

Pkprof: プロセスごとに分類するカーネルプロファイラ

7M-4

— 4.4BSD 上での実装 —

西岡真吾 川口敦生 元田浩†

(株)日立製作所基礎研究所

1. はじめに

本稿では、Pkprof(Process-labeled Kernel PROFiler) [1][2]を 4.4BSD-Lite オペレーティングシステム上に試験的に実装した例(以下、PKPROF と呼ぶ)について述べる。使われた技術の多く(モニター、サンプラ、プロファイルデータバッファ管理など)は、4.4BSD-Lite カーネルとは独立であり、他のオペレーティングシステムで pkprof を実装する際にも有用であろう。

2. 概要

4.4BSD-Lite カーネルは、KPROF と呼ばれるカーネルプロファイラを提供している。KPROF は、

- 各カーネルルーチンが呼び出された回数
- 各カーネルルーチンが実行に費やした時間の二つを計測する。

前者は各カーネルルーチンに埋め込まれたモニターで、また後者はCPUが実行中の機械語のアドレスをサンプリングするサンプラによって計測される。この KPROF を改造して PKPROF を実装した。

4.4BSD-Lite カーネルは、次の三つの事象が起こったときに、カーネルモードに移行する。

1. プロセスがシステムコールを発行したとき
2. プロセスが物理メモリが割り当てられてないあるいは不正なアドレスを参照したとき
3. 周辺機器が割り込みを起こしたとき

Pkprof: Process-labeled Kernel Profiler — Sample Implementation on 4.4BSD-Lite Operating System —
Shingo Nishioka, Atsuo Kawaguchi, and Hiroshi Motoda
Advanced Research Laboratory, Hitachi, Ltd.

Akanuma, Hatoyama, Hiki, Saitama, 350-03 Japan.

†現在、大阪大学産業科学研究所

1, 2 が起こったときの処理は現在実行中のプロセスのために行われるものと見なせる。一方 3 が起こったときの処理は、必ずしも現在実行中のプロセスのためとは限らない。

プロファイルデータバッファの切り替えを制御するために、新しく二つの広域変数、svdproc と svdevent を加え、それらの値を更新するコードを必要個所に埋め込むこととした。前者は現在カーネルがサービスしているプロセスを示し、後者は同様に非同期事象を示す。これら二変数の値に応じてプロファイルデータバッファを選択するようモニターとサンプラを変更した。

3. サービス中のプロセスの追跡

図1は、4.4BSD-Lite カーネルのコンテキスト切り替えに関連するカーネルルーチンの呼び出し関係を示したものである。図中、*を付したルーチンが現在実行中のプロセスを切り替える。これらに svdproc および svdevent の値を更新す

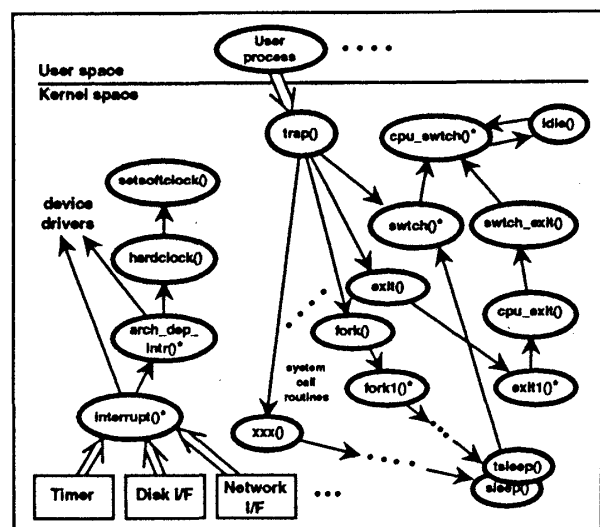


図1 コンテキスト切り替えに関連するカーネルルーチンの呼び出し関係

るコードを埋め込んだ。割り込みを受け付ける `interrupt()` では `svdproc` の値を `NULL` にして、以後の処理の計測データが非同期事象計測データバッファに格納されるようにした。

4. 非同期事象の追跡

図2は、4.4BSD-Lite カーネルにおける割り込み処理の概略を示したものである。割り込みは、まず最下層の割り込みハンドラによって受け取られ、その種類に応じて適当なデバイスドライバの割り込み処理ルーチンが呼び出される。割り込み処理ルーチンは処理中に `buf` 構造体や `mbuf` 構造体といった内部データを作成する。処理の最後に割り込み処理ルーチンはこれらの内部データを適当なキューに格納し、`wakeup()` を呼んでそのキューで `sleep()` していたプロセスを実行可能状態にする。

PKPROF では、処理中に作成される内部データのアドレスを非同期事象の識別子に用い、ハッシュ表を使って非同期事象計測データバッファと関連付けた。内部データを切り替える箇所すべてに、

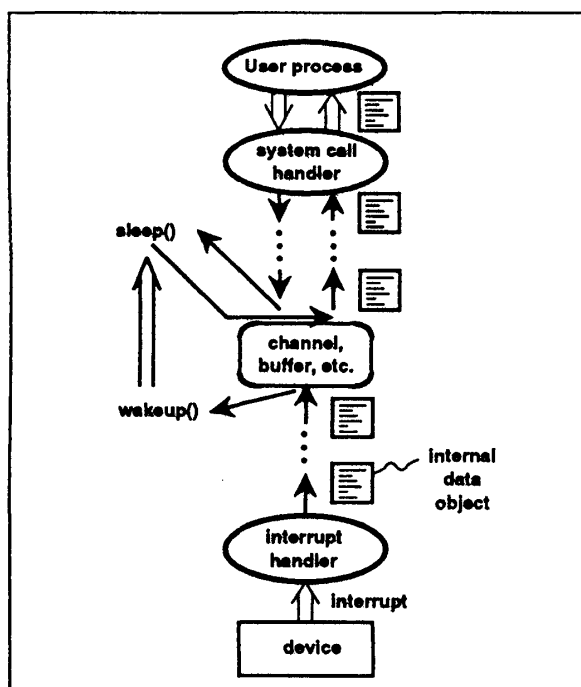


図2 割り込み処理の概要

`svdevent` が適切な非同期事象計測データバッファを指すよう更新するコードを埋め込んだ。

また `wakeup()` 後、実行を再開したプロセスが内部データを参照するところに、その内部データに関連付けられた計測データをプロセスのプロファイルデータバッファに移すコードを埋め込んだ。すなわち PKPROF では、割り込み処理にかかった資源消費の計測データを、それによって作成された内部データを最初に参照するプロセスのものとして分類する。

5. オーバヘッド

モニターの実行時間は、KPROF のものと比較して 16% 増加した。サンプラーの実行時間増加も同程度と推定している。プロセス切り替えに関連するルーチンの実行時間は、PKPROF なしのとときと比較して 0.04% 増加した。

非同期事象の計測のオーバヘッドは、事象の種類に強く依存した。ディスク装置の読み出しスループットは、ほとんど低下しなかった。一方 Ethernet から UDP パケットを受信するスループットは、PKPROF によって 50-80% の低下を見た。

6. おわりに

PKPROF のコードは下記の Web ページで公開している。

<http://www.har1.hitachi.co.jp/~nis/>

参考文献

- [1] S. Nishioka, A. Kawaguchi, and H. Motoda, "Process-labeled kernel profiling: a new facility to profile system activities", 1996 USENIX Technical Conference Proceedings, 1996.
- [2] 川口他, 「Pkprof: プロセスごとに分類するカーネルプロファイラ — 概念 —」, 情報処理学会第 52 回全国大会論文集 (講演番号 7M-3), 1996.