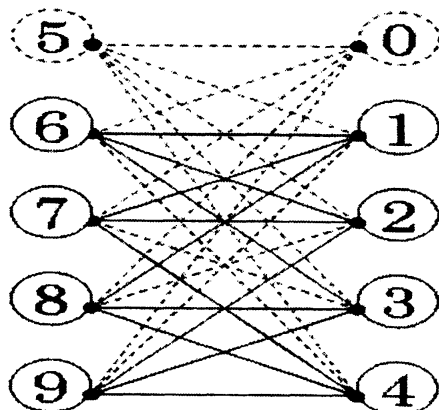


図 4.5 Disk0 の辺(青破線)と頂点0, 5に接続する辺(赤破線)を除いたグラフ

Disk0	Disk1	Disk2	Disk3	Disk4
	{6, 0}	{6, 1}	{6, 2}	{6, 3}
	{7, 4}	{7, 0}	{7, 1}	{7, 2}
	{8, 3}	{8, 4}	{8, 0}	{8, 1}
	{9, 2}	{9, 3}	{9, 4}	{9, 0}

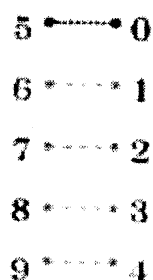
図 4.6 頂点5に接続する辺に対応するデータを除いた配置

Disk0	Disk1	Disk2	Disk3	Disk4
		{6, 1}	{6, 2}	{6, 3}
	{7, 4}		{7, 1}	{7, 2}
	{8, 3}	{8, 4}		{8, 1}
	{9, 2}	{9, 3}	{9, 4}	

図 4.7 頂点0に接続する辺に対応するデータを除いた配置

ここで、上図の配置において、データを取り除かれたブロックは、バツ印と無印がある。この違いについて説明する。本節のディスクアレイでは、最初に、 $N^2 (=5^2)$ 個のブロックにデータを配置しているが、求めるディスクアレイは、 $(N-1)^2 (=4^2)$ 個のブロックにデータとパリティを配置することである。そこで、データだけでなくブロックごとに取り除く場合にはバツ印を記し、データを取り除いたブロックに後でパリティを配置する場合には無印で表す。

さらに、図 4.7 から、4 個のデータブロック $\{v+5, v\}$ ($1 \leq v \leq 4$)、すなわち、 $\{6, 4\}$, $\{7, 3\}$, $\{8, 2\}$, $\{9, 1\}$ を取り除くと、図 4.9 になる。このデータブロックの配置に対応するグラフは、図 4.8 であり、取り除かれたデータブロックに対応する辺は緑破線で示している。



Disk0	Disk1	Disk2	Disk3	Disk4
			{6, 2}	{6, 3}
	{7, 4}		{7, 1}	
		{8, 4}		{8, 1}
	{9, 2}	{9, 3}		

図 4.8 辺 $\{v+5, v\}$ (緑破線) を除いたグラフ 図 4.9 データ $\{v+5, v\}$ を除いた配置

グラフ $Q_{5,5}$ に対応する $5^2=25$ 個のデータブロックの配置から、 $5+8+4=17$ 個のデータブロックを取り除くことにより、パリティブロックを配置するブロック(領域)が確保された。4 個のディスクの $4^2=16$ 個の領域に、8 個のデータブロック $\{6, 2\}$, $\{6, 3\}$, $\{7, 1\}$, $\{7, 4\}$, $\{8, 1\}$, $\{8, 4\}$, $\{9, 2\}$, $\{9, 3\}$ と、8 個のパリティブロック $\{1, 1\}$, $\{2, 2\}$, $\{3, 3\}$, $\{4, 4\}$, $\{6, 6\}$, $\{7, 7\}$, $\{8, 8\}$, $\{9, 9\}$ を配置すると、図 4.11 のディスクアレイになる。パリティは青地で表示している。対応するグラフは図 4.10 である。

ここで、パリティブロックが配置されるディスクの定め方を説明する。グラフ $Q_{N,N}$ におけるデータブロックの配置のディスク番号を M ($M = 0, 1, 2, \dots, N-1$) とすると、パリティブロック $\{w, w\}$ が配置されるディスクは、

$$w + w \equiv M \pmod{N}$$

を満たすディスクである。例示しているのは $N=P=5$ の場合であり、パリティブロック $\{9, 9\}$ は、

$$9 + 9 \equiv 3 \pmod{5}$$

であるため、Disk3 に配置される。他のパリティブロックについても同様に配置されるディスク番号が定まる。

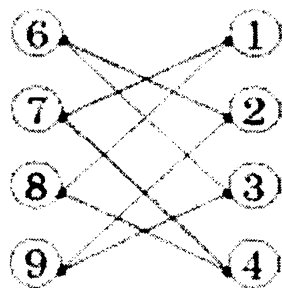


図 4.10 $Q_{9,5}$ から派生したグラフ

{3, 3}	{6, 6}	{6, 2}	{6, 3}
{7, 4}	{1, 1}	{7, 1}	{7, 7}
{8, 8}	{8, 4}	{4, 4}	{8, 1}
{9, 2}	{9, 3}	{9, 9}	{2, 2}

図 4.11 $Q_{9,5}$ からのディスクアレイ

このようにして、 N が素数 P である場合について、完全二部グラフ $K_{N,N}$ の各頂点に自己ループを加えたグラフ $Q_{N,N}$ からある条件で辺（データブロック）を取り除いていくことにより、 $N-1(=P-1)$ 個のディスク、すなわち $(N-1)^2(=(P-1)^2)$ 個のブロックに、データとパリティを配置することができる。

次に、このディスクアレイにおいて、ディスクが二個故障した際での復旧の仕組みを説明する。Disk 1, 2 が故障したとすると、データとパリティの配置は図 4.13 になる。損失したデータのグラフは図 4.12 となる。

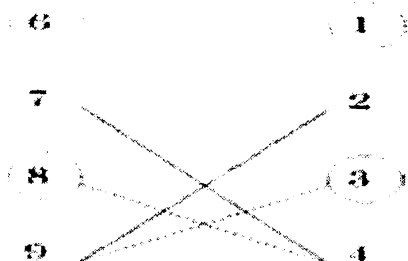


図 4.12 Disk1(実線), Disk2(破線)のグラフ

Disk1	Disk2		
{3, 3}	{6, 6}	{6, 2}	{6, 3}
{7, 4}	{1, 1}	{7, 1}	{7, 7}
{8, 8}	{8, 4}	{4, 4}	{8, 1}
{9, 2}	{9, 3}	{9, 9}	{2, 2}

図 4.13 Disk1, 2 故障時のアレイ

損失したデータ $\{ \{3, 3\}, \{7, 4\}, \{8, 8\}, \{9, 2\}, \{6, 6\}, \{1, 1\}, \{8, 4\}, \{9, 3\} \}$ は、各頂点の一本の辺しか失われていなければ、残りデータから復旧可能である。このことより頂点 7 に注目すると $\{7, 4\}$ しか損失されていないため復旧される。 $\{7, 4\}$ が復旧されたことにより頂点 4 から損失された辺が $\{4, 8\}$ の一本になるため $\{4, 8\}$ が復旧可能になり、 $\{4, 8\}$ が復旧される。 $\{4, 8\}$ が復旧されたことにより頂点 8 の損失された辺が $\{8, 8\}$ だけになり、 $\{8, 8\}$ が復旧される。同様に頂点 2 に注目すると $\{2, 9\}$ が復旧され、 $\{2, 9\}$ が復旧されたことにより $\{9, 3\}$ が復旧可能になり、 $\{9, 3\}$ が復旧される。そして $\{9, 3\}$ が復旧されたことにより $\{3, 3\}$ が復旧される。 $\{6, 6\}, \{1, 1\}$ は Disk3, Disk4 の残っているデータブロックから復旧可能である。このステップを繰り返すことにより、すべての辺が復旧される。よって、このディスクアレイは、ディスク二個の故障を許容する。

5. $2P-2$ 個のディスクでのデータとパリティの配置

本節では, N が $2P-1$ (P は素数) である場合について取り扱う. $2N(=2P)$ 個の頂点を持つ完全二部グラフを用い, $N-1(=2P-2)$ 個のディスクによるデータとパリティの配置のモデル化を説明する.

完全二部グラフ $K_{N,N}$ の各頂点に自己ループを加えたグラフ $Q_{N,N}$ を考える. 素数 $P=5$ の場合, $N=2P-1=9$ となる. 完全二部グラフ $K_{9,9}$ の各頂点に自己ループを加えたグラフ $Q_{9,9}$ を考え, 前節と同様にデータとパリティの配置を行う.

Disk0 に対応する 9 個のデータブロック $\{9, 0\}, \{10, 8\}, \{11, 7\}, \{12, 6\}, \{13, 5\}, \{14, 4\}, \{15, 3\}, \{16, 2\}, \{17, 1\}$ (青破線に対応) を取り除くと, 下図が得られる.

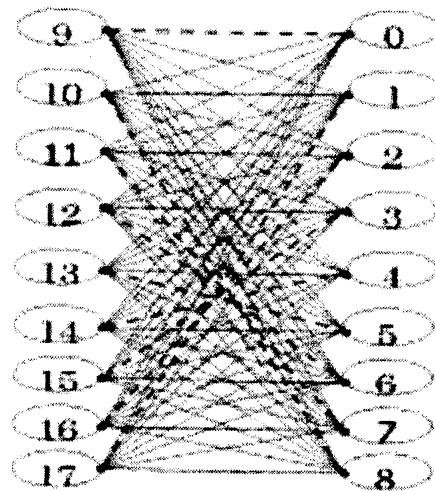


図 5.1 Disk 0 に対応する辺 (青破線) を除いたグラフ

Disk0	Disk1	Disk2	Disk3	Disk4	Disk5	Disk6	Disk7	Disk8
$\{9, 0\}$	$\{9, 1\}$	$\{9, 2\}$	$\{9, 3\}$	$\{9, 4\}$	$\{9, 5\}$	$\{9, 6\}$	$\{9, 7\}$	$\{9, 8\}$
$\{10, 8\}$	$\{10, 0\}$	$\{10, 1\}$	$\{10, 2\}$	$\{10, 3\}$	$\{10, 4\}$	$\{10, 5\}$	$\{10, 6\}$	$\{10, 7\}$
$\{11, 7\}$	$\{11, 8\}$	$\{11, 0\}$	$\{11, 1\}$	$\{11, 2\}$	$\{11, 3\}$	$\{11, 4\}$	$\{11, 5\}$	$\{11, 6\}$
$\{12, 6\}$	$\{12, 7\}$	$\{12, 8\}$	$\{12, 0\}$	$\{12, 1\}$	$\{12, 2\}$	$\{12, 3\}$	$\{12, 4\}$	$\{12, 5\}$
$\{13, 5\}$	$\{13, 6\}$	$\{13, 7\}$	$\{13, 8\}$	$\{13, 0\}$	$\{13, 1\}$	$\{13, 2\}$	$\{13, 3\}$	$\{13, 4\}$
$\{14, 4\}$	$\{14, 5\}$	$\{14, 6\}$	$\{14, 7\}$	$\{14, 8\}$	$\{14, 0\}$	$\{14, 1\}$	$\{14, 2\}$	$\{14, 3\}$
$\{15, 3\}$	$\{15, 4\}$	$\{15, 5\}$	$\{15, 6\}$	$\{15, 7\}$	$\{15, 8\}$	$\{15, 0\}$	$\{15, 1\}$	$\{15, 2\}$
$\{16, 2\}$	$\{16, 3\}$	$\{16, 4\}$	$\{16, 5\}$	$\{16, 6\}$	$\{16, 7\}$	$\{16, 8\}$	$\{16, 0\}$	$\{16, 1\}$
$\{17, 1\}$	$\{17, 2\}$	$\{17, 3\}$	$\{17, 4\}$	$\{17, 5\}$	$\{17, 6\}$	$\{17, 7\}$	$\{17, 8\}$	$\{17, 0\}$

図 5.2 Disk0 を除いたデータの配置

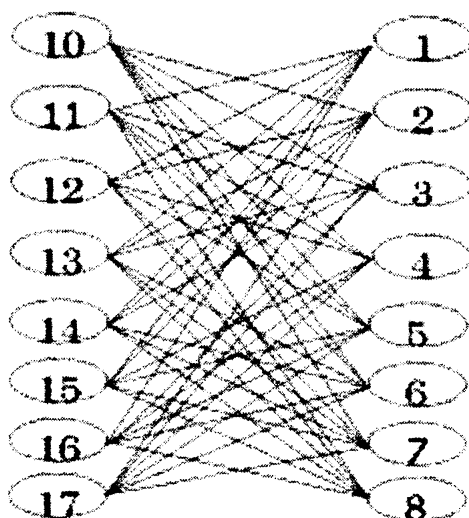
グラフ $Q_{9,9}$ から, Disk0 に配置されたデータブロックに対応する 9 本の辺は取り除かれている. 頂点 0 と頂点 9 に接続している 16 本の辺 $\{9, 1\}, \{9, 2\}, \{9, 3\}, \{9, 4\}, \{9, 5\}, \{9, 6\}, \{9, 7\}, \{9, 8\}, \{10, 0\}, \{11, 0\}, \{12, 0\}, \{13, 0\}, \{14, 0\}, \{15, 0\}, \{16, 0\}, \{17, 0\}$ を取り除き, さらに, 辺 $\{v+9, v\}$ ($1 \leq v \leq 8$), すなわち $\{10, 1\}, \{11, 2\}, \{12, 3\}, \{13, 4\}, \{14, 5\}, \{15, 6\}, \{16, 7\}, \{17, 8\}$ を取り除くと下図になる.

Disk1	Disk2	Disk3	Disk4	Disk5	Disk6	Disk7	Disk8
		{10, 2}	{10, 3}	{10, 4}	{10, 5}	{10, 6}	{10, 7}
{11, 8}		{11, 1}		{11, 3}	{11, 4}	{11, 5}	{11, 6}
{12, 7}	{12, 8}		{12, 1}	{12, 2}		{12, 4}	{12, 5}
{13, 6}	{13, 7}	{13, 8}		{13, 1}	{13, 2}	{13, 3}	
	{14, 6}	{14, 7}	{14, 8}		{14, 1}	{14, 2}	{14, 3}
{15, 4}	{15, 5}		{15, 7}	{15, 8}		{15, 1}	{15, 2}
{16, 3}	{16, 4}	{16, 5}	{16, 6}		{16, 8}		{16, 1}
{17, 2}	{17, 3}	{17, 4}	{17, 5}	{17, 6}	{17, 7}		

図 5.3 頂点 0, 9, $\{v+9, v\}$ を取り除いた配置

前節同様、対応するパリティを配置させると下図になる。

Disk1	Disk2	Disk3	Disk4	Disk5	Disk6	Disk7	Disk8
{5, 5}	{10, 10}	{10, 2}	{10, 3}	{10, 4}	{10, 5}	{10, 6}	{10, 7}
{11, 8}	{1, 1}	{11, 1}	{11, 11}	{11, 3}	{11, 4}	{11, 5}	{11, 6}
{12, 7}	{12, 8}	{6, 6}	{12, 1}	{12, 2}	{12, 12}	{12, 4}	{12, 5}
{13, 6}	{13, 7}	{13, 8}	{2, 2}	{13, 1}	{13, 2}	{13, 3}	{13, 13}
{14, 14}	{14, 6}	{14, 7}	{14, 8}	{7, 7}	{14, 1}	{14, 2}	{14, 3}
{15, 4}	{15, 5}	{15, 15}	{15, 7}	{15, 8}	{3, 3}	{15, 1}	{15, 2}
{16, 3}	{16, 4}	{16, 5}	{16, 6}	{16, 16}	{16, 8}	{8, 8}	{16, 1}
{17, 2}	{17, 3}	{17, 4}	{17, 5}	{17, 6}	{17, 7}	{17, 17}	{4, 4}

図 5.4 $Q_{8,9}$ からのディスクアレイ図 5.5 グラフ $Q_{8,9}$ から派生したグラフ

6. N個のディスクでのデータとパリティの配置

第3節から第5節では、N-1個のディスクによるデータとパリティの配置のモデル化を説明した。N-1個のディスクでの配置では、Nは素数Pまたは $2P-1$ (Pは素数)であった。本節では、Nが素数Pである場合に、 $2N$ 個の頂点の完全二部グラフを用い、 $N(N=P)$ 個のディスクによるデータとパリティの配置のモデル化を説明する。

完全二部グラフ $K_{N,N}$ の各頂点に自己ループを加えたグラフ $Q_{N,N}$ を考える。このグラフ $Q_{N,N}$ は、 $2N$ 個の頂点と N^2+2N 本の辺(うち $2N$ 本は自己ループ)を持つ。グラフ $Q_{N,N}$ の $2N$ 本の自己ループをパリティブロックに対応させ、自己ループ以外の N^2 をデータブロックに対応させる。グラフ $Q_{N,N}$ に対応させ、自己ループ以外の N^2 個のデータブロック $\{u, v\}$ ($N \leq u \leq 2N-1$, $0 \leq v \leq N-1$) を N 個のディスク Disk0, Disk1, Disk2, ..., Disk N-1 に配置する。

素数 $P=5$ の場合、完全二部グラフ $K_{5,5}$ の各頂点に自己ループを加えたグラフ $Q_{5,5}$ は図 6.1 になる。N-1個のディスク時には、頂点0と頂点P、それに対応する辺を取り除いてきたが、今回はN個のディスクであるためPの値をそのまま用いることができる。

グラフ $Q_{5,5}$ のデータブロックの配置は図 6.2 になる。グラフ $Q_{5,5}$ データブロックの配置にはグラフ $Q_{5,5}$ の自己ループ(パリティデータ) $\{\{0, 0\}, \{1, 1\}, \{2, 2\}, \{3, 3\}, \{4, 4\}, \{5, 5\}, \{6, 6\}, \{7, 7\}, \{8, 8\}, \{9, 9\}\}$ は含まれていない。

ここで、データブロックの配置については、前節と同様であり、

$$u + v \equiv M \pmod{N}$$

を満たすディスクにデータを配置する。

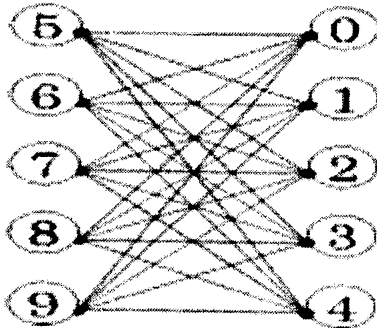


図 6.1 グラフ $Q_{5,5}$

Disk0	Disk1	Disk2	Disk3	Disk4
{5, 0}	{5, 1}	{5, 2}	{5, 3}	{5, 4}
{6, 4}	{6, 0}	{6, 1}	{6, 2}	{6, 3}
{7, 3}	{7, 4}	{7, 0}	{7, 1}	{7, 2}
{8, 2}	{8, 3}	{8, 4}	{8, 0}	{8, 1}
{9, 1}	{9, 2}	{9, 3}	{9, 4}	{9, 0}

図 6.2 グラフ $Q_{5,5}$ に対応するデータブロックの配置

第3節から第5節でのN-1個のディスク時には、頂点0と頂点Pに対応する辺を取り除くことによりパリティの配置する領域を確保したが、本節ではN個のディスク時によるものなので、N-1個のディスク時の構成法をそのまま用いるわけにはいかない。しかし、次のグラフ理論の知見—グラフ $Q_{n,p}$ の任意の二つの1-因子の組み合わせはハミルトン閉路になる—は、本節でも同様に用いることとする。この知見を用いることにより、パリティブロックを配置できる領域が、この1-因子の組み合わせによるハミルトン閉路を取り除いたブロックによって確保される。

グラフ $Q_{5,5}$ に対応するデータブロックの配置(図 6.2)からパリティに対応する二つの1-因子の組み合わせ(図 6.3)を取り除く。

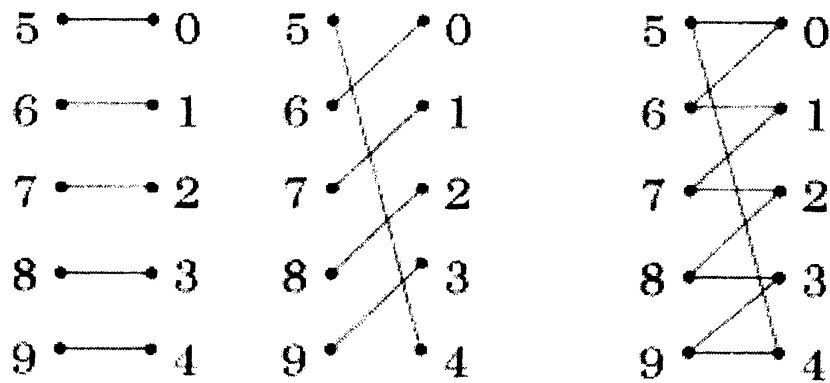


図 6.3 $Q_{5,5}$ の二つの 1-因子 (左図・中央図) と
その組み合わせによるハミルトン閉路 (右図)

1-因子の組み合わせによるハミルトン閉路に対応しているデータブロック $\{\{5, 0\}, \{6, 1\}, \{7, 2\}, \{8, 3\}, \{9, 4\}, \{5, 4\}, \{6, 0\}, \{7, 1\}, \{8, 2\}, \{9, 3\}\}$ (図 6.3 右図) を, グラフ $Q_{5,5}$ に対応するデータブロックの配置 (図 6.2) から取り除くと, ハミルトン閉路を除いたアレイ (図 6.4) になる.

Disk 0	Disk 1	Disk 2	Disk 3	Disk 4
	{5, 1}	{5, 2}	{5, 3}	
{6, 4}			{6, 2}	{6, 3}
{7, 3}	{7, 4}	{7, 0}		
		{8, 4}	{8, 0}	{8, 1}
{9, 1}	{9, 2}			{9, 0}

図 6.4 ハミルトン閉路を除いたアレイ

一般に, グラフ $Q_{N,N}$ に対応する N^2 個のデータブロックの配置から, $2N$ 個のデータブロックを取り除くことにより, パリティブロックを配置するブロック (領域) が確保される. N 個のディスクの N^2 個の領域に, $N^2 - 2N$ 個のデータブロックと, $2N$ 個のパリティブロックが配置される.

ここでは, $N=P=5$ の場合を考えているので, グラフ $Q_{5,5}$ に対応する $5^2=25$ 個のデータブロックの配置から, 10 個のデータブロックを取り除くことにより, パリティブロックを配置するブロック (領域) が確保された. 5 個のディスクの $5^2=25$ 個の領域に, 15 個のデータブロック (図 6.1.4) と, 10 個のパリティブロック $\{\{0, 0\}, \{1, 1\}, \{2, 2\}, \{3, 3\}, \{4, 4\}, \{5, 5\}, \{6, 6\}, \{7, 7\}, \{8, 8\}, \{9, 9\}\}$ を配置すると, 図 6.5 のディスクアレイになる. パリティブロックは青地で表示している.

ここで, パリティブロックが配置されるディスクの定め方を説明する. グラフ $Q_{N,N}$ におけるデータブロックの配置のディスク番号を M ($M = 0, 1, 2, \dots, N-1$) とすると, パリティブロック $\{w, w\}$ が配置されるディスクは,

$$w + w \equiv M \pmod{N}$$

を満たすディスクである. 例示している図は $N=P=5$ の場合であり, パリティブロック $\{9, 9\}$ は,

$$9 + 9 \equiv 3 \pmod{5}$$

であるため, Disk3 に配置される. 他のパリティブロックについても同様に配置されるディスク番号が定まる.

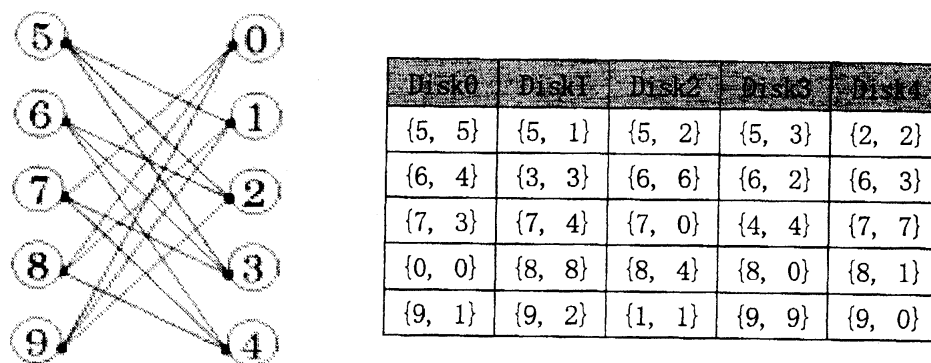


図 6.5 $Q_{5,5}$ から派生したグラフとそのディスクアレイ

次に, このディスクアレイ (図 6.5) において, ディスクが二個故障した際の復旧の仕組みを説明する. Disk 2, 4 が故障したとすると, データとパリティの配置は図 6.6 になる. 損失したデータに対応するグラフの辺を図 6.7 に示す.

Disk0	Disk1	Disk2	Disk3	Disk4
{5, 5}	{5, 1}	{5, 2}	{5, 3}	{2, 2}
{6, 4}	{3, 3}	{6, 6}	{6, 2}	{6, 3}
{7, 3}	{7, 4}	{7, 0}	{4, 4}	{7, 7}
{0, 0}	{8, 8}	{8, 4}	{8, 0}	{8, 1}
{9, 1}	{9, 2}	{1, 1}	{9, 9}	{9, 0}

図 6.6 Disk2, 4 の故障

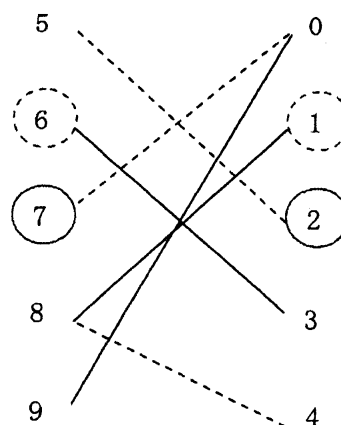


図 6.7 Disk2 (破線), Disk4 (実線) のグラフ

図 6.7 より, 損失したデータは頂点から一本辺が失われていなければ復旧可能である.

{5, 2}, {2, 2}.
 {3, 3}, {6, 6}.
 {9, 0}, {0, 7}, {7, 7}.
 {4, 4}, {8, 1}, {1, 1}.

のように順に復旧可能であるため, すべての辺が復旧される. 上で記したデータブロック, パリティブロックは, 一連のステップにより復旧できるブロックを各行に表示している. この一連のステップにより, N が素数 P である場合に, N 個のディスクでのデータとパリティの配置は, 二個のディスク故障を許容する.

7. 結論

ディスクでのブロックは、グラフでのマッチングの辺に対応し、任意の二個のディスクのブロックの組み合わせはグラフの閉路に対応する。この条件を満たすグラフは、二個のディスクの故障を許容する。このようなグラフの構成法、すなわち、 N が素数 P または $2P-1$ の場合に、 $N-1$ 個のディスクでのアレイの構成法を与え、 N が素数 P の場合に、 N 個のディスクでのアレイの構成法を与えた。そして、各グラフが二個のディスクの故障に許容性があるアレイをモデル化する際に、前述の状態を満たすことを示すことによって、それらの正当性を立証した。

参考文献

- [1] B. A. Anderson : "Symmetry Groups of Some Perfect 1-Factorizations of Complete Graphs". *Discrete Mathematics* 18(1977), 227-234.
- [2] N. Deo and S. Nanda : "One-Factors and Hamiltonian Paths in Modeling Data and Parity Placement in Disk Arrays". *CONGRESSUS NUMERANTUM* 176(2005), 191-199.
- [3] 恵羅 博・土屋 守正 : 『グラフ理論』. 産業図書, 1997.
- [4] 伊勢 雅英 : 『伊勢 雅英のストレージ最前線』. 2005.
URL: <http://enterprise.watch.impress.co.jp/cda/storage/2005/05/09/5191.html>
- [5] 小林 章彦 : 『普及が始まった「RAID 6」とは』. 2006.
URL: <http://www.atmarkit.co.jp/fsys/keyword/019raid6/019raid6.html>
- [6] M. Kobayashi : "On Perfect One-Factorization of the Complete Graph K_{2p} ". *Graphs and Combinatorics* 5(1989), 351-353.
- [7] S. Nanda and N. Deo : "Methods for Placing Data and Parity to Tolerate Two Disk Failures in Disk Arrays Using Complete Bipartite Graphs". *CONGRESSUS NUMERANTUM* 179(2006), 167-179.
- [8] 宇野 俊夫 : 『ディスクアレイテクノロジー RAID』. エーアイ出版株式会社, 2000.