Linuxファイルシステムの信頼性検証技術

電源断が起きても安心できるシステムの構築を目指して

近年，Linux(注1)の組込みシステムへの適用が注目されていますが，適用に際しては品質面の十分な評価が必要です。例えば，データをハードディスク装置（HDD）などに記録しているシステムがありますが，それらのシステムでは，不慮の電源断などが起きても，書き込んだデータが壊れないようにしなければなりません。しかし，この要求にLinuxがどの程度まで応えられるかという一般的な解はありません。また，一般的な指標を得ることが難しいため，個々に評価する必要がありますが，そのためには膨大な人手と時間が必要です。

東芝は，Linuxのファイルシステムのデータ整合性を自動的に評価できる環境を構築しました。また，評価結果はCE Linux Forumなどを通して公開しています。

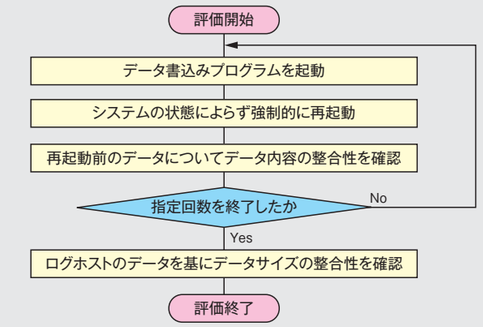

図1．データ信頼性評価の流れ ― 評価は数千回行われることもありますが，すべての流れは自動化されています。

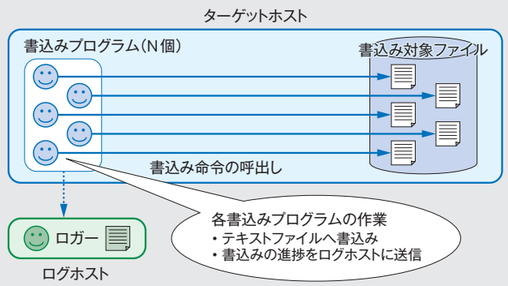

図2．データ書込みプログラムと書込み進捗確認プログラムの関係 ― 書込みを行うプログラムが動作中にもかかわらず，強制的にシステムを再起動させてしまうことで，データが不安定となる状況を作り出しています。

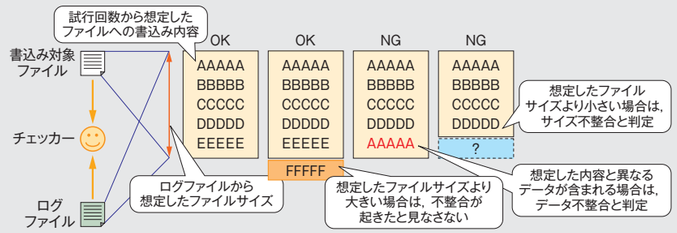

図3．データ整合性の確認方法 ― チェッカーはファイル内容の整合性とファイルサイズの整合性について，それぞれのログを基に自動的に検査を行います。

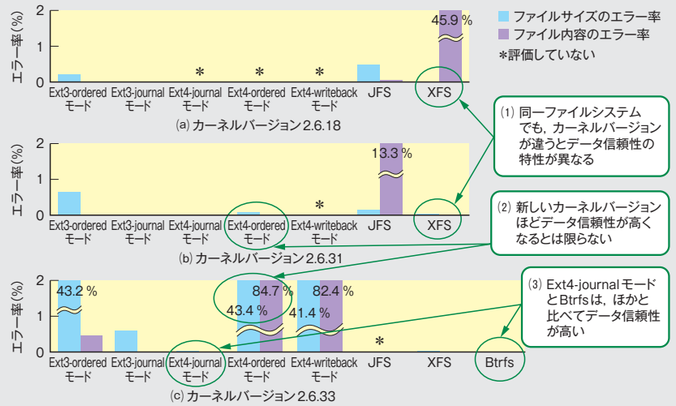

図4．各種ファイルシステムのデータ信頼性評価結果 ― 三つのバージョンのLinuxカーネルを利用して評価しています。ファイルシステムの種類やカーネルバージョンの違いによって，データの信頼性に大きな差があることがわかります。グラフの横軸はファイルシステムの種類です。

ファイルシステムとデータの信頼性

オープンソース ソフトウェアの代表例とも言えるLinuxは，目覚ましい進化を遂げており，現在，約3か月に一度Linuxカーネル(注2)のバージョンアップが行われています。その際，新しい機能の追加や不具合への対応が行われますが，こうした変更には，ファイルシステムに関連するものも含まれます。

Linuxで提供されるファイルシステムは，実に多様です。例えば，読み書きできるデバイスを対象としたものだけでも10種類程度あり，システムによって適切なものを選択して利用します。このとき，利用方法やデバイスとの相性，パフォーマンスなどの要素を総合的に判断する必要があります。そして，その要素の一つに，“データの信頼性”があります。データの信頼性とは，システムが電源断や強制リセットなどによって突然停止してしまった場合でも，それまでにアプリケーションプログラムが書き込んだと判断したデータが正しく書き込まれていることを保証するものです。

（注1） Linuxは，Linus Torvalds 氏の米国及びその他の国における登録商標。
（注2） カーネルは，オペレーティングシステムの中核となるソフトウェアで，CPU，メモリ，HDDなどのシステムの資源を管理するとともに，アプリケーションから使えるようにするためのインタフェースを提供している。

信頼性評価への環境構築

評価の流れとそれに使用した環境は，それぞれ**図1**と**図2**のようになります。評価対象のシステム（ターゲットホスト）では，データを書き込み続けるプログラムが動作しており，このシステムをランダムなタイミングで強制的にリセットして，電源断に似た状況を作ります。このとき，データ書込み進捗（しんちょく）の取得方法と不正終了後のデータ整合性の確認方法が問題となります。

まず，データ書込み進捗の取得方法ですが，この情報を同一のシステム上に保存した場合，その進捗状況に関係するデータは信頼できません。なぜなら，不正終了したシステムではデータが正しく書き込まれている保証がないからです。そこで，データ書込み進捗情報は別のシステム（ログホスト）で取得するようにしました。

ターゲットホストで動作する書込みプログラムは，データを書き込むたびに書込み回数の情報をログホストに送信します。ここで，書込み後に書込み回数送信という順番を必ず守ることで，強制リセットで突然システムが停止しても，書込み回数を少なく見積る不整合が起きなくなります。

次に，不正終了後のデータ整合性の確認では，書込み内容の整合性（データ整合性）と，ファイルサイズの整合性（サイズ整合性）の評価を行います（**図3**）。

データ整合性の評価では，ファイルの内容をすべてチェックし，実際に書き込まれたデータが想定するものと一致するかどうかを確認します。ここで，一致しなかった場合は，データ不整合としてエラーとなります。

サイズ整合性の評価では，まずターゲットホストのファイルから書込み回数を算出し，次にログホスト上に記録された書込み回数と比較します。ここで，ターゲットホストの書込み回数のほうがログホストに記録されたものより少ない場合には，サイズ不整合としてエラーとなります。逆に，書込み回数が一致するか，ターゲットホストの書込み回数のほうが多い場合は，正しいと判定します。

なお，ターゲットホストの書込み回数のほうが多い場合を許容しているのは，書込み後に書込み回数送信という順で処理されるためです。書込み直後にリセットした場合，ログホスト上には最後に書込みを行ったときの情報が記録されません。しかし，最後の書込みが実際に行われたかどうかは，ターゲットホスト上のファイルから検出することができるため問題とはなりません。

取得データから見るLinuxファイルシステムのデータ信頼性

前の章で述べた評価方法に基づき，実際に三つのバージョンのLinuxカーネルで，ファイルシステムの評価を複数回行い，平均を取りました（**図4**）。Ext4ファイルシステムのjournalモードとBtrfsで良い結果が得られています。これらのファイルシステムは，動作原理上データ信頼性が高いと言われており，実際にそれを示しています。しかし，Ext4ファイルシステムのorderedモードやXFSのように，カーネルのバージョンの差により，特性が大きく異なるものもあります。

例えば図4で，カーネルバージョン2.6.33のExt3ファイルシステム及びExt4ファイルシステムのorderedモードのエラー率が大きく上がっています。これは2.6.31から2.6.32の間の変更で，ファイルシステムの下層にあたるブロックデバイス層が大幅に書き換えられたときに混入したバグが原因です。このようなケースでは，ファイルの読み書きが正しく行われているように見えるため，気づくことは困難です。そこで，システム適用前に評価を実施することで，データの信頼性が高いシステムを提供できます。

今後の展望

今後は，ファイルシステムだけでなく，Linuxのほかの機能に対しても，信頼性の面からの評価を行い，Linuxを利用するシステムの品質向上に取り組んでいきたいと考えています。

小林　良岳
ソフトウェア技術センター
先端ソフトウェア開発担当主務