<table>
<thead>
<tr>
<th>Title</th>
<th>Detecting corrupted pages in M replicated large files</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Hwang, FK; Zang, W</td>
</tr>
<tr>
<td>Citation</td>
<td>IEEE Transactions On Parallel And Distributed Systems, 1997, v. 8 n. 12, p. 1241-1245</td>
</tr>
<tr>
<td>Issued Date</td>
<td>1997</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10722/42990">http://hdl.handle.net/10722/42990</a></td>
</tr>
<tr>
<td>Rights</td>
<td>©1997 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.; This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.</td>
</tr>
</tbody>
</table>
Detecting Corrupted Pages in $M$ Replicated Large Files

F.K. Hwang and Wenan Zang

Abstract—A file in a distributed database system is replicated on $M$ sites and may contain corrupted pages. Abdel-Ghaffar and El Abbadi gave a detection scheme assuming that the number of corrupted pages $f < M/2$. We replace this assumption by a much weaker one, that, for each page, the majority of copies are correct. Our schemes are based on the structure of the Reed-Solomon code, as proposed by Abdel-Ghaffar and El Abbadi for $M = 2$.

Index Terms—Data consistency, data corruption, fault detection, file comparison, Reed-Solomon code.

1 INTRODUCTION

It is often desirable to keep replicated copies of large files at remote sites to prevent loss of information and to permit easy access. However, data in a file can be corrupted due to various reasons. Thus, it is necessary to compare the files from time to time so that data faults can be detected. Metzner and Abidi [10] first proposed the method of combined signatures. Divide a file into pages of standard size. A binary parity sequence, called a signature, is derived for each page. It is assumed that two copies of a page agree if and only if their signatures agree. However, due to the typically large number of pages in a file, it is still too much work to exchange signatures on all pages for comparison. Instead, combined signatures are exchanged, where a combined signature is a weighted sum of a subset of signatures; when the subsets are properly chosen, all disagreeing pages can be detected.

Many schemes have been proposed (cf. [1], [2], [3], [6], [8], [9], [10], [11]) for detecting disagreeing pages between two replicated files. Let $N$ denote the number of pages and let $f$ denote the number of disagreeing pages, which is assumed to be known. In particular, Abdel-Ghaffar and El Abbadi [2], following an idea of Metzner and Kapturowski [11] of using linear block codes, showed that a one-round exchange of $\min\{N, 2f\}$ combined signatures suffice. Abdel-Ghaffar and El Abbadi [1] also proposed a detection scheme for $M$ replicated files. With $M \geq 3$, it is assumed that, for each page, the majority of copies agree (and are presumed to be correct). A corrupted page, or a fault, is now defined to be a page whose signature disagrees with the majority version. In one model, one site is designated as the coordinator, say site 1, while the other sites 2, ..., $M$, are referred to as the participants; the coordinator exchanges messages with the participants to detect all corrupted pages in the replicated file. Communication is between the coordinator and the participants, i.e., participants do not exchange any information. Abdel-Ghaffar and El Abbadi gave a detection scheme which transmits at most $(M - 2)\min\{N, f\} + \min\{N, 2f\}$ combined signatures (in possibly two rounds) under the constraint $M \geq 2f + 1$.

In many applications, $f$ is relatively small compared to $N$, but $M$ can also be a small number. Thus, the constraint $M \geq 2f + 1$ imposes a severe limit on the number of detectable faults. For example, for three sites, only one fault is allowed, even though there are thousands of pages. In this paper, we replace this constraint by the assumption that, for each page, the majority of copies are correct. Note that any model which does not presume an incorruptible copy and allows identical errors must make this assumption, lest the correctness of a page become indeterminable. Thus, our assumption is “minimal” in that sense. We shall also provide insights in the last section on how likely it is that this assumption will be violated and how our algorithms can be modified to detect the violation.

Let us introduce some notions before presenting our schemes. Throughout, $f_m$ stands for the number of corrupted pages at site $m$, $p_{n,m}$ with $1 \leq n \leq N$ and $1 \leq m \leq M$, stands for the $n$th page of the copy residing in site $m$, and $p_{n,m}$ stands for the signature of $p_{n,m}$. Let $\alpha$ be a primitive element in the finite field $GF(2^k)$. Each signature is composed of $b$ bits and, thus, can be considered as an element in $GF(2^k)$. Since, usually, $b$ is much larger than $\log_2 N$, we may assume that different pages have distinct signatures. For each site $m$ and for each positive integer $j$, we define the combined signature $\text{sig}_{j,m} = \sum_{n=1}^{N} p_{n,m} \alpha^n$. These signatures are the syndromes of a Reed-Solomon code corresponding to the vector $(p_{1,m}, \ldots, p_{N,m})$ [4]. Finally, given a vector $(e_1, \ldots, e_N)$, its weight is defined to be the number of nonzero entries. Our schemes will heavily rely on the following three facts:

**FACT 1.** For any given $E_1, \ldots, E_N \in GF(2^k)$, the system of equations $\sum_{n=1}^{N} e_n \alpha^n = E_j$, where $j = 1, 2, \ldots, N$, has a unique solution.

\[1045-9219/97/$10.00 © 1997 IEEE\]
FACT 2. If \((e_1, \ldots, e_N)\) has weight, at most, \(f\) and if \(\sum_{n=1}^{N} e_n \alpha^{m n} = 0\), where \(j = 1, 2, \ldots, f\), then \(e_n = 0\) for each \(n = 1, 2, \ldots, N\).

FACT 3. For any given \(E_1, \ldots, E_f \in \text{GF}(2^k)\), where \(1 \leq j \leq N\), the system of equations \(\sum_{n=1}^{N} e_n \alpha^{m n} = E_j\) for \(j = 1, \ldots, f\), has, at most, one solution of \((e_1, \ldots, e_N)\) with weight less than or equal to \(\lceil f/2 \rceil\). This solution can be obtained using a Reed-Solomon decoder [4].

It follows from Fact 3 that, if the number of disagreeing pages between any pair of sites \(i\) and \(j\), where \(1 \leq i, j \leq M\), is a clique. Note that the maximum clique problem can be solved in linear time by depth-first search (see, for instance, [12]). By assumption, the majority of sites agree on each page \(p\). Therefore, \(G_p\) contains a (unique) maximum clique \(C_p\) as \(\left| C_p \right| \geq \lceil (M + 1)/2 \rceil\), where \(\lceil x \rceil\) is the smallest integer not less than \(x\). Let \(F_m = \{p : m \notin C_p\}\). Clearly, \(F_m\) is the set of faults at site \(m\).

THEOREM 1. The above one-round scheme requires the transmission of \((M - 1)\text{min}[N, 2f]\) combined signatures to identify \(f\) corrupted pages. Moreover, \((M - 1)\text{min}[N, 2f]\) is the minimum number of combined signatures needed to be transmitted for any one-round scheme if \(N \geq f\).

PROOF. Abdel-Ghaffar and El Abbadi [2] proved that a site with \(f\) faults must transmit \(\text{min}[N, 2f]\) combined signatures to identify all \(f\) faults. Suppose that the \(M - 1\) participants transmit a total of fewer than \((M - 1)\text{min}[N, 2f]\) combined signatures. Then, there exists a participant transmitting fewer than \(\text{min}[N, 2f]\) combined signatures. Consider the case in which this participant contains \(f\) faults (we need the assumption \(N \geq f\) here). There is no guarantee that the \(f\) faults will be identified.

2 A One-Round Scheme

The algorithm at the coordinator \(s_1\) is shown below.

ALGORITHM (One-Round)

\begin{itemize}
  \item \textsc{send} (request \(\text{sig}_{1,m}, \ldots, \text{sig}_{\text{min}[N,2f],m}\)) to all participants \(m\)
  \item \textsc{receive} (\(\text{sig}_{1,m}, \ldots, \text{sig}_{\text{min}[N,2f],m}\)) from all participants \(m\)
  \item \textsc{for} each pair of sites \(i\) and \(j\), where \(1 \leq i, j \leq M\)
    \item compute a vector \((e_{1,i}, \ldots, e_{N,j})\) of weight, at most, \(f\)
    \item as a solution of the system of equations \(\sum_{n=1}^{N} e_{n,i} \alpha^{m n} = \text{sig}_{q,i} - \text{sig}_{q,j}\), where \(q = 1, 2, \ldots, \text{min}[N,2f]\)
  \end{itemize}

\begin{equation}
\textsc{end}(\text{for})
\end{equation}

\begin{itemize}
  \item \textsc{for} each page \(p\), where \(1 \leq p \leq N\)
    \item \textsc{construct} a graph \(G_p\) with vertex-set \([1, 2, \ldots, M]\)
    \item such that \([i, j]\) is an edge in \(G_p\) iff \(e_{p,i,j} = 0\)
    \item \textsc{find} a maximum clique \(C_p\) in \(G_p\)
  \end{itemize}

\begin{equation}
\textsc{end}(\text{for})
\end{equation}

\begin{itemize}
  \item \textsc{for} each site \(m\), where \(1 \leq m \leq M\)
  \item \textsc{return} \(F_m = \{p : m \notin C_p\}\) (\(F_m\) is the set of corrupted pages at site \(m\))
\end{itemize}

\begin{equation}
\textsc{end}(\text{for})
\end{equation}

Our algorithm is based on the scheme of Abdel-Ghaffar and El Abbadi [1]. Each participant sends \(\text{min}[N, 2f]\) combined signatures to the coordinator. It follows from Fact 1 and Fact 3 that, by comparing the \(\text{min}[N, 2f]\) combined signatures for any pair of sites \(i\) and \(j\), the coordinator can obtain the unique solution \((e_{1,i}, \ldots, e_{N,j})\) of (1) with weight, at most, \(f\). Since, for each page \(p\), \(e_{p,i,j} = 0\) if and only if site \(i\) and site \(j\) agree on page \(p\), the coordinator can thus identify all the disagreeing pages between any pair of sites \(i\) and \(j\). Let \(G_p\) be the graph constructed in the algorithm. Then, each connected component of \(G_p\) is a clique. Note that the maximum clique problem can be solved in linear time by depth-first search (see, for instance, [12]). By assumption, the majority of sites agree on each page \(p\). Therefore, \(G_p\) contains a (unique) maximum clique \(C_p\) as \(\left| C_p \right| \geq \lceil (M + 1)/2 \rceil\), where \(\lceil x \rceil\) is the smallest integer not less than \(x\). Let \(F_m = \{p : m \notin C_p\}\). Clearly, \(F_m\) is the set of faults at site \(m\). Let us proceed to the case \(N \geq f\).

ALGORITHM (Two-Round with \(N \geq f\))

\begin{itemize}
  \item \textsc{send} (request \(\text{sig}_{1,m}, \ldots, \text{sig}_{f,m}\)) to all participants \(m\)
  \item \textsc{receive} (\(\text{sig}_{1,m}, \ldots, \text{sig}_{f,m}\)) from all participants \(m\)
  \item \textsc{for} each pair of sites \(i\) and \(j\), where \(1 \leq i, j \leq M\)
    \item Try to compute a vector \(e_{i,j} = (e_{i,j,1}, \ldots, e_{N,j})\) of weight at most \(\lceil f/2 \rceil\) as a solution of the system of equations \(\sum_{n=1}^{N} e_{n,i,j} \alpha^{m n} = \text{sig}_{q,i} - \text{sig}_{q,j}\), where
\end{itemize}
Let \( w_{ij} \) denote the weight of the vector 
\((e_{1,i}, \ldots, e_{N,i,j})\) if such a solution exists, and 
let \( W \) be the set of all pairs \([i, j]\) of sites such that either 
(a) no solution of (2) with weight at most \( \left\lfloor f/2 \right\rfloor \) 
exists or 
(b) there is a solution of (2) with weight \( w_{ij} \) at most 
\( \left\lfloor f/2 \right\rfloor \) satisfying \( w_{ij} > w_{i,j} \) for any pair \([s, t]\) of 
sites disjoint from \([i, j]\).

\[ q = 1, 2, \ldots, f \]  

(2)

Let

\[ \alpha^m_q = \sigma_{q,t} - \sigma_{q,s} \]

where \( q = 1, 2, \ldots, \min\{N, 2\} \)

END(“FOR”)

FOR each page \( p \), where \( 1 \leq p \leq N \)

CONSTRUCT a graph \( G_p \) with vertex-set \([1, 2, \ldots, M] \) such that \([s, t]\) is an edge in \( G_p \) iff 
\( e_{p,s,t} = 0 \)

FIND a maximum clique \( C_p \) in \( G_p \)

END(“FOR”)

ELSE \( \Omega \neq \emptyset \)

IF \( \Omega = \{(i, j), (j, k), (k, i)\} \) for some three sites \( i, j, \) and 
\( k \)

SEND (REQUEST \( \sigma_{q,t}^{(1)}, \ldots, \sigma_{q,m}^{(\min\{N, 2\})} \)) to 
all sites \( i, j, \) and \( k \)

RECEIVE (\( \sigma_{q,t}^{(1)}, \ldots, \sigma_{q,m}^{(\min\{N, 2\})} \)) from all 
sites \( i, j, \) and \( k \)

FOR each pair \([s, t]\) \( \in \Omega \)

replace \( e_{s,t} \) by the 
solution \( (e_{s,t}, \ldots, e_{N,s,t}) \) with weight at most 
\( f \) of the system of equations

\[ \sum_{m=1}^{N} e_{n,s,t} \alpha^m_q = \sigma_{q,s} - \sigma_{q,t} \]

where \( q = 1, 2, \ldots, \min\{N, 2\} \)

END(“FOR”)

FOR each page \( p \), where \( 1 \leq p \leq N \)

CONSTRUCT a graph \( G_p \) with vertex-set 
\([1, 2, \ldots, M] \) such that \([s, t]\) is an edge in \( G_p \) iff 
\( e_{p,s,t} = 0 \)

FIND a maximum clique \( C_p \) in \( G_p \)

END(“FOR”)

ELSE some site \( i \) is contained in each pair in \( \Omega \)

FOR each page \( p \), where \( 1 \leq p \leq N \)

CONSTRUCT a graph \( G_p \) with vertex-set 
\([1, 2, \ldots, M] \) such that \([s, t]\) is an edge in 
\( G_p \) iff \( e_{p,s,t} = 0 \)

FIND a maximum clique \( C_p \) in \( G_p \)

END(“FOR”)

END(“FOR”)

END(“IF”)

END(“FOR”)

FOR each site \( m \), where \( 1 \leq m \leq M \)

RETURN \( F_m = \{p; m \in C_p\} \) (\( F_m \) is the set of 
corrupted pages at site \( m \))

END(“FOR”)

In the present case, instead of \( \min\{N, 2\} \) combined signa-
tures, only \( f \) combined signatures \( \sigma_{q,m} = \sum_{n=1}^{N} p_{n,m} \alpha^m_q \), 
are transmitted from each participant \( m \) to the coordinator, 
where \( 1 \leq q \leq f \). If a pair of sites contains no more than \( f/2 \) 
disagreeing pages, then the comparison of \( f \) combined signa-
tures between the pair is sufficient to detect all the disagree-
ging pages. Any pair violating this condition is called 
abnormal (other pairs are called normal).

For each pair of sites \( i \) and \( j \), the coordinator tries to 
compute a vector \( (e_{i,j}, \ldots, e_{N,i,j}) \) of weight at most \( \left\lfloor f/2 \right\rfloor \) 
as a solution of (2).

FACT 4. If \([i, j]\) is an abnormal pair, then \( [i, j] \) is in \( \Omega \) (def-
ined in the algorithm).

PROOF. We aim to prove that either (a) or (b) stated in 
the algorithm holds for \([i, j]\).

Let \( A \) denote the nonempty set of all abnormal 
pairs. Then \( A \) can not contain two disjoint pairs or 
the total number of faults would be greater than \( f \). 
If \([i, j]\) is an abnormal pair, then each pair \([s, t]\) dis-
joint from \([i, j]\) is normal, hence, there is a solution 
with weight \( w_{i,j} \) at most \( \left\lfloor f/2 \right\rfloor \) of (2) with \([s, t]\) in 
place of \([i, j]\). Assume that \((e_{i,j}, \ldots, e_{N,i,j})\) is a solu-
tion of (2) with weight no more than \( \left\lfloor f/2 \right\rfloor \). Since 
\( (p_{i,j} - p_{i,j}, \ldots, p_{N,i,j} - p_{N,j}) \) is a solution of (2) 
with weight greater than \( \left\lfloor f/2 \right\rfloor \), these two solutions are 
different. Note that \( \sum_{n=1}^{N} (e_{n,i,j} - (p_{n,i} - p_{n,j})) \alpha^m_q = 0 \), 
where \( q = 1, 2, \ldots, f \). By Fact 2, there are at least \( f + 1 \) 
values of \( n \) for which \( e_{n,i,j} \neq p_{n,i} - p_{n,j} \). For each two
sites \( k \) and \( l \), let \( d_{ij} \) denote the number of disagreeing pages between them. Then, \( w_{ij} + d_{ij} \geq f + 1 \), whence \( w_{ij} + f_i + f_j \geq f + 1 \). So, for each pair \( \{s, t\} \) disjoint from \( \{i, j\} \), we have \( w_{ij} \geq f - f_i - f_j + 1 > f_i + f_j \), the last equality holds because \( \{i, j\} \) is an abnormal pair, \( f_i + f_j > \left\lfloor \sqrt{f} \right\rfloor \), hence, \( d_{ij} \leq f_i + f_j \leq \left\lfloor \sqrt{f} \right\rfloor \). \( \square \)

Using the fact that \( \Omega \) cannot contain two disjoint pairs, we immediately have the following statement.

**FACT 5.** Let \( \{i, j\} \) be an arbitrary pair of sites in \( \Omega \). Then one of the following three cases occurs.

- Case 1. There is a site \( k \) outside \( \{i, j\} \) such that \( \Omega = \{i, j, k\} \).
- Case 2. Each pair in \( \Omega \) contains site \( i \).
- Case 3. Each pair in \( \Omega \) contains site \( j \).

Let us consider Case 1, now. By Fact 4, any pair of sites outside \( \Omega \) is normal. By Fact 3, all the disagreeing pages between any two sites in \( \{i, j, k\} \) are detected. For each ambiguous page \( p \), the total number of combined signatures transmitted is \( \min(N - f, f) \) combined signatures to identify \( f \) corrupted pages.

**THEOREM 2.** The above two-round scheme requires the transmission of at most

\[
(M - 1)\min(N, f) + \min(1 + \left\lfloor \sqrt{f} \right\rfloor, 1 + \left\lfloor (M + 1)/2 \right\rfloor)\min(N - f, f)
\]

combined signatures to identify \( f \) corrupted pages.

In Case 2 and Case 3, we proposed to find a minimal (with respect to set inclusion rather than size) set \( \Delta \) of sites excluding \( i \) such that \( \Delta \cap C_p \neq \emptyset \) for each nonambiguous page \( p \). Clearly, the smaller the size of \( \Delta \), the more efficient the algorithm. The reason why we did not try to find a minimum \( \Delta \) (with respect to size) with the desired property is as follows: Let \( T \) denote the set of all nonambiguous pages. For each site \( m \neq i \), let \( T_m \) denote the set of all nonambiguous pages \( p \) such that \( m \in C_p \). To obtain \( \Delta \), we have to solve the following problem:

**Input:** A finite set \( T \) and a family \( \{T_m: m \neq i\} \) of subsets of \( T \)

such that \( T = \cup_{m \neq i} T_m \).

**Output:** A minimum set \( \Delta \) such that \( T = \cup_{m \in \Delta} T_m \).

which is the well-known set-covering problem and, therefore, \( \text{NP} \)-hard. Johnson [7] showed that a greedy heuristic algorithm returns a set cover with a ratio bound of \( \ln |T| + 1 \). Feige [5] proved that the set-covering problem can not be approximated within \( (1 - \epsilon)\ln |T| \) for any \( \epsilon > 0 \), unless \( \text{NP} \subseteq \text{DTIME}(n^{n^{o(\log n)}}) \).

**4 Remarks**

Our algorithms are based on the assumption that for each page, the majority of copies are correct. Without this assumption, we are unable to distinguish between the correct pages and the corrupted pages, and thus, detection failure occurs. Let us give a rough estimate of the probability that
the assumption fails. For \( f \geq \lceil M/2 \rceil \), the probability that a given page has at least \( \lceil M/2 \rceil \) corrupted pages is

\[
\sum_{i=\lceil M/2 \rceil}^{\min(M,f)} \frac{M!}{i!(M-i)!} \left( \frac{N-1}{f-i} \right)^i \frac{NM}{f} - i, 
\]

which is dominated by the term \( i = \lceil M/2 \rceil \); this term is roughly \( \left( \frac{M}{f+1} \right)^{f+1} \). Multiplying this by \( N \), we get an estimate of the probability that there exists a page with at least \( \lceil M/2 \rceil \) corrupted copies. For \( N = 100, f = M = 3 \), this estimate is 0.03; when \( f = M = 5 \), it is 0.001. For \( N = 1,000, f = M^2 = 9 \), this estimate is 0.036; when \( f = M^2 = 25 \), it is 0.0023. For \( N = 1,000, f = 0.01NM = 30 \), this estimate is 0.45; when \( f = 0.01NM = 50 \), it is 0.021.

Observe that our algorithms can be modified to detect the case when a majority of copies of some page are corrupted.

In the one-round scheme and in the cases \( \Omega = \emptyset \) and \( \Omega = \{i, j, k, \ldots \} \) in the two-round scheme, after obtaining \( C_p \) for each page \( p \), let us check the size of \( C_p \). If \( |C_p| < \lceil (M+1)/2 \rceil \), we declare detection failure; otherwise, output \( F_m \) for each site \( m \).

In the two-round scheme, in case some site \( i \) is contained in each pair in \( \Omega \), let us modify \( S \) to be a minimal subset of \( \{1, 2, \ldots, M\} - \{i\} \) that intersects every maximum clique \( C \) with \( |C| \geq (M-1)/2 \) in each \( C_p - \{i\} \). Possibly some \( G_p - \{i\} \) contains two maximum cliques of size \( (M-1)/2 \). Since each maximum clique is a connected component in \( G_p - \{i\} \), we can find these cliques by depth-first search. Then, we modify the loop “for each page \( p \) with \( |C_p| = (M-1)/2 \), where \( 1 \leq p \leq N \) ...” (right above the loop for returning \( F_m \)) as follows: For each page \( p \) with \( C_p = (M-1)/2 \), let \( e_{p,t} = 0 \) for each \( t \in S \cap C_p \). If \( e_{p,j} = 0 \), then set \( C_p = C_p \cup \{j\} \); if \( e_{p,j} \neq 0 \), then check if \( \{1, \ldots, M\} - (C_p \cup \{i\}) \) is a clique. If yes, denote this clique by \( C \) and take an arbitrary \( t' \in S \cap C \). If \( e_{p,j, t'} = 0 \), then set \( C_p = C_p \cup \{i\} \).

After obtaining \( C_p \) for each page \( p \), let us check the size of \( C_p \). If \( |C_p| < \lceil (M+1)/2 \rceil \) for some page \( p \), we declare detection failure; otherwise, output \( F_m \) for each site \( m \).

In this paper, we proposed two schemes for efficient detection of corrupted pages in a large replicated file. Our approach is based on the previous schemes of Abdel-Ghaffar and El Abbadi [1], [2]. In these two papers, the lower bounds of the communication complexity of the coordinator-based model were derived and proved to be tight. However, the communication complexity of our two-round scheme is bigger than the lower bound presented in the previous papers. The tight lower bound for the present model remains unknown. We close the paper with this open problem.

ACKNOWLEDGMENTS

Wenan Zang’s research was supported in part by CRCG grant 33/024/0010.

REFERENCES


Wenan Zang obtained his PhD degree in operations research from Rutgers University in 1995. He has been an assistant professor at the University of Hong Kong since 1996. His main research interests are in combinatorics and optimization.