Pkprof：プロセスごとに分類するカーネルプロファイラ — 概念 —

7M－3

川口敦生　西岡真吾　元田浩†
(株)日立製作所基礎研究所

1. はじめに

Pkprof(Process-labeled Kernel PROFiler)は，オペレーティングシステムのカーネルが，どのプロセス(あるいはスレッド)のためにどれだけの資源を消費したかを測定する，カーネルプロファイラである[1].

アプリケーションプログラムの開発に用いられるユーザレベルプロファイラと，オペレーティングシステムカーネルの開発に用いられるカーネルプロファイラは，これまで個別に用いられてきた．ユーザレベルプロファイラは，システムコールの処理を不可分の処理として扱う．一方カーネルプロファイラでは，特定のプロセスによって一時的にもたらされた過負荷を分離して扱うことは困難であった．

近年のオペレーティングシステムは，サーバと呼ばれる多くのユーザレベルプログラムがシステムレベルのサービスを提供している．こういった構成では，カーネルレベルとユーザレベルのプロファイル結果を同時に検討することが，システム構成の決定や性能改良に有益である．

Pkprof はこのような検討のもとに考案された新しいカーネルプロファイラである．Pkprof は，今カーネルがどのプロセスのためにサービスを実行しているかを追跡する．そしてどのプロセスのためかという情報を，サービス実行に要した資源消費にラベル付けして記録する．

Pkprof: Process-labeled Kernel Profiler — Concepts —
Atsuo Kawaguchi, Shingo Nishioka, 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

```
                                  called/total       parents
index  %time    self  descendents  called+self    name           index
                                  called/total       children

.....

                0.00        2.86      65/65          137u:raw_cat [7]
[8]     30.2    0.00        2.86         65        137u:read [8]
                0.00        2.86      65/65          137k:read [9]

-----------------------------------------------

                0.00        2.86      65/65          137u:read [8]
[9]     30.2    0.00        2.86         65        137k:read [9]
                0.01        2.85      65/65          137k:vn_read [10]

-----------------------------------------------

                0.01        2.85      65/65          137k:read [9]
[10]    30.2    0.01        2.85         65        137k:vn_read [10]
                0.00        2.84      65/76          137k:nfs_read [2]
                0.00        0.02     65/163          137k:lease_check [90]
                0.00        0.00     65/105          137k:nfs_lock [282]
                0.00        0.00     65/101          137k:nfs_unlock [283]
```

図1 継ぎ目なしのプロファイルの例

ユーザレベルプロファイラと併用すると，一つのプログラムがユーザ空間ならびにカーネル空間で消費した資源を計測することが可能になる．このような「継ぎ目なしのプロファイル」(Seamless profile)の測定結果例を図1に示す．この例ではプロセス137が raw_cat() 関数の実行に費やしたCPU時間の内訳が，gprof[2]形式でカーネル内の処理(137k:と付された関数)も含めて示されている．継ぎ目なしのプロファイルは，システム全体の視点からプログラムの性能改善を進めていくのに極めて有用である．

2. システム内で起こる事象と Pkprof

システムコールの処理はユーザプロセスのシステムコール発行で開始され，カーネル空間内で実行される．したがって，そのような実行による資源消費は，システムコールを発行したプロセスに帰することができる．このようなプロセスの実行に同期して生じるカーネル空間内での資源消費のきっかけとしては，他に仮想記憶を実現するためのページフォールト処理などがあげられる．

カーネルは周辺機器からの割り込み処理も担

当している．例えばネットワークインタフェースはパケットを受信すると割り込みをかけ，通常の実行を中断する．カーネルはパケット受信処理を行った後，中断された実行を再開する．このパケット受信処理に要した資源消費は，受信したパケットを利用したプロセスに帰することができる．ただしこのようなプロセスの実行と非同期に起こる資源消費にラベル付けを行うには，その資源消費によって恩恵を得るプロセスが判明するまで待たねばならない．

3. 基本構造

Pkprof は従来のプロファイラと同様に，各コードブロック（例えば関数）に埋め込まれるコード（モニターと呼ぶ）と現在の実行箇所をサンプリングするサンプラーから構成できる．モニターやサンプラーの計測データは，各プロセス毎に用意されたプロファイルデータバッファに記録する．

同期事象の処理の開始およびプロセスの切り替えを行うすべてのカーネルコードにプロファイルデータバッファを適切に切り替えるためのコードを埋め込む．

非同期事象の処理に要した資源消費については先に述べたように，ラベル付けを遅らせる必要がある．そのためにまず，非同期事象計測データバッファを設ける．非同期事象にシステム内で唯一の識別子を付与し，同期事象の計測と同様にこの識別子を使って非同期事象計測データバッファを切り替える．処理の結果をあるプロセスが利用した時点で，識別子に対応する非同期事象計測データバッファに格納された計測データを取り出し，そのプロセスの計測データに追加する．このようにして，非同期事象の処理に要した資源消費の測定データは，その処理の恩恵を受けたプロセス分と分類される．なお，この場合の唯一の識別子としては，例えば非同期事象の処理の際に生成される内部データの先頭アドレスなどが考えられる．

4. 実現

カーネルプロファイラを既に持っているオペレーティングシステムで pkprof を実装するのは，比較的容易な作業である．ただし非同期事象処理の計測については，非同期事象の処理が何レベルかにわたって行われる場合（例えばネットワークのプロトコル処理）など，実装が複雑になることが考えられる．また，非同期事象計測データバッファ用メモリの確保，再利用や恩恵を受けるプロセスが判明したときの処理が高速にできなければ，pkprof なしの時の動作状況との乖離が大きくなることが予想される．

5. おわりに

本稿では pkprof の基本的な考えについて述べた．具体的な実装例については，4.4BSD-Lite オペレーティングシステムに pkprof を実装した例を別稿[4]に述べたので，参照されたい．

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] L. Graham, P. B. Kessler and M. K. McKusick, "gprof: a Call Graph Execution Profiler", ACM SIGPLAN Notices 17(6), 120-126, 1982.

[3] J. Hall and A. J. Goldberg, "Call Path Profiling of Monotonic Program Resources in UNIX", 1993 Summer USENIX Technical Conference Proceedings, 1-13, 1993.

[4] 西岡他，「Pkprof:プロセスごとに分類するカーネルプロファイラ —4.4BSD 上での実装 —」，情報処理学会第 52 回全国大会論文集（講演番号 7M-4），1996.