

ハードウェア障害統計情報の集約方式

3Y-5

山本 邦雄* 矢澤 良一* 安保 進**
 *NTTソフトウェア(株) **NTT情報システム本部

1. はじめに

HARTSシステムにおいて、各サイトで収集されるハードウェア障害情報は膨大な量となることから、各サイトから遠隔監視センタへの障害情報転送に伴う通信コストの負担も多大となる。このため、転送データ量の削減を図る必要があった。そこで、ハードウェア障害情報のうち、障害統計情報について情報量の削減方式の検討を行った。本稿ではその集約方式の概要と適用効果について報告する。

2. ハードウェア障害統計情報とは

ハードウェア障害統計情報は図1に示す形式からなり、リードバイト数等の統計項目のいずれかの項目でオーバーフローが発生した時点で統計情報としてログファイルに出力される。従って、使用頻度の高い装置では統計情報が頻繁に出力されることとなる。

F5L71 #02 ... PDN=F000 LDN=L EID=EE ... CSW=CCC... SSB=aaaabbcdeeff.....	
F5L71:統計情報を示すID	c :再試行データチェック数
F000:装置番号	d :シークエラー回数
aaaa:リードバイト数	ee:シーク回数
bb :訂正可能データチェック数	ff:シーク移動回数

図1 ハードウェア障害統計情報の形式

3. 方式概要

統計情報量を削減する方法として以下のようなものが考えられる。

- ①コード化圧縮・・・使用頻度の大きいデータ値に短いコードを割り付け、小さくなるに従って長いコードを割り付ける
- ②冗長データ圧縮・・・不要な冗長データを削除する
- ③意味圧縮・・・特定文字列を単一文字または特定形式に置き換える (ex. 'PDN=xxxx'→'Axxxx')
- ④置き換え圧縮・・・使用頻度の大きい文字の文字列を特定形式に置き換える (ex. '0000'→'470')
- ⑤集約方式・・・同じ意味・形式を持つ複数の情報を単一の情報に集約する

これらの圧縮方法に関する比較検討結果を表1に示す。

表1 情報圧縮方式各案の比較

方式	長所	短所	評価	評価理由
①コード化圧縮	・全データを対象とすると効果大	・伸張が必要 ・メモリ使用率が高い	×	サイト資源が高負荷となる
②冗長データ圧縮	・2つ以上の区切りのブロックを1つとすることで効果あり	・参照時に相対位置で参照している場合に影響あり ・伸張ができない	×	削減効果が少ない(10%以下)
③意味圧縮	・特定データを対象とすると効果あり	・復元が必要 ・処理が複雑	×	処理がやや難
④置き換え圧縮	・圧縮処理が単純 ・変換辞書が不要 ・対象文字列がゼロだけでも、かなりの効果あり	・伸張が必要 ・置き換え文字を多くすると、圧縮/伸張が遅くなる	△	遠隔監視センタでの伸張プログラムが必要となる
⑤集約方式	・データ形式、内容を損なわない(復元不要) ・サイト資源をほとんど使用しない	・統計情報のみの圧縮となる	○	遠隔監視センタでの伸張プログラムが不要である

A compress method of hardware-statistics data

Kunio YAMAMOTO*, Ryoichi YAZAWA*, Susumu ANBO**

*NTT Software Corporation, **NTT Information Systems Head quarters

HARTSでは以下の観点から、⑤の集約方式を採用した。本方式は、元のデータ形式を変えることなく、情報量の削減を実現している。

- ・分析に必要な情報について、意味・内容を損なわないように情報量を削減できること
- ・圧縮、復元の処理に伴うシステムの資源負荷（CPU・メモリ）が無視できるレベルであること

4. 処理方式

本方式の処理方式は以下のとおりである。

- (1) 障害統計情報を格納するテーブル（以下、エントリテーブルと示す）をメモリ上に確保する。（エントリテーブル内のエントリ数はシステム内装置数と統計情報発生頻度により決定する）
- (2) ログファイルの先頭より読み込み、障害統計情報が存在するまで読み飛ばす。
- (3) 障害統計情報が存在し、かつ、該情報がエントリテーブル上に存在しない場合は、該情報をエントリテーブル上に登録する。[図2①]
- (4) 該情報が既にエントリテーブル上に存在する場合は、必要な項目のみ加算を行い、それぞれの項目内において、桁あふれ（以下、オーバフローと示す）が発生するかどうかを調べる。[図2②]
- (5) (4)の結果、オーバフローすると判断した場合は、エントリテーブル上の情報を障害情報抽出ファイル（以下、抽出ファイルと示す）に出力し、そのエントリ位置に該情報を登録する。
- (6) (4)の結果、オーバフローしないと判断した場合は、必要な項目のみ加算を行い、元のエントリ位置に加算結果後の情報を登録[図2③]する。また、発生回数を1加算[図2④]する。
- (7) エントリテーブル上の登録情報数がエントリ数の上限を超えた場合は、最初に検出した発生回数が一番少ない情報を抽出ファイルに出力し、そのエントリ位置に該情報を登録する。
- (8) 日跨りが発生した場合とログファイル終了時には、エントリテーブル上の全情報を抽出ファイルに抽出する。[図2⑤]（HARTSでの分析単位は1日単位としているため、日跨り毎にエントリテーブルを初期化する）

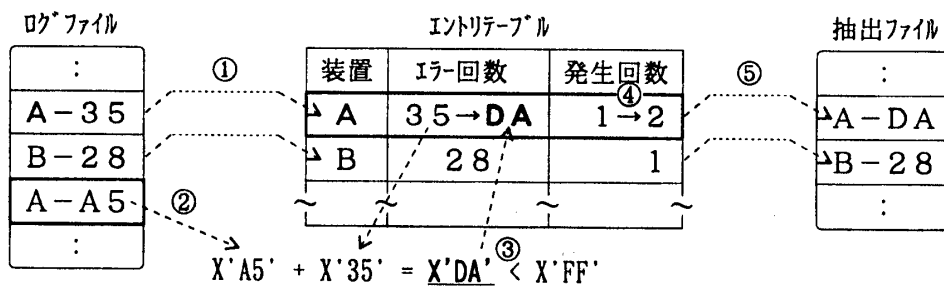


図2 集約方式の概略図

5. 適用効果

障害統計情報は1装置当たり1日に2~5回発生するが、本方式によりこれが1つとなる。システムを構成する装置の台数、種類、稼働時間および負荷状況によるが、あるシステムにおいて、本方式適用前後の情報量を比較したところ以下の通りであった。

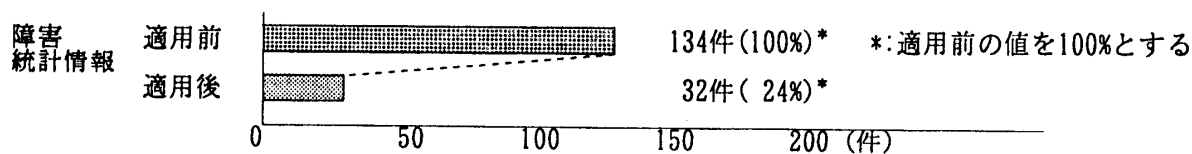


図3 本方式適用前後の情報量

6. おわりに

ハードウェア障害統計情報の集約方式について述べた。本方式の適用により、具体事例としてハード障害統計情報が76%削減でき、その有効性が確認できた。今後、具体事例に基づいた評価を継続するとともに、さらに改良を進めていく予定である。