



# Red Hat Enterprise Linux 6

## SystemTap ビギナーズガイド

SystemTap 入門



# Red Hat Enterprise Linux 6 SystemTap ビギナーズガイド

---

SystemTap 入門

Robert Krátký

Red Hat Customer Content Services

rkratky@redhat.com

Mirek Jahoda

Red Hat Customer Content Services

mjahoda@redhat.com

Don Domingo

Engineering Services and Operations Content Services

William Cohen

Engineering Services and Operations Performance Tools

Red Hat, Inc.

## 編集者

Jacquelynn East

Red Hat Engineering Content Services

## 法律上の通知

Copyright © 2016 Red Hat, Inc. and others.

This document is licensed by Red Hat under the [Creative Commons Attribution-ShareAlike 3.0 Unported License](https://creativecommons.org/licenses/by-sa/3.0/). If you distribute this document, or a modified version of it, you must provide attribution to Red Hat, Inc. and provide a link to the original. If the document is modified, all Red Hat trademarks must be removed.

Red Hat, as the licensor of this document, waives the right to enforce, and agrees not to assert, Section 4d of CC-BY-SA to the fullest extent permitted by applicable law.

Red Hat, Red Hat Enterprise Linux, the Shadowman logo, the Red Hat logo, JBoss, OpenShift, Fedora, the Infinity logo, and RHCE are trademarks of Red Hat, Inc., registered in the United States and other countries.

Linux<sup>®</sup> is the registered trademark of Linus Torvalds in the United States and other countries.

Java<sup>®</sup> is a registered trademark of Oracle and/or its affiliates.

XFS<sup>®</sup> is a trademark of Silicon Graphics International Corp. or its subsidiaries in the United States and/or other countries.

MySQL<sup>®</sup> is a registered trademark of MySQL AB in the United States, the European Union and other countries.

Node.js<sup>®</sup> is an official trademark of Joyent. Red Hat is not formally related to or endorsed by the official Joyent Node.js open source or commercial project.

The OpenStack<sup>®</sup> Word Mark and OpenStack logo are either registered trademarks/service marks or trademarks/service marks of the OpenStack Foundation, in the United States and other countries and are used with the OpenStack Foundation's permission. We are not affiliated with, endorsed or sponsored by the OpenStack Foundation, or the OpenStack community.

All other trademarks are the property of their respective owners.

## 概要

本ガイドでは、SystemTap を使用して Red Hat Enterprise Linux の各種サブシステムを監視する基本的な方法を詳細に説明します。

## 目次

<b>前書き</b> .....	<b>3</b>
1. ドキュメント規則	3
1.1. 誤字規則	3
1.2. 引用規則	4
1.3. 注記および警告	5
2. ヘルプの利用とフィードバック提供	5
2.1. ヘルプが必要ですか?	5
2.2. ご意見をお寄せください	6
<b>第1章 はじめに</b> .....	<b>7</b>
1.1. 本ガイドの目的	7
1.2. SYSTEMTAP の機能	7
<b>第2章 SYSTEMTAP の使用</b> .....	<b>8</b>
2.1. インストールと設定	8
2.1.1. SystemTap のインストール	8
2.1.2. 必要なカーネル情報 RPM パッケージのインストール	8
2.1.3. 初期テスト	9
2.2. 他のコンピューター用のインストールメンテーション生成	10
2.3. SYSTEMTAP スクリプトの実行	12
2.3.1. SystemTap フライトレコーダーモード	13
2.3.1.1. メモリー内フライトレコーダー	14
2.3.1.2. ファイルフライトレコーダー	14
<b>第3章 SYSTEMTAP の作動方法</b> .....	<b>16</b>
3.1. アーキテクチャー	16
3.2. SYSTEMTAP スクリプト	16
3.2.1. イベント	18
3.2.2. Systemtap ハンドラー/ボディー	20
3.3. 基本的な SYSTEMTAP ハンドラーコンストラクト	23
3.3.1. 変数	23
3.3.2. 条件付き (conditional) ステートメント	24
3.3.3. コマンドラインの引数	26
3.4. 連想アレイ	26
3.5. SYSTEMTAP でのアレイ演算	27
3.5.1. 関連する値の割り当て	27
3.5.2. アレイからの値の読み取り	27
3.5.3. 関連する値の増加	28
3.5.4. アレイ内での複数要素の処理	28
3.5.5. アレイおよびアレイ要素の消去/削除	29
3.5.6. 条件付きステートメントにおけるアレイの使用	30
3.5.7. 統計集計の計算	31
3.6. TAPSETS	33
<b>第4章 便利な SYSTEMTAP スクリプト</b> .....	<b>34</b>
4.1. ネットワーク	34
4.1.1. ネットワークのプロファイリング	34
4.1.2. ネットワークソケットコードで呼び出された関数の追跡	36
4.1.3. 着信 TCP 接続の監視	37
4.1.4. カーネルでのネットワークパケットドロップの監視	38
4.2. ディスク	39
4.2.1. ディスク読み取り/書き込みトラフィックの要約	39

---

4.2.2. ファイル読み取り/書き込みの I/O 時間の追跡	41
4.2.3. 累積 I/O の追跡	44
4.2.4. I/O 監視 (デバイスごと)	45
4.2.5. ファイルの読み取りおよび書き込みの監視	46
4.2.6. ファイル属性の変更の監視	47
4.3. プロファイリング	48
4.3.1. 関数呼び出しのカウンタ	48
4.3.2. 呼び出し先の追跡	49
4.3.3. カーネルおよびユーザースペースで費やした時間の判定	51
4.3.4. ポーリングアプリケーションの監視	52
4.3.5. 最もよく使われるシステムコールの追跡	55
4.3.6. プロセスごとのシステムコールボリュームの追跡	56
4.4. 競合ユーザースペースのロックの特定	58
<b>第5章 SYSTEMTAP のエラーを理解する</b> .....	<b>61</b>
5.1. 解析エラーとセマンティックエラー	61
5.2. ランタイムエラーおよび警告	62
<b>第6章 リファレンス</b> .....	<b>64</b>
<b>付録A 改訂履歴</b> .....	<b>65</b>
<b>索引</b> .....	<b>66</b>

# 前書き

## 1. ドキュメント規則

本ガイドでは、いくつかの規則を使用して特定の単語やフレーズを強調表示し、特定の情報への注意を促しています。

### 1.1. 誤字規則

特定の単語や句に注意を促すために4つの誤字規則を使用しています。これらの規則や、これらが適用される状況は以下のとおりです。

#### Mono-spaced Bold

シェルコマンド、ファイル名、パスなど、システム入力を強調表示するために使用されます。キーとキーの組み合わせを強調表示するためにも使用されます。以下に例を示します。

現在の作業ディレクトリーのファイル **my\_next\_bestselling\_novel** の内容を表示するには、シェルプロンプトで **cat my\_next\_bestselling\_novel** コマンドを入力し、**Enter** を押してコマンドを実行します。

上記には、ファイル名、シェルコマンドおよびキーが含まれます。これはすべて mono-spaced bold で示され、コンテキストにより区別可能なものになります。

キーの組み合わせは、キーの組み合わせの各パーツをつなげるプラス記号によって個別のキーと区別できます。以下に例を示します。

**Enter** を押してコマンドを実行します。

**Ctrl+Alt+F2** を押して、仮想端末に切り替えます。

最初の例では、押す特定のキーを強調表示しています。2つ目の例は、同時に押す3つのキーのセットというキーの組み合わせを強調表示しています。

ソースコードについて記述では、クラス名、メソッド、関数、変数名、および段落内で記述された戻り値は、**mono-spaced bold**で示されます。以下に例を示します。

ファイル関連クラスには、ファイルシステムの **filesystem**、ファイルの **file**、ディレクトリーの **dir** が含まれます。各クラスには、独自の関連付けられたパーミッションセットがあります。

#### Proportional Bold

これは、アプリケーション名、ダイアログボックステキスト、ラベルが付いたボタン、チェックボックスおよびラジオボタン、メニュータイトルおよびサブメニュータイトルなど、システムで発生した単語またはフレーズを示します。以下に例を示します。

メインメニューバーから **System** → **Preferences** → **Mouse** を選択して、**Mouse Preferences** を起動します。**Buttons** タブで、**Left-handed mouse** のチェックボックスを選択し、**Close** をクリックして、左から右に主要なマウスボタンを切り替えます (左側のマウスは適切なマウスになります)。

特殊文字を `gedit` ファイルに挿入するには、メインメニューバーから **Applications** → **Accessories** → **Character Map** を選択します。次に、**Character Map** メニューバーから **Search** → **Find...** を選択し、**Search** フィールドに文字の名前を入力して **Next** をクリックします。目的の文字は **Character Table** に強調表示されます。この強調表示した文字をダブルクリックして、**Text to copy** フィールドに配置し、**Copy** ボタンをクリックします。次に、ドキュメントに戻り、`gedit` メニューバーから **Edit** → **Paste** を選択します。

上記のテキストにはアプリケーション名、システム全体のメニュー名および項目、アプリケーション固有のメニュー名、GUI インターフェース内のボタンおよびテキストなどがあります。すべては proportional bold で示され、コンテキストと区別できます。

### ***Mono-spaced Bold Italic*** または ***Proportional Bold Italic***

mono-spaced bold または proportional bold のいずれでも、イタリックを追加すると、置換または変数テキストが表示されます。イタリックは、状況に応じて変化するテキストや、文字を入力しないテキストを表します。以下に例を示します。

`ssh` を使用してリモートマシンに接続するには、シェルプロンプトで **ssh** **username@domain.name** を入力します。リモートマシンが **example.com** で、そのマシンのユーザー名が **john** の場合は、**ssh john@example.com** と入力します。

**mount -o remount file-system** コマンドは、名前付きのファイルシステムを再マウントします。たとえば、**/home** ファイルシステムを再マウントする場合、コマンドは **mount -o remount /home** になります。

現在インストールされているパッケージのバージョンを表示するには、**rpm -q package** コマンドを使用します。結果は以下のようになります: **package-version-release**。

上記の太字のイタリック体の用語、`username`、`domain.name`、`file-system`、`package`、`version`、および `release` に注意してください。各単語はプレースホルダーで、コマンドの発行時に入力するテキストまたはシステムによって表示されるテキストのどちらかになります。

作業のタイトルを示す標準的な使用法のほかに、イタリックは新用語と重要な用語の最初の使用を示します。以下に例を示します。

Publican は *DocBook* 公開システムです。

## 1.2. 引用規則

端末の出力およびソースコードの一覧は、周りのテキストから視覚的に表示されます。

端末に送信される出力は **mono-spaced roman** にセットされて表現されます。

```
books      Desktop documentation drafts mss photos stuff svn
books_tests Desktop1 downloads  images notes scripts svgs
```

ソースコードの一覧も **mono-spaced roman** にセットされますが、構文の強調表示が以下のように追加されます。

```
static int kvm_vm_ioctl_deassign_device(struct kvm *kvm,
                                       struct kvm_assigned_pci_dev *assigned_dev)
{
    int r = 0;
    struct kvm_assigned_dev_kernel *match;
```



```

mutex_lock(&kvm->lock);

match = kvm_find_assigned_dev(&kvm->arch.assigned_dev_head,
                             assigned_dev->assigned_dev_id);
if (!match) {
    printk(KERN_INFO "%s: device hasn't been assigned before, "
           "so cannot be deassigned\n", __func__);
    r = -EINVAL;
    goto out;
}

kvm_deassign_device(kvm, match);

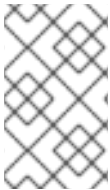
kvm_free_assigned_device(kvm, match);

out:
mutex_unlock(&kvm->lock);
return r;
}

```

### 1.3. 注記および警告

最後に、3つの視覚的スタイルを使用して、見落とす可能性のある情報に注意を促します。



#### 注記

注記とは、タスクへのヒント、ショートカット、または代替アプローチです。注意を無視しても悪い結果を招くことはありませんが、便利なヒントを見逃してしまう可能性があります。



#### 重要

見落としやすい詳細のある重要なボックス: 現行セッションにのみ適用される設定変更や、更新を適用する前に再起動が必要なサービスなどです。「Important」というラベルが付いたボックスを無視しても、データが失われることはありませんが、スムーズな操作が行えないことがあります。



#### 警告

警告は無視すべきではありません。警告を無視すると、データが失われる可能性があります。

## 2. ヘルプの利用とフィードバック提供

### 2.1. ヘルプが必要ですか?

本ガイドで説明されている手順で問題が発生した場合は、Red Hat カスタマーポータル <http://access.redhat.com> にアクセスしてください。カスタマーポータルでは、以下を行うことができます。

- Red Hat 製品に関する技術サポート記事の知識ベースの検索または閲覧。
- Red Hat グローバルサポートサービス (GSS) へのサポートケースの送信。
- その他の製品ドキュメントへのアクセス。

Red Hat は、Red Hat のソフトウェアおよびテクノロジーについて、多くの電子メーリングリストも提供しています。一般に公開されているメーリングリストの一覧は、<https://www.redhat.com/mailman/listinfo>を参照してください。メーリングリストの名前をクリックして、その一覧をサブスクライブするか、またはメーリングリストのアーカイブにアクセスします。

## 2.2. ご意見をお寄せください

本ガイドで誤字脱字を発見されたり、このマニュアルを改善するための提案をお持ちの場合は、弊社までご連絡ください。Red\_Hat\_Enterprise\_Linux の製品に対して、<http://bugzilla.redhat.com/> から Bugzilla レポートを送信してください。

バグレポートを送信する際には、マニュアル識別子 (『doc-SystemTap\_Beginners\_Guide』) を指定してください。

本ガイドを改善するためのご意見やご提案をお寄せいただく場合は、できるだけ具体的にご説明ください。エラーが見つかった場合は、簡単に確認できるように、セクション番号と前後のテキストを含めてください。

## 第1章 はじめに

SystemTap は、オペレーティングシステム (特にカーネル) の動作を詳細に観察および監視できる追跡およびプロービングのツールです。**netstat**、**ps**、**top**、**iostat**などのツールの出力に似た情報を提供します。ただし SystemTap は、収集した情報に対して、より多くのフィルタリングと分析オプションを提供するように設計されています。

### 1.1. 本ガイドの目的

SystemTap は、分析を詳細に行うために、稼働中の Linux システムを監視するインフラストラクチャーを提供します。これは、管理者や開発者がバグやパフォーマンス問題の根本的な原因を特定する手助けとなります。

SystemTap を使用しないで、実行中のカーネルのアクティビティーを監視しようとする、インストールメント化、再コンパイル、インストール、および再起動という面倒なプロセスが必要になります。SystemTap を使用するとこのプロセスが不要になり、ユーザーが作成する SystemTap スクリプトを実行するだけで同様の情報を収集できるようになります。

ただし、SystemTap は当初、Linux カーネルに関する中級から上級の知識を備えたユーザーのために設計されました。このため、SystemTap は Linux カーネルに関する知識と経験があまりない管理者や開発者にはそれほど有用なものではありません。さらに、既存の SystemTap ドキュメントのほとんどは、同様に知識と経験が豊富なユーザーを対象としています。これにより、ツールの学習も困難になります。

そこで、『SystemTap ビギナーズガイド』では初級者を対象に、以下を説明します。

- SystemTap の概要を紹介するほか、ユーザーにアーキテクチャーについて説明し、全カーネルタイプの設定手順を提供します。
- システムの異なるコンポーネントで詳細なアクティビティーを監視するための事前作成された SystemTap スクリプトを提供し、それらの実行方法と出力の分析方法も提供します。

### 1.2. SYSTEMTAP の機能

- 柔軟性 - SystemTap のフレームワークを使用すると、幅広いカーネル機能、システムコール、およびカーネルスペースで発生する他のイベントについて、調査および監視目的のシンプルなスクリプトを開発できます。つまり、SystemTap は **ツール** というよりも、独自のカーネル固有のフォレンジックおよび監視ツールの開発を可能にするシステムといえます。
- 容易な使用 - 上記で説明したように、SystemTap を使うことでユーザーはカーネルプロセスのインストールメント化、再コンパイル、インストール、および再起動という面倒なプロセスを経ずにカーネルスペースのイベントをプローブできるようになります。

「[4章 便利な SystemTap スクリプト](#)」で挙げられている SystemTap スクリプトのほとんどは、他の同様のツール (**top**、**oprofile**、または **ps** など) ではネイティブで利用可能ではないシステムフォレンジックや監視機能の例になります。これらのスクリプトは、SystemTap のアプリケーションの豊富な例を示すために提供されており、これによりユーザーは、独自の SystemTap スクリプト作成時に使用可能な機能について学ぶことができます。

## 第2章 SYSTEMTAP の使用

本章では、SystemTap のインストール方法と SystemTap スクリプトの実行方法を説明します。

### 2.1. インストールと設定

SystemTap のデプロイにインストールが必要となるのは、SystemTap パッケージと対応するカーネルの **-devel**、**-debuginfo**、および **-debuginfo-common-arch** パッケージです。システムに複数のカーネルがインストールされていて、それらのカーネル上で SystemTap を使用するには、それらの各カーネルバージョン用に **-devel** と **-debuginfo** パッケージをインストールします。

これらの手順は、以下のセクションで詳細に説明します。



#### 重要

多くのユーザーは、**-debuginfo** と **-debug** を混同します。SystemTap のデプロイに必要なとなるのは、カーネルの **-debuginfo** パッケージのインストールであって、カーネルの **-debug** バージョンではない点に注意してください。

#### 2.1.1. SystemTap のインストール

SystemTap をデプロイするには、以下の RPM パッケージをインストールします。

- **systemtap**
- **systemtap-runtime**

システムに **yum** がインストールされていることを前提として、これら 2 つの RPM パッケージは **yum install systemtap systemtap-runtime** を使用してインストールすることができます。SystemTap を使用する前に、必要なカーネル情報パッケージをインストールします。

#### 2.1.2. 必要なカーネル情報 RPM パッケージのインストール

SystemTap は、カーネル内にインストールメンターションを配置する (プローブする) ためにカーネルの情報が必要になります。この情報により、SystemTap はインストールメンターションのコード生成が可能になります。この情報は、一致するカーネル用の **-devel**、**-debuginfo**、および **-debuginfo-common-arch** パッケージに含まれています。通常の「vanilla」カーネルに必要な **-devel** および **-debuginfo** パッケージは、以下のとおりです。

- **kernel-debuginfo**
- **kernel-debuginfo-common-arch**
- **kernel-devel**

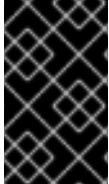
同様に、PAE カーネルに必要なパッケージは、**kernel-PAE-debuginfo**、**kernel-PAE-debuginfo-common-arch**、および **kernel-PAE-devel** になります。

システムで現在使用中のカーネルを確認するには、以下のコマンドを実行します。

```
uname -r
```

たとえば、32 ビット AMD および Intel アーキテクチャーのマシン上のカーネルバージョン **2.6.32-53.el6** で SystemTap を使用する場合は、以下の RPM パッケージをダウンロードおよびインストールする必要があります。

- **kernel-debuginfo-2.6.32-53.el6.i686.rpm**
- **kernel-debuginfo-common-i686-2.6.32-53.el6.i686.rpm**
- **kernel-devel-2.6.32-53.el6.i686.rpm**



### 重要

**-devel**、**-debuginfo**、および **-debuginfo-common-arch** パッケージのバージョン、バリエーション、アーキテクチャーは、SystemTap によって **正確に** プローブされるためにカーネルと一致する必要があります。

システムで SystemTap が必要とするチャンネルの一覧を取得するには、以下のスクリプトを使用します。

```
#!/bin/bash
pkg=`rpm -q --whatprovides "redhat-release"`
releasever=`rpm -q --qf "%{version}" $pkg`
variant=`echo $releasever | tr -d "[:digit:]" | tr "[:upper:]" "[:lower:]"`
if test -z "$variant"; then
  echo "No Red Hat Enterprise Linux variant (workstation/client/server) found."
  exit 1
fi
version=`echo $releasever | tr -cd "[:digit:]"`
base=`uname -i`
echo "rhel-$base-$variant-$version"
echo "rhel-$base-$variant-$version-debuginfo"
echo "rhel-$base-$variant-optional-$version-debuginfo"
echo "rhel-$base-$variant-optional-$version"
```

チャンネルを追加したら、コマンド **debuginfo-install kernelname-version** を使用して、カーネルに必要な **-devel**、**debuginfo**、および **debuginfo-install arch** パッケージをインストールします。**kernelname** を適切なカーネルバリエーション名 (**kernel-PAE** など) に置き換え、**version** をターゲットカーネルのバージョンに置き換えます。たとえば、**kernel-PAE-2.6.32-53.el6** カーネルに必要なカーネル情報パッケージをインストールするには、**debuginfo-install kernel-PAE-2.6.32-53.el6** を実行します。

### 2.1.3. 初期テスト

SystemTap でプローブするカーネルが現在使用中であれば、デプロイメントが成功したかどうかを直ちにテストできます。別のカーネルをプローブする場合は、再起動して該当カーネルを読み込みます。

テストを開始するには、**stap -v -e 'probe vfs.read {printf("read performed\n"); exit();}'** のコマンドを実行します。このコマンドは単に、SystemTap に **read performed** をプリントして、仮想ファイルシステムの読み込みが検出されると、正常に終了するよう指示します。SystemTap が正常にデプロイされていれば、以下のような出力になります。

```
Pass 1: parsed user script and 45 library script(s) in 340usr/0sys/358real ms.
Pass 2: analyzed script: 1 probe(s), 1 function(s), 0 embed(s), 0 global(s) in 290usr/260sys/568real ms.
```

```

Pass 3: translated to C into "/tmp/stapiArgLX/stap_e5886fa50499994e6a87aacdc43cd392_399.c" in
490usr/430sys/938real ms.
Pass 4: compiled C into "stap_e5886fa50499994e6a87aacdc43cd392_399.ko" in
3310usr/430sys/3714real ms.
Pass 5: starting run.
read performed
Pass 5: run completed in 10usr/40sys/73real ms.

```

(**Pass 5** で始まる) 出力の最後の 3 行は、SystemTap がカーネルをプローブするインストルメンテーションを正常に作成できたこと、インストルメンテーションを実行したこと、プローブしているイベントを検出したこと (このケースでは、仮想ファイルシステムの読み込み)、および有効なハンドラーを実行したこと (テキストをプリントし、エラーなしで終了) を示しています。

## 2.2. 他のコンピューター用のインストルメンテーション生成

ユーザーが SystemTap スクリプトを実行すると、そのスクリプトからカーネルモジュールが構築されます。すると SystemTap はそのモジュールをカーネルに読み込み、カーネルから直接指定されたデータを抽出できるようにします (詳細は、「[アーキテクチャー](#)」の [手順3.1「SystemTap セッション](#)」を参照)。

通常、SystemTap スクリプトは (「[インストールと設定](#)」にあるように) SystemTap がデプロイされているシステムでのみ実行できます。つまり、SystemTap を 10 台のシステムで実行するには、これら **すべての** システムに SystemTap をデプロイする必要があります。場合によっては、これは実現不可能もしくは望ましくないこともあります。たとえば、企業のポリシーで管理者が特定マシンにコンパイラーやデバッグ情報を提供する RPM のインストールを禁止されていれば、SystemTap のデプロイはできなくなります。

この状況を避けるためには、[クロスインストルメンテーション](#)を使用します。これは、1 台のコンピューター上の SystemTap スクリプトから別のコンピューターで使用する SystemTap インストルメンテーションモジュールを生成するプロセスです。このプロセスは、以下の利点をもたらします。

- 各種マシンのカーネル情報パッケージを単一の **ホストマシン** にインストールできます。
- 生成された SystemTap インストルメンテーションモジュールを使用するために各 **ターゲットマシン** にインストールする必要があるのは 1 つの RPM **systemtap-runtime** のみです。

### 注記

本セクションでは分かりやすくするために、以下の用語を使用します。

- **instrumentation module** – SystemTap スクリプトから構築されるカーネルモジュールです。**SystemTap モジュール** は **ホストシステム** 上に構築され、**ターゲットシステム** の **ターゲットカーネル** に読み込まれます。
- **ホストシステム** – このシステム上で (SystemTap スクリプトから) インストルメンテーションモジュールがコンパイルされ、**ターゲットシステム** に読み込まれます。
- **ターゲットシステム** – このシステム内で (SystemTap スクリプトから) **インストルメンテーションモジュール** が構築されます。
- **ターゲットカーネル** – **ターゲットシステム** のカーネルです。このカーネルが **インストルメンテーションモジュール** の読み込み、実行を行います。

### 手順2.1 ホストシステムとターゲットシステムの設定

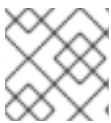
1. 各 ターゲットシステム に **systemtap-runtime** RPM をインストールします。
2. 各 ターゲットシステム で **uname -r** を実行して、各 ターゲットシステム で実行中のカーネルを確認します。
3. SystemTap を ホストシステム にインストールします。インストールメンテーションモジュールは、ホストシステム 上で ターゲットシステム 用に構築されます。SystemTap のインストール方法は、「[SystemTap のインストール](#)」を参照してください。
4. 上記で判明した ターゲットカーネル のバージョンを使用して、ターゲットカーネル と関連 RPM を「[必要なカーネル情報 RPM パッケージのインストール](#)」に説明されている方法でホストシステム にインストールします。複数の ターゲットシステム で異なる ターゲットカーネル を使用している場合は、ターゲットシステム で使用しているカーネルごとにこのステップを繰り返します。

手順2.1「[ホストシステムとターゲットシステムの設定](#)」が完了すると、(いずれの ターゲットシステム用の) インストールメンテーションモジュール も ホストシステム での構築が可能になります。

インストールメンテーションモジュール を構築するには、ホストシステム で以下のコマンドを実行します (適切な値を指定してください)。

```
stap -r kernel_version script -m module_name -p4
```

ここでの **kernel\_version** は、ターゲットカーネル のバージョンを指します (ターゲットマシンで実行した **uname -r** の出力)。**script** は、インストールメンテーションモジュール に変換されるスクリプトです。**module\_name** は、希望する インストールメンテーションモジュール の名前です。



#### 注記

実行中のカーネルのアーキテクチャー表記を確認するには、**uname -m** を実行します。

インストールメンテーションモジュールがコンパイルされたら、ターゲットシステム にコピーして、以下のコマンドを使用して読み込みます。

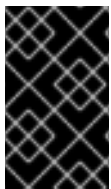
```
staprun module_name.ko
```

たとえば、ターゲットカーネル 2.6.32-53.el6 用に **simple.stp** という名前が付けられた SystemTap スクリプトから インストールメンテーションモジュール **simple.ko** を作成するには、以下のコマンドを使用します。

```
stap -r 2.6.32-53.el6 -e 'probe vfs.read {exit()}' -m simple -p4
```

これで **simple.ko** という名前のモジュールが作成されます。インストールメンテーションモジュール **simple.ko** を使用するには、これを ターゲットシステム にコピーして、(ターゲットシステム 上で) 以下のコマンドを実行します。

```
staprun simple.ko
```



#### 重要

構築された インストールメンテーションモジュール が機能するには、ホストシステム と ターゲットシステム が同一アーキテクチャーで同じ Linux ディストリビューションを実行している必要があります。



## 2.3. SYSTEMTAP スクリプトの実行

SystemTap スクリプトは **stap** コマンドで実行されます。**stap** を使うと、SystemTap スクリプトを標準入力またはファイルから実行できます。

**stap** および **staprun** を実行するには、システムに対する権限の昇格が必要になります。ただし、すべてのユーザーが SystemTap 実行のためだけに root アクセスを付与されるわけではありません。たとえば、権限のないユーザーが自身のマシン上で SystemTap インストールメンターションを実行する必要がある場合もあります。

通常のユーザーが root アクセスなしで SystemTap を実行できるようにするには、ユーザーを以下の両方のユーザーグループに追加します。

### stapdev

このグループのメンバーは **stap** を使用して SystemTap スクリプトを実行したり、**staprun** を使用して SystemTap インストールメンターションモジュールを実行することができます。

**stap** の実行では、SystemTap スクリプトがカーネルモジュールにコンパイルされ、それがカーネルに読み込まれます。これにはシステムに対する権限の昇格が必要となり、**stapdev** メンバーにはそれが付与されます。ただし、この権限は **stapdev** メンバーに有効な root アクセスも付与することになります。このため、**stapdev** グループのメンバーシップは、root アクセスを信頼して付与できるメンバーにのみ許可してください。

### stapusr

このグループのメンバーが SystemTap インストールメンターションモジュールの実行に使用できるのは、**staprun** のみです。さらに、`/lib/modules/kernel_version/systemtap/` から実行できるのは、これらのモジュールだけになります。このディレクトリーを所有できるのは root ユーザーのみで、書き込みが可能なのも root ユーザーのみでなければならない点に注意してください。



### 注記

SystemTap スクリプトを実行するには、ユーザーは **stapdev** と **stapusr** の両方のグループに属する必要があります。

**stap** で使用する一般的なオプションには、以下のものがあります。

#### -v

SystemTap セッションの出力をさらに詳細なものにします。このオプション (たとえば、**stap -vvv script.stp**) は、スクリプトの実行に関するさらなる詳細を提供するために、繰り返すことができます。スクリプトの実行中にエラーが発生すると、これは特に便利なものになります。このオプションは、スクリプトの実行中にエラーが発生した場合に特に便利です。

SystemTap スクリプトの一般的なエラーに関する詳細は、「[5章 SystemTap のエラーを理解する](#)」を参照してください。

#### -o filename

標準出力をファイル (*filename*) に送信します。

#### -S size、count

ファイルのサイズを *size* で指定されたメガバイト数に制限し、ファイル数を *count* の数に制限します。ファイル名には、連続番号の接尾辞が付きます。このオプションは、SystemTap に logrotate 演算を実装します。



**-o** と使用すると、**-S** はログファイルのサイズを制限します。

#### **-x process ID**

SystemTap ハンドラー関数の **target()** を指定されたプロセス ID に設定します。**target()** の詳細については、「[SystemTap 関数](#)」を参照してください。

#### **-c command**

SystemTap ハンドラー関数の **target()** を指定されたコマンドに設定します。指定コマンドへの完全パスを使用する必要があります。たとえば、**cp** を指定する代わりに、(**stap script -c /bin/cp** のように) **/bin/cp** を使用します。**target()** の詳細については、「[SystemTap 関数](#)」を参照してください。

#### **-e 'script'**

systemtap 翻訳の入力に、ファイルではなく **script** 文字列を使用します。

#### **-F**

SystemTap のフライトレコーダーモードを使用し、スクリプトをバックグラウンドプロセスにします。フライトレコーダーモードの詳細は、「[SystemTap フライトレコーダーモード](#)」を参照してください。

スイッチ **-** を使用して、標準入力からスクリプトを実行するように **stap** を指示することも可能です。例を示します。

#### 例2.1 標準入力からスクリプトを実行

```
echo "probe timer.s(1) {exit()}" | stap -
```

例2.1「標準入力からスクリプトを実行」では、**echo** が標準入力に渡したスクリプトを実行するように **stap** に指示します。**stap** で使用するオプションはすべて、**-** スイッチの前に挿入する必要があります。たとえば、例2.1「標準入力からスクリプトを実行」をより詳細な出力にするには、以下のようなコマンドになります。

```
echo "probe timer.s(1) {exit()}" | stap -v -
```

**stap** の詳細については、**man stap** を参照してください。

SystemTap インストールメンテーション (クロスインストールメンテーション中に SystemTap スクリプトから構築されたカーネルモジュール) を実行するには、**staprun** を代わりに使用します。**staprun** およびクロスインストールメンテーションの詳細は、「[他のコンピューター用のインストールメンテーション生成](#)」を参照してください。



#### 注記

**stap** の **-v** オプションと **-o** オプションは **staprun** でも機能します。**staprun** の詳細については、**man staprun** を参照してください。

### 2.3.1. SystemTap フライトレコーダーモード

SystemTap のフライトレコーダーモードを使用すると、SystemTap スクリプトは長期間の実行が可能になり、最近の出力のみにフォーカスできるようになります。フライトレコーダーモード (**-F** オプション) は、生成される出力の量を制限します。フライトレコーダーモードには、メモリー内モードとファ

イルモードの2種類があります。どちらの場合も、SystemTap スクリプトはバックグラウンドプロセスとして実行されます。

### 2.3.1.1. メモリー内フライトレコーダー

フライトレコーダーモード (**-F** オプション) をファイル名なしで使用すると、SystemTap はカーネルメモリー内のバッファを使用してスクリプトの出力を保存します。次に、SystemTap インストールメンテーションモジュールが読み込まれてプローブの実行が開始し、インストールメンテーションが外されてバックグラウンドに置かれます。関心のあるイベントが発生すると、インストールメンテーションは再度アタッチされ、メモリーバッファ内の最近の出力と継続中の出力が閲覧可能となります。以下のコマンドでは、フライトレコーダーのメモリー内モードを使用してスクリプトが開始されます。

```
stap -F /usr/share/doc/systemtap-version/examples/io/iotime.stp
```

スクリプトが開始されると、実行中のスクリプトに再接続するためのコマンドを示すメッセージが表示されます。

```
Disconnecting from systemtap module.  
To reconnect, type "staprun -A stap_5dd0073edcb1f13f7565d8c343063e68_19556"
```

関心のあるイベントが発生したら、実行中のスクリプトに再度アタッチしてメモリーバッファ内の最近のデータを出力し、継続中の出力を得るために、以下のコマンドを使用します。

```
staprun -A stap_5dd0073edcb1f13f7565d8c343063e68_19556
```

カーネルバッファはデフォルトでは 1MB のサイズですが、**-s** オプションを使用してバッファのメガバイト単位のサイズを指定する (2 の累乗に切り上げる) ことが可能です。たとえば、SystemTap コマンドライン上で **-s2** とすると、バッファを 2MB に指定します。

### 2.3.1.2. ファイルフライトレコーダー

フライトレコーダーモードは、ファイルにデータを保存することもできます。保持するファイルの数とサイズは、**-S** オプションにコンマ区切りの2つの数字の引数を続けて制御します。最初の引数は、各出力ファイルのメガバイト単位の最大サイズです。2つ目の引数は、保存する最新ファイルの数です。ファイル名は、**-o** オプションの後に名前を付けて指定します。SystemTap は数値の接尾辞をファイル名に付けてファイルの順序を示します。以下のコマンドでは、SystemTap がフライトレコーダーモードで開始され、出力は **/tmp/pfaults.log.[0-9]+** という名前のファイルに保存されます。各ファイルのサイズは 1MB 以下となり、最新の2ファイルが保存されます。

```
stap -F -o /tmp/pfaults.log -S 1,2 pfaulst.stp
```

コマンドがプリントする数字は、プロセス ID です。プロセスに SIGTERM を送信すると、SystemTap スクリプトはシャットダウンし、データ収集が停止されます。たとえば、以前のコマンドでプロセス ID が 7590 となっている場合、以下のコマンドを実行すると SystemTap スクリプトはシャットダウンします。

```
kill -s SIGTERM 7590
```

スクリプトが生成した最新の2ファイルのみが保存され、それより古いファイルは削除されます。このため、**ls -sh /tmp/pfaults.log.\*** を実行して表示されるのは、2ファイルのみです。

```
1020K /tmp/pfaults.log.5 44K /tmp/pfaults.log.6
```

---

最新のデータは数字の大きい方のファイルで、上記の場合は /tmp/pfaults.log.6 になります。

## 第3章 SYSTEMTAP の作動方法

SystemTap を使うとユーザーはシンプルなスクリプトを作成および再使用して、実行中の Linux システムのアクティビティーを徹底的に調べることができます。このスクリプトは、データを抽出し、フィルターをかけ、素早く安全にデータをまとめるように設計できます。これにより、複雑なパフォーマンス (または機能的な) 問題の診断が可能になります。

SystemTap スクリプトにおける本質的な考えは、イベントに名前を付け、それにハンドラーを与えることです。SystemTap がスクリプトを実行すると、SystemTap はイベントを監視します。イベントが発生したら、Linux カーネルがハンドラーをサブルーチンとして実行し、その後通常動作を再開します。

イベントには関数の開始/終了、タイマーの有効期限切れ、セッション終了など数種類のものがあります。ハンドラーは一連のスクリプト言語のステートメントで、イベント発生時に実行する作業を指定します。この作業には通常、イベントコンテキストからのデータ抽出、それらの内部変数への保存、結果のプリントなどが含まれます。

### 3.1. アーキテクチャー

SystemTap のセッションは、SystemTap スクリプトを実行すると始まります。このセッションは、以下の順番で生じます。

#### 手順3.1 SystemTap セッション

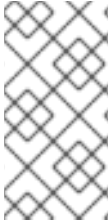
1. SystemTap はまず、使用されている tapset について既存の tapset ライブラリー (通常は `/usr/share/systemtap/tapset/`) に対してスクリプトをチェックします。次に SystemTap は、見つかった tapset を tapset ライブラリー内の対応する定義で置き換えます。
2. SystemTap はスクリプトを C に変換し、システム C コンパイラーを実行してそこからカーネルモジュールを作成します。このステップを実行するツールは、**systemtap** パッケージに含まれています (詳細は「[SystemTap のインストール](#)」を参照してください)。
3. SystemTap はモジュールを読み込み、スクリプト内の全プローブ (イベントおよびハンドラー) を有効にします。**systemtap-runtime** パッケージ (詳細は「[SystemTap のインストール](#)」を参照) 内の **staprun** は、この機能を提供します。
4. イベントが発生すると、それに対応するハンドラーが実行されます。
5. SystemTap セッションが終了すると、プローブは無効になり、カーネルモジュールは読み込み解除されます。

このステップは、コマンドラインプログラム **stap** で実行されます。このプログラムは、SystemTap のメインのフロントエンドツールです。**stap** の詳細は、(SystemTap がマシンに適切にインストールされてから) **man stap** を参照してください。

### 3.2. SYSTEMTAP スクリプト

SystemTap スクリプトはそのほとんどにおいて、各 SystemTap セッションのベースになっています。SystemTap スクリプトが SystemTap に対してどのタイプの情報を収集するか、収集後に何をするかを指示します。

「[3章 SystemTap の作動方法](#)」の説明にあるように、SystemTap スクリプトは **イベント** と **ハンドラー** という 2 つのコンポーネントで構成されています。SystemTap セッションが開始されたら、SystemTap はオペレーティングシステムで指定されたイベントを監視し、イベントが発生したらハンドラーを実行します。



## 注記

イベントとそれに対応するハンドラーは、まとめて **プローブ** と呼ばれます。SystemTap スクリプトには複数のプローブを備えることができます。

プローブのハンドラーは一般的に **プローブボディー** と呼ばれます。

アプリケーションの開発という面では、イベントとハンドラーの使用は診断プリントステートメントをコマンドのプログラムシーケンスに挿入するというコードのインストルメント化に似ています。診断プリントステートメントを使用すると、プログラムの実行後に発行されたコマンドの履歴をみることができます。

SystemTap スクリプトでは、コードを再コンパイルすることなくインストルメンテーションコードの挿入が可能で、ハンドラーに関する柔軟性が広がります。イベントは、ハンドラー実行の引き金となります。ハンドラーは指定されたデータを記録して、特定の 방법으로プリントするよう指定できます。

## 形式

SystemTap スクリプトは、ファイル拡張子 **.stp** を使用し、以下の書式のプローブを含んでいます。

```
probe event {statements}
```

SystemTap は1つのプローブにつき複数のイベントをサポートしており、複数イベントはコンマ(,)で区切ります。1つのプローブで複数のイベントが指定された場合、SystemTap は指定されたイベントが発生するとそのハンドラーを実行します。

プローブにはそれぞれ、対応する **ステートメントブロック** があります。このステートメントブロックは中括弧({})で囲まれており、イベントごとに実行されるステートメントが含まれています。SystemTap はこれらのステートメントを順番に実行し、通常は複数のステートメントを分ける特別なセパレーターやターミネーターは必要ありません。



## 注記

SystemTap スクリプト内のステートメントブロックは、C プログラミング言語と同じ構文とセマンティクスを使用します。ステートメントブロックは、別のステートメントブロック内の入れ子状態にすることができます。

Systemtap では、多くのプローブが使用するコードを外に括り出して関数を作成することができます。つまり、複数のプローブで同じステートメントを何度も繰り返し書くのではなく、以下のように *function* 内に指示を配置することができます。

```
function function_name(arguments) {statements}
probe event {function_name(arguments)}
```

**statements** in *function\_name* は、*event* のプローブの実行時に実行されます。*arguments* は、*function* に渡されるオプションの値です。



## 重要

「[SystemTap スクリプト](#)」では、SystemTap スクリプトの基本を説明します。SystemTap スクリプトの理解を深めるには、「[4章 便利な SystemTap スクリプト](#)」を参照してください。この章の各セクションでは、スクリプト、イベント、ハンドラー、および予測される出力を詳細に説明しています。

### 3.2.1. イベント

SystemTap のイベントは大まかに、*同期*と *非同期*に分けられます。

#### 同期イベント

*同期*イベントは、プロセスがカーネルコード内の特定の場所で指示を実行する際に発生します。これは他のイベントの参照ポイントとなり、ここからさらにコンテキストデータが入手可能になります。

同期イベントの例には以下のようなものがあります。

#### `syscall.system_call`

システムコール `system_call` へのエントリ。システムコールの終了を希望する場合は、`.return` をイベントに追加すると、システムコールの終了を監視するようになります。たとえば、システムコール `close` のエントリと終了を指定するには、`syscall.close` および `syscall.close.return` をそれぞれ使用します。

#### `vfs.file_operation`

仮想ファイルシステム (VFS) の `file_operation` イベントへのエントリ。`syscall` イベントと同様に、イベントに `.return` を追加すると、`file_operation` 動作の終了を監視します。

#### `kernel.function("function")`

カーネル関数 `function` へのエントリ。たとえば、`kernel.function("sys_open")` は、カーネル関数 `sys_open` がシステム内のスレッドによって呼び出される際に発生する「イベント」を参照します。カーネル関数 `sys_open` の `return` を指定するには、`kernel.function("sys_open").return` のように、`return` 文字列をイベントステートメントに追加します。

プローブイベントを定義する際には、アスタリスク (\*) をワイルドカードに使用できます。また、カーネルソースファイル内の関数のエントリまたは終了も追跡可能です。以下の例を見てみましょう。

#### 例3.1 wildcards.stp

```
probe kernel.function("**@net/socket.c") {}
probe kernel.function("**@net/socket.c").return {}
```

この例では、最初のプローブのイベントは `net/socket.c` カーネルソースファイル内の全関数のエントリを指定しています。2つ目のプローブでは、これら全関数の終了を指定しています。この例では、ハンドラーにステートメントがないことに注意してください。このため、情報が収集されず、表示されることもありません。

#### `kernel.trace("tracepoint")`

`tracepoint` の静的プローブ。最近のカーネル (2.6.30 以降) には、カーネル内の特定イベント用のインストルメンテーションが含まれています。これらのイベントは、トレースポイントで静的にマークが付けられています。SystemTap で利用可能なトレースポイントの例としては、`kernel.trace("kfree_skb")` があります。これは、カーネル内でネットワークバッファが解放されると合図します。

#### `module("module").function("function")`

モジュール内の関数のプローブを可能にします。以下に例を示します。

#### 例3.2 moduleprobe.stp

```
probe module("ext3").function("") { }
probe module("ext3").function("").return { }
```

例3.2 「[moduleprobe.stp](#)」の最初のプローブは、**ext3** モジュールの全関数のエントリーを指しています。2つ目のプローブは、同じモジュールの全関数の終了を指しています。**.return** の接尾辞の使用は、**kernel.function()** と類似します。例3.2 「[moduleprobe.stp](#)」のプローブは、プローブハンドラーにステートメントがないことに注意してください。このため、有用なデータは表示されません(例3.1 「[wildcards.stp](#)」の場合と同様)。

システムのカーネルモジュールは通常 **/lib/modules/kernel\_version** にあります。ここでの *kernel\_version* は、現在読み込まれているカーネルのバージョンを指します。モジュールは、ファイル名拡張子 **.ko** を使用します。

## 非同期イベント

非同期イベントは、コード内の特定の指示や場所に関連付けられていません。このタイプのプローブイベントは、主にカウンターやタイマー、同様のコンストラクトで構成されています。

非同期イベントの例には以下のようなものがあります。

### begin

SystemTap スクリプトの実行と同時に、SystemTap セッションが起動します。

### end

SystemTap セッションが終了します。

### timer イベント

ハンドラーの定期実行を指定するイベントです。以下に例を示します。

#### 例3.3 timer-s.stp

```
probe timer.s(4)
{
  printf("hello world\n")
}
```

例3.3 「[timer-s.stp](#)」は、4秒ごとに **hello world** をプリントするプローブの例です。以下のような timer イベントも使用できます。

- **timer.ms(*milliseconds*)**
- **timer.us(*microseconds*)**
- **timer.ns(*nanoseconds*)**
- **timer.hz(*hertz*)**
- **timer.jiffies(*jiffies*)**

情報を収集する他のプローブと併せて timer イベントを使用すると、定期的な更新が表示でき、その情報の変遷が分かります。



## 重要

SystemTap では、多数のプロブイベントの使用をサポートしています。サポートされるイベントの詳細は、**man stapprobes** を参照してください。**man stapprobes** の『SEE ALSO』セクションには、特定のサブシステムやコンポーネントでサポートされているイベントについて説明している他の **man** ページへのリンクも含まれています。

### 3.2.2. Systemtap ハンドラー/ボディー

以下のサンプルスクリプトを見てみましょう。

#### 例3.4 helloworld.stp

```
probe begin
{
  printf ("hello world\n")
  exit ()
}
```

例3.4 「[helloworld.stp](#)」では、**begin** イベント (セッションの開始) が `{ }` で囲まれているハンドラーを始動させます。これは単に **hello world** をプリントして改行し、終了するものです。



## 注記

SystemTap スクリプトの実行は、**exit()** 関数が実行されるまで継続されます。スクリプトの実行を停止したい場合は、手動で **Ctrl+C** と入力すると中断できます。

### printf () ステートメント

**printf ()** ステートメントは、データをプリントする最も簡単な関数の1つです。**printf ()** を以下の書式で使用すると、多くの SystemTap 関数を使用するデータを表示できます。

```
printf ("format string\n", arguments)
```

*format string* では、*arguments* のプリント方法を指定します。例3.4 「[helloworld.stp](#)」の書式文字列は、SystemTap に **hello world** のプリントを指示するだけで、書式指定子は含みません。

引数によっては、**%s** (文字列用) や **%d** (数字用) といった書式指定子を書式文字列に使用することもできます。書式文字列には複数の書式指定子を使用することが可能で、それぞれを対応する引数に一致させます。複数の引数はコンマ (,) で区切ります。



## 注記

セマンティックの面では、SystemTap **printf** 関数は、C 言語の関数に非常によく似ています。上記の SystemTap の **printf** 関数における構文と書式は、C 言語の **printf** と同一のものであります。

以下のプロブの例を見てみましょう。

#### 例3.5 variables-in-printf-statements.stp



```
probe syscall.open
{
    printf ("%s(%d) open\n", execname(), pid())
}
```

例3.5 「[variables-in-printf-statements.stp](#)」では、SystemTap がシステムコール **open** への全エントリをプローブするように指示しています。各イベントでは、現行の **execname()** (実行可能ファイル名の付いた文字列) と **pid()** (現行のプロセス ID 番号) に続けて **open** という単語をプリントします。このプローブ出力の抜粋は以下のようになります。

```
vmware-guestd(2206) open
hald(2360) open
hald(2360) open
hald(2360) open
df(3433) open
df(3433) open
df(3433) open
hald(2360) open
```

### SystemTap 関数

SystemTap は、**printf ()** 引数として使用可能な多くの関数をサポートしています。例3.5 「[variables-in-printf-statements.stp](#)」では、SystemTap 関数 **execname()** (カーネル関数を呼び出した、またはシステムコールを実行したプロセス名) および **pid()** (現行プロセス ID) を使用しています。

一般的に使用される SystemTap 関数を以下に挙げます。

#### tid()

現行スレッドの ID。

#### uid()

現行ユーザーの ID。

#### cpu()

現行の CPU 番号。

#### gettimeofday\_s()

Unix epoch (1970 年 1 月 1 日) からの秒数。

#### ctime()

UNIX epoch からの秒数を日にちに換算。

#### pp()

現在処理されているプローブポイントを記述する文字列。

#### thread\_indent()

この関数はプリント結果をうまく整理するので、便利なものです。この関数はインデント差分の引数を取ります。これは、スレッドの「インデントカウンター」に追加する、またはそこから取り除くスペースの数を示すものです。その後、適切なインデントスペースの数と一般的な追跡データの文字列を返します。

ここで返された文字列の一般的なデータに含まれるのは、タイムスタンプ (スレッドの `thread_indent()` への最初の呼び出しからのマイクロ秒)、プロセス名、およびスレッド ID です。これによりどの関数が呼び出されたか、誰が呼び出したか、各関数呼び出しの長さが特定できます。

それぞれの呼び出しが終わり次第、次の呼び出しが始まれば、エントリーと終了の一致は容易です。ただし、ほとんどの場合は、最初の関数呼び出しエントリーがなされた後、これが終了する前に他の複数の呼び出しがエントリーおよび終了することがあります。インデントカウンターがあることで、最初の呼び出しが終了していない場合、次の関数呼び出しをインデントして、エントリーとそれに対応する終了が一致しやすくなります。

以下で `thread_indent()` の使用例を見てみましょう。

### 例3.6 `thread_indent.stp`

```
probe kernel.function("**@net/socket.c")
{
  printf ("%s -> %s\n", thread_indent(1), probefunc())
}
probe kernel.function("**@net/socket.c").return
{
  printf ("%s <- %s\n", thread_indent(-1), probefunc())
}
```

例3.6 「`thread_indent.stp`」では、各イベントでの `thread_indent()` とプローブの関数を以下の書式でプリントします。

```
0 ftp(7223): -> sys_socketcall
1159 ftp(7223): -> sys_socket
2173 ftp(7223): -> __sock_create
2286 ftp(7223): -> sock_alloc_inode
2737 ftp(7223): <- sock_alloc_inode
3349 ftp(7223): -> sock_alloc
3389 ftp(7223): <- sock_alloc
3417 ftp(7223): <- __sock_create
4117 ftp(7223): -> sock_create
4160 ftp(7223): <- sock_create
4301 ftp(7223): -> sock_map_fd
4644 ftp(7223): -> sock_map_file
4699 ftp(7223): <- sock_map_file
4715 ftp(7223): <- sock_map_fd
4732 ftp(7223): <- sys_socket
4775 ftp(7223): <- sys_socketcall
```

このサンプル出力には、以下の情報が含まれています。

- スレッド (`thread_indent()` から文字列に追加) の最初の `thread_indent()` 呼び出しからの時間 (マイクロ秒単位)。
- 関数呼び出し (`thread_indent()` から文字列に追加) を実施したプロセス名 (およびその対応 ID)。
- 呼び出しがエントリー (<-) か終了 (->) かを示す矢印。インデントがあることで、特定の関数呼び出しのエントリーとそれに対応する終了が一致しやすくなります。

- プロセスが呼び出した関数名。

#### name

特定のシステムコールの名前を識別します。この変数は、イベント `syscall.system_call` を使用するプローブでのみ使用可能です。

#### target()

`stap script -x process ID` または `stap script -c command` と併用されます。プロセス ID またはコマンドの引数を取るスクリプトを指定したい場合、スクリプト内で参照先となる変数として `target()` を使用します。以下に例を示します。

#### 例3.7 targetexample.stp

```
probe syscall.* {
  if (pid() == target())
    printf("%s/n", name)
}
```

例3.7 「targetexample.stp」を引数 `-x process ID` で実行すると、(`syscall.*` イベントで指定された)すべてのシステムコールを監視し、指定されたプロセスで実行された全システムコールの名前をプリントします。

これは、特定のプロセスをターゲットとしたい場合に毎回 `if (pid() == process ID)` と指定することと同様の効果があります。ただし、`target()` を使用すると、スクリプトの再利用が容易になり、スクリプトを実行するたびに引数としてプロセス ID を渡すだけで済みます (例: `stap targetexample.stp -x process ID`)。

サポートされる SystemTap 関数の詳細情報は、`man stapfuncs` を参照してください。

## 3.3. 基本的な SYSTEMTAP ハンドラーコンストラクト

SystemTap は、ハンドラーでいくつかの基本的なコンストラクトの使用をサポートしています。ほとんどのこれらハンドラーコンストラクトの構文は、C および `awk` 構文に基づいています。本セクションでは、SystemTap ハンドラーコンストラクトで最も有用なものをいくつか説明します。これで、簡潔かつ便利な SystemTap スクリプトの作成ができるようになります。

### 3.3.1. 変数

ハンドラーでは、変数を自由に使うことができます。名前を選択し、関数もしくは式から値を割り当て、式内で値を使用します。SystemTap は、割り当てられた値のタイプに基づいて、自動的に変数が文字列か整数かを識別します。たとえば、変数 `foo` を `gettimeofday_s()` に (`foo = gettimeofday_s()` として) 設定すると、`foo` は数字として識別され、`printf()` では整数の書式指定子 (`%d`) でプリントされます。

ただしデフォルトでは、変数は使用されているプローブのみのローカルとなります。つまり、変数はプローブハンドラーが呼び出されるたびに初期化され、使用され、破棄されます。複数のプローブで変数を共有するには、プローブの外で `global` を使用して変数名を宣言します。以下の例を見てください。

#### 例3.8 timer-jiffies.stp

```

global count_jiffies, count_ms
probe timer.jiffies(100) { count_jiffies ++ }
probe timer.ms(100) { count_ms ++ }
probe timer.ms(12345)
{
  hz=(1000*count_jiffies) / count_ms
  printf ("jiffies:ms ratio %d:%d => CONFIG_HZ=%d\n",
    count_jiffies, count_ms, hz)
  exit ()
}

```

例3.8 「timer-jiffies.stp」では、jiffies およびミリ秒をカウントするタイマーを使用してカーネルの **CONFIG\_HZ** 設定を計算し、それに応じて計算しています。global ステートメントにより、スクリプトは変数 **count\_jiffies** および **count\_ms** (各プローブで設定) を **probe timer.ms(12345)** と共有して使用することができます。

### 注記

例3.8 「timer-jiffies.stp」の ++ 表記 (**count\_jiffies ++** と **count\_ms ++**) は、変数の値を1増やすために使用されます。以下のプローブで **count\_jiffies** は、100 jiffies ごとに1つ増えます。

```

probe timer.jiffies(100) { count_jiffies ++ }

```

この場合、SystemTap は **count\_jiffies** が整数であると理解します。**count\_jiffies** には初期値が割り当てられていないことから、デフォルトでは初期値はゼロになります。

## 3.3.2. 条件付き (conditional) ステートメント

場合によっては、SystemTap スクリプトの出力が大きすぎることもあるかもしれません。これに対処するには、スクリプトの論理を細分化して、出力をプローブに関連するもしくは有用なものにする必要があります。

これはハンドラーで **条件** を使うことで実行できます。SystemTap は以下のタイプの条件付きステートメントを受け付けます。

### If/Else ステートメント

書式は以下のようになります。

```

if (condition)
  statement1
else
  statement2

```

**condition** 式がゼロ以外の場合、**statement1** が実行されます。**condition** 式がゼロの場合、**statement2** が実行されます。**else** 節 (**else statement2**) はオプションです。**statement1** と **statement2** は両方とも、ステートメントブロックになります。

### 例3.9 ifelse.stp

```

global countread, countnonread
probe kernel.function("vfs_read"),kernel.function("vfs_write")

```

```

{
  if (probefunc()=="vfs_read")
    countread ++
  else
    countnonread ++
}
probe timer.s(5) { exit() }
probe end
{
  printf("VFS reads total %d\n VFS writes total %d\n", countread, countnonread)
}

```

例3.9 「ifelse.stp」は、5秒間にシステムが実行する仮想ファイルシステムの読み込み (**vfs\_read**) と書き込み (**vfs\_write**) 回数をカウントするスクリプトです。この実行時には、プローブした関数の名前が (条件 **if (probefunc()=="vfs\_read")**) の **vfs\_read** と一致する場合、スクリプトは変数 **countread** の値を1つ増やします。一致しない場合は、**countnonread (else {countnonread ++})** を増やします。

### While ループ

書式は以下のようになります。

```

while (condition)
  statement

```

**condition** がゼロ以外であれば、**statement** 内のステートメントのブロックは実行されます。**statement** はステートメントブロックであることが多く、**condition** が最終的にゼロとなるように値を変更する必要があります。

### For ループ

書式は以下のようになります。

```

for (initialization; conditional; increment) statement

```

**for** ループは、単に **while** ループの短縮形です。以下が同等の **while** ループになります。

```

initialization
while (conditional) {
  statement
  increment
}

```

### 条件演算子

等号 **==** のほかに、以下の演算子も条件付きステートメントに使用できます。

**>=**

より大か等しい

**<=**

より小か等しい

!=

等しくない

### 3.3.3. コマンドラインの引数

SystemTap スクリプトでは、**\$** または **@** の直後にコマンドライン上の引数の番号を続けることで、単純なコマンドライン引数を受け付けるようにすることができます。コマンドライン引数としてユーザーが整数を入力すると思われる場合は **\$** を、文字列が予測される場合は **@** を使用します。

#### 例3.10 commandlineargs.stp

```
probe kernel.function(@1) {}
probe kernel.function(@1).return {}
```

例3.10 「`commandlineargs.stp`」は例3.1 「`wildcards.stp`」と似ていますが、(**stap commandlineargs.stp kernel function** のように) プローブするカーネル関数をコマンドライン引数として渡すことができる点で異なります。また、**@1**、**@2** のように、ユーザーが入力した順番で複数のコマンドライン引数をスクリプトが受け付けるように指定することもできます。

## 3.4. 連想アレイ

SystemTap は、連想アレイの使用もサポートします。通常の変数は単一の値を表しますが、連想アレイでは値の集合を表すことができます。簡単に言うと、連想アレイは一意的鍵の集合です。アレイ内のそれぞれの鍵にそれに関連する値があります。

(これより後で紹介するように) 連想アレイは通常複数のプローブで処理されるので、SystemTap スクリプトでは **global** 変数として宣言されるべきです。連想アレイ内の要素にアクセスする構文は **awk** の構文と似ており、以下のようになります。

```
array_name[index_expression]
```

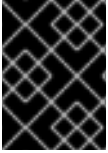
ここで **array\_name** は、アレイが使用する任意の名前です。**index\_expression** は、アレイ内の特定の一意の鍵を見るために使用されます。例として、**tom**、**dick**、および **harry** という3人の年齢(一意の鍵)を指定する、**foo** という名前のアレイを構築してみましょう。以下のアレイステートメントを使用して、3人にそれぞれ23、24、25という年齢(関連する値)を割り当てることができます。

#### 例3.11 基本的なアレイステートメント

```
foo["tom"] = 23
foo["dick"] = 24
foo["harry"] = 25
```

アレイステートメント内では最大9つのインデックス式を指定することができ、それぞれをコンマ(,)で区切ります。これは、鍵に複数の情報が含まれる場合に便利です。`disktop.stp` からの以下の行では、プロセスID、実行可能ファイル名、ユーザーID、親プロセスのID、および文字列「W」という5つの要素を使用しています。ここでは、**devname** の値を鍵に関連付けています。

```
device[pid(),execname(),uid(),ppid(),"W"] = devname
```



## 重要

連想アレイは、単一プローブで使用されるか複数プローブで使用されるかに関わらず、すべてを **global** と宣言する必要があります。

## 3.5. SYSTEMTAP でのアレイ演算

本セクションでは、SystemTap で使用される最も一般的なアレイ演算を説明しています。

### 3.5.1. 関連する値の割り当て

インデックス化された一意のペアに関連する値を設定するには、等号 `=` を使います。

```
array_name[index_expression] = value
```

例3.11「基本的なアレイステートメント」では、明示的な関連する値を一意の鍵に設定する非常に基本的な例を示しています。 **index\_expression** と **value** の両方に、ハンドラー関数を使用することもできます。たとえば以下のように、アレイを使用して、タイムスタンプをプロセス名(これを一意の鍵として使用)への関連する値として設定することができます。

#### 例3.12 タイムスタンプをプロセス名に関連付ける

```
foo[tid()] = gettimeofday_s()
```

例3.12「タイムスタンプをプロセス名に関連付ける」ではイベントがステートメントを呼び出すと、SystemTap は適切な **tid()** 値(つまり、スレッドの ID。これは一意の鍵として使用されます)を返します。同時に SystemTap は関数 **gettimeofday\_s()** を使用して、対応するタイムスタンプを関数 **tid()** で定義されている一意の鍵への関連する値として設定します。これで、スレッド ID とタイムスタンプを含む鍵のペアで構成されるアレイが作成されます。

この例では、**tid()** がアレイ **foo** ですすでに定義されている値を返すと、この演算子はその値に関連付けられている元の値を破棄し、**gettimeofday\_s()** からの現行タイムスタンプで置き換えます。

### 3.5.2. アレイからの値の読み取り

アレイからの値の読み取りは、変数値の読み取りと同じ方法でできます。これを行うには、**array\_name[index\_expression]** ステートメントを数式に要素として含めます。以下に例を示します。

#### 例3.13 単純計算でのアレイ値の使用

```
delta = gettimeofday_s() - foo[tid()]
```

この例では、(「関連する値の割り当て」からの)例3.12「タイムスタンプをプロセス名に関連付ける」のコンストラクトを使用してアレイ **foo** が構築されていることを想定しています。これで参照ポイントとなるタイムスタンプが設定され、**delta** の計算に使用されます。

例3.13「単純計算でのアレイ値の使用」のコンストラクトは、現行の **gettimeofday\_s()** から鍵 **tid()** の関連する値を差し引くことで、変数 **delta** の値を計算します。このコンストラクトは、**tid()** の値をアレイから読み取ることで計算を行います。このコンストラクトは、読み取り操作の開始と完了など、2



つのイベント間の時間を判定する際に便利なものです。



### 注記

*index\_expression* が一意の鍵を見つけられない場合は、0 の値 (例3.13 「単純計算でのアレイ値の使用」 などの数値演算の場合) もしくは null/空の文字列の値 (文字列の演算の場合) がデフォルトで返されます。

### 3.5.3. 関連する値の増加

アレイ内の一意の鍵の関連する値を増やすには、**++** を使用します。

```
array_name[index_expression] ++
```

ここでも、*index\_expression* にハンドラー関数を使用できます。たとえば、仮想ファイルシステムへの読み込み (**vfs.read** イベントを使用) を特定のプロセスが実行した回数を計算したい場合は、以下のプローブを使用します。

#### 例3.14 vfsreads.stp

```
probe vfs.read
{
  reads[execname()] ++
}
```

例3.14 「**vfsreads.stp**」 では、プローブが最初にプロセス名 **gnome-terminal** を返す際 (つまり、**gnome-terminal** が初めて VFS 読み込みを実行する際)、そのプロセス名は一意の鍵 **gnome-terminal** に関連する値 1 が付いたものになります。プローブがプロセス名 **gnome-terminal** を次に返す際には、SystemTap は **gnome-terminal** の関連する値を 1 増やします。SystemTap は、プローブがプロセス名を返す際にこの演算を **すべての** プロセス名に対して実行します。

### 3.5.4. アレイ内での複数要素の処理

アレイで十分な情報を収集したら、それを有用なものにするためにアレイで全要素を取得して処理する必要があります。例3.14 「**vfsreads.stp**」 を見てみましょう。このスクリプトは各プロセスが何回 VFS 読み込みを行ったかという情報を収集しますが、その情報をどうするかについては指定していません。例3.14 「**vfsreads.stp**」 を有用にする簡単な方法は、**reads** アレイで鍵のペアをプリントすることです。

アレイ内の鍵のペアすべてを処理する最善の方法 (反復として) は、**foreach** ステートメントを使用することです。以下の例を見てみましょう。

#### 例3.15 cumulative-vfsreads.stp

```
global reads
probe vfs.read
{
  reads[execname()] ++
}
probe timer.s(3)
{
```



```

foreach (count in reads)
  printf("%s : %d \n", count, reads[count])
}

```

例3.15 「cumulative-vfsreads.stp」の2つ目のプローブでは、**foreach** ステートメントが **count** 変数を使用して **reads** アレイ内の一意の鍵の反復を参照しています。同じプローブ内の **reads[count]** アレイステートメントは、一意の鍵に関連する値を取得します。

例3.15 「cumulative-vfsreads.stp」の最初のプローブでは、スクリプトはVFS-readの統計情報を3秒ごとにプリントし、VFS-readを実行したプロセス名とその回数を表示します。

例3.15 「cumulative-vfsreads.stp」の **foreach** ステートメントは、順不同でアレイ内のプロセス名の全反復をプリントすることに注意してください。+(昇順)または-(降順)を使用すると、スクリプトに特定の順番で反復をプロセスするよう指示することができます。さらに、**limit value** オプションを使うと、スクリプトがプロセスする反復数を制限することもできます。

以下の代替プローブ例を見てみましょう。

```

probe timer.s(3)
{
  foreach (count in reads- limit 10)
    printf("%s : %d \n", count, reads[count])
}

```

この **foreach** ステートメントは、スクリプトにアレイ **reads** 内の要素を (関連する値の) 降順で処理するよう指示します。**limit 10** オプションは、**foreach** に最初の10の反復のみを処理するよう指示します (つまり、値の高い上位10位の反復のみをプリントします)。

### 3.5.5. アレイおよびアレイ要素の消去/削除

別のプローブに再使用するために、アレイ要素内の関連する値を消去したり、アレイ全体をリセットする必要がある場合もあります。「アレイ内での複数要素の処理」の例3.15 「cumulative-vfsreads.stp」では、時間の経過とともにプロセスごとのVFS readsの増加が分かりますが、3秒間で各プロセスが実行したVFS readsの数は表示されません。

これを実行するには、アレイが累積した値を消去する必要があります。**delete** 演算子を使用してアレイ内の要素またはアレイ全体を削除すると、これが実行できます。以下の例を見てみましょう。

#### 例3.16 noncumulative-vfsreads.stp

```

global reads
probe vfs.read
{
  reads[execname()] ++
}
probe timer.s(3)
{
  foreach (count in reads)
    printf("%s : %d \n", count, reads[count])
  delete reads
}

```

例3.16 「`noncumulative-vfsreads.stp`」では、2つ目のプローブが調査した3秒間のみで各プロセスが実行したVFS reads 数をプリントします。`delete reads` ステートメントは、プローブ内の `reads` アレイを消去します。

### 注記

同一プローブ内には、複数のアレイ演算を設置することが可能です。「[アレイ内での複数要素の処理](#)」および「[アレイおよびアレイ要素の消去/削除](#)」の例を使用すると、3秒間でプロセスが実行したVFS reads の数を追跡し、かつそれらのプロセスでの累積VFS reads 数を集計することができます。以下の例を見てみましょう。

```
global reads, totalreads

probe vfs.read
{
  reads[execname()] ++
  totalreads[execname()] ++
}

probe timer.s(3)
{
  printf("=====\n")
  foreach (count in reads-)
    printf("%s : %d \n", count, reads[count])
  delete reads
}

probe end
{
  printf("TOTALS\n")
  foreach (total in totalreads-)
    printf("%s : %d \n", total, totalreads[total])
}
```

この例では、`reads` と `totalreads` のアレイが同じ情報を追跡し、同様の方式でプリントします。唯一の違いは、`reads` は3秒ごとに消去されるのに対して、`totalreads` は増え続けるという点です。

### 3.5.6. 条件付きステートメントにおけるアレイの使用

連想アレイは `if` ステートメントでも使用することができます。これは、アレイ内の値が特定の条件に一致した場合にサブルーチンを実行するという場合に便利です。以下の例を見てみましょう。

#### 例3.17 `vfsreads-print-if-1kb.stp`

```
global reads
probe vfs.read
{
  reads[execname()] ++
}

probe timer.s(3)
{
  printf("=====\n")
```

```

foreach (count in reads-)
  if (reads[count] >= 1024)
    printf("%s : %dkB \n", count, reads[count]/1024)
  else
    printf("%s : %dB \n", count, reads[count])
}

```

例3.17 「`vfsreads-print-if-1kb.stp`」では、全プロセス一覧と各プロセスが VFS read を実行した回数が 3 秒ごとにプリントされます。プロセス名の関連する値が 1024 以上の場合、スクリプト内の `if` ステートメントがこれを変換し、**kB** でプリントします。

### メンバーシップのテスト

特定の一意の鍵がアレイのメンバーかどうかをテストすることもできます。以下の例のように、`if` ステートメント内でアレイ内のメンバーシップを使用することも可能です。

```
if([index_expression] in array_name) statement
```

以下の例を見てみましょう。

### 例3.18 `vfsreads-stop-on-stapio2.stp`

```

global reads

probe vfs.read
{
  reads[execname()] ++
}

probe timer.s(3)
{
  printf("=====\n")
  foreach (count in reads+)
    printf("%s : %d \n", count, reads[count])
  if(["stapio"] in reads) {
    printf("stapio read detected, exiting\n")
    exit()
  }
}

```

`if(["stapio"] in reads)` ステートメントは、一意の鍵 **stapio** がアレイ **reads** に追加されたら、**stapio read detected, exiting** をプリントするようにスクリプトに指示します。

### 3.5.7. 統計集計の計算

統計集計は、新規データを素早くかつ大量に累積 (集計ストリーム統計情報のみを保存) することが重要な場合に、数値の統計情報を収集するために使用されます。統計集計はグローバル変数またはアレイ内の要素として使用できます。

統計集計に値を追加するには、演算子 `<<< value` を使用します。

### 例3.19 `stat-aggregates.stp`

```
global reads
probe vfs.read
{
  reads[execname()] <<< count
}
```

例3.19 「stat-aggregates.stp」では、演算子 `<<< count` が、`reads` アレイ内の対応する `execname()` の関連する値に `count` が返した数字を保存します。これらの値は保存されるのであって、各一意の鍵の関連する値に追加されたり、現行の関連する値に置き換わるものではありません。各一意の鍵 (`execname()`) に複数の関連する値があり、ハンドラーが実行するプローブで累積していると考えればよいでしょう。



### 注記

例3.19 「stat-aggregates.stp」のコンテキストでは、`count` は、仮想ファイルシステムに返された `execname()` が書き込んだデータ量を返します。

統計集計が収集したデータを抽出するには、構文書式 `@extractor(variable/array index expression)` を使用します。`extractor` は、以下のいずれかの整数抽出になります。

#### count

変数やアレイのインデックス式に保存されたすべての値の数を返します。例3.19 「stat-aggregates.stp」のサンプルプローブでは、`@count(writes[execname()])` の式は、アレイ `writes` の各一意の鍵に保存されている値の数を返します。

#### sum

変数やアレイのインデックス式に保存されたすべての値の合計を返します。例3.19 「stat-aggregates.stp」のサンプルプローブでは、`@sum(writes[execname()])` の式は、アレイ `writes` の各一意の鍵に保存されている値すべての合計を返します。

#### min

変数やアレイのインデックス式に保存されたすべての値で最も小さいものを返します。

#### max

変数やアレイのインデックス式に保存されたすべての値で最も大きいものを返します。

#### avg

変数やアレイのインデックス式に保存されたすべての値の平均を返します。

統計集計を使用する際には、複数のインデックス式 (最大5つ) を使用するアレイコンストラクトを構築することもできます。これは、プローブ中に追加のコンテキスト情報を捕捉する際に便利です。以下に例を示します。

### 例3.20 複数のアレイインデックス

```
global reads
probe vfs.read
{
  reads[execname(),pid()] <<< 1
}
```

```

}
probe timer.s(3)
{
  foreach([var1,var2] in reads)
    printf("%s (%d) : %d \n", var1, var2, @count(reads[var1,var2]))
}

```

例3.20 「複数のアレイインデックス」では、最初のプローブで各プロセスが実行した VFS read の数を追跡します。この例が他と異なる点は、このアレイは実行された read をプロセス名とそれに対応するプロセス ID の両方に関連付けしている点です。

例3.20 「複数のアレイインデックス」の2つ目のプローブは、アレイ **reads** が収集した情報を処理してプリントする方法を示しています。**foreach** ステートメントは、最初のプローブのアレイ **reads** の最初のインスタンスに含まれる変数 (**var1** および **var2**) と同じ数を使用している点に注意してください。

### 3.6. TAPSETS

Tapsets は、SystemTap スクリプトで使用する事前作成されたプローブおよび関数のライブラリーを形成するスクリプトです。ユーザーが SystemTap スクリプトを実行すると、SystemTap はスクリプトのプローブイベントとハンドラーを tapset ライブラリーに対して確認します。そして SystemTap はスクリプトを C 言語に変換する前に対応するプローブと関数を読み込みます (SystemTap で発生するセッションの情報は、「アーキテクチャー」を参照してください)。

SystemTap スクリプトと同様に、tapsets はファイル名拡張子 **.stp** を使用します。tapsets の標準ライブラリーは、デフォルトで **/usr/share/systemtap/tapset/** にあります。ただし SystemTap スクリプトとは異なり、tapsets は直接実行するものではなく、他のスクリプトがそこから定義をプルできるライブラリーを構成します。

簡単に言うと、tapset ライブラリーは抽象化レイヤーで、ユーザーによるイベントおよび関数の定義を容易にするように設計されています。言うなれば、Tapsets は、ユーザーがイベントとして特定を希望する可能性のある関数に便利なエイリアスを提供します。適切なエイリアスを知ることは、ほとんどの場合、カーネルのバージョンごとに異なる特定のカーネル関数を覚えるよりも容易です。

「イベント」および「SystemTap 関数」のいくつかのハンドラーと関数は、tapsets で定義されています。たとえば、**thread\_indent()** は **indent.stp** で定義されています。

## 第4章 便利な SYSTEMTAP スクリプト

本章では、各種のサブシステムの監視および調査に使用可能な SystemTap スクリプトをいくつか説明します。これらのスクリプトはすべて、**systemtap-testsuite** RPM をインストールした時点 `/usr/share/systemtap/testsuite/systemtap.examples/` で利用できます。

### 4.1. ネットワーク

以下のセクションでは、ネットワーク関連の関数を追跡し、ネットワークアクティビティのプロファイルを構築するスクリプトを説明します。

#### 4.1.1. ネットワークのプロファイリング

このセクションでは、ネットワークアクティビティのプロファイルを実行する方法を説明します。`nettop.stp` では、マシン上で各プロセスが生成しているネットワークトラフィックの量が確認できます。

##### nettop.stp

```
#!/usr/bin/env stap

global ifxmit, ifrecv
global ifmerged

probe netdev.transmit
{
  ifxmit[pid(), dev_name, execname(), uid()] <<< length
}

probe netdev.receive
{
  ifrecv[pid(), dev_name, execname(), uid()] <<< length
}

function print_activity()
{
  printf("%5s %5s %-7s %7s %7s %7s %-15s\n",
    "PID", "UID", "DEV", "XMIT_PK", "RECV_PK",
    "XMIT_KB", "RECV_KB", "COMMAND")

  foreach ([pid, dev, exec, uid] in ifrecv) {
    ifmerged[pid, dev, exec, uid] += @count(ifrecv[pid,dev,exec,uid]);
  }
  foreach ([pid, dev, exec, uid] in ifxmit) {
    ifmerged[pid, dev, exec, uid] += @count(ifxmit[pid,dev,exec,uid]);
  }
  foreach ([pid, dev, exec, uid] in ifmerged-) {
    n_xmit = @count(ifxmit[pid, dev, exec, uid])
    n_recv = @count(ifrecv[pid, dev, exec, uid])
    printf("%5d %5d %-7s %7d %7d %7d %7d %-15s\n",
      pid, uid, dev, n_xmit, n_recv,
      n_xmit ? @sum(ifxmit[pid, dev, exec, uid])/1024 : 0,
      n_recv ? @sum(ifrecv[pid, dev, exec, uid])/1024 : 0,
      exec)
  }
}
```

```

print("\n")

delete ifxmit
delete ifrecv
delete ifmerged
}

probe timer.ms(5000), end, error
{
  print_activity()
}

```

**function print\_activity()** は、以下の式を使用する点に注意してください。

```

n_xmit ? @sum(ifxmit[pid, dev, exec, uid])/1024 : 0
n_recv ? @sum(ifrecv[pid, dev, exec, uid])/1024 : 0

```

これらの式は、if/else 条件です。最初のステートメントは、以下の擬似コードをより簡潔にしたものです。

```

if n_recv != 0 then
  @sum(ifrecv[pid, dev, exec, uid])/1024
else
  0

```

**nettop.stp** では、どのプロセスがシステム上でネットワークトラフィックを生成しているかを追跡し、各プロセスについて以下の情報を提供します。

- **PID** – プロセスの ID。
- **UID** – ユーザー ID。ユーザー ID が **0** の場合は、root ユーザーを参照します。
- **DEV** – プロセスがデータの送受信に使用したイーサネットデバイス (例: eth0、eth1)。
- **XMIT\_PK** – プロセスが送信したパケット数。
- **RECV\_PK** – プロセスが受信したパケット数。
- **XMIT\_KB** – プロセスが送信したデータ量 (キロバイト単位)。
- **RECV\_KB** – サービスが受信したデータ量 (キロバイト単位)。

**nettop.stp** では、5 秒ごとにネットワークプロファイルのサンプルが提供されます。**probe timer.ms(5000)** を編集すると、この間隔を変更できます。例4.1「**nettop.stp** のサンプル出力」には、**nettop.stp** からの 20 秒間に渡る出力を引用してあります。

#### 例4.1 nettop.stp のサンプル出力

```

[...]
PID UID DEV XMIT_PK RECV_PK XMIT_KB RECV_KB COMMAND
0 0 eth0 0 5 0 0 swapper
11178 0 eth0 2 0 0 0 synergyc

```

```

PID UID DEV XMIT_PK RECV_PK XMIT_KB RECV_KB COMMAND
2886 4 eth0 79 0 5 0 cups-polld
11362 0 eth0 0 61 0 5 firefox
0 0 eth0 3 32 0 3 swapper
2886 4 lo 4 4 0 0 cups-polld
11178 0 eth0 3 0 0 0 synergyc

PID UID DEV XMIT_PK RECV_PK XMIT_KB RECV_KB COMMAND
0 0 eth0 0 6 0 0 swapper
2886 4 lo 2 2 0 0 cups-polld
11178 0 eth0 3 0 0 0 synergyc
3611 0 eth0 0 1 0 0 Xorg

PID UID DEV XMIT_PK RECV_PK XMIT_KB RECV_KB COMMAND
0 0 eth0 3 42 0 2 swapper
11178 0 eth0 43 1 3 0 synergyc
11362 0 eth0 0 7 0 0 firefox
3897 0 eth0 0 1 0 0 multiload-apple
[...]
```

#### 4.1.2. ネットワークソケットコードで呼び出された関数の追跡

このセクションでは、カーネルの `net/socket.c` ファイルから呼び出された関数を追跡する方法を説明します。このタスクでは、各プロセスがカーネルレベルでネットワークと対話する様子が詳細に分かります。

##### socket-trace.stp

```

#!/usr/bin/stap

probe kernel.function("**@net/socket.c").call {
  printf ("%s -> %s\n", thread_indent(1), probefunc())
}
probe kernel.function("**@net/socket.c").return {
  printf ("%s <- %s\n", thread_indent(-1), probefunc())
}
```

`socket-trace.stp` は、`thread_indent()` が機能する様子の説明に使用した [SystemTap 関数の例 3.6 「thread\\_indent.stp」](#) と同じものです。

##### 例4.2 socket-trace.stp のサンプル出力

```

[...]
```

```

0 Xorg(3611): -> sock_poll
3 Xorg(3611): <- sock_poll
0 Xorg(3611): -> sock_poll
3 Xorg(3611): <- sock_poll
0 gnome-terminal(11106): -> sock_poll
5 gnome-terminal(11106): <- sock_poll
0 scim-bridge(3883): -> sock_poll
3 scim-bridge(3883): <- sock_poll
0 scim-bridge(3883): -> sys_socketcall
```



```

4 scim-bridge(3883): -> sys_recv
8 scim-bridge(3883): -> sys_recvfrom
12 scim-bridge(3883):-> sock_from_file
16 scim-bridge(3883):<- sock_from_file
20 scim-bridge(3883):-> sock_recvmsg
24 scim-bridge(3883):<- sock_recvmsg
28 scim-bridge(3883): <- sys_recvfrom
31 scim-bridge(3883): <- sys_recv
35 scim-bridge(3883): <- sys_socketcall
[...]
```

例4.2 「[socket-trace.stp のサンプル出力](#)」には、3秒間の [socket-trace.stp](#) の出力を引用してあります。[thread\\_indent\(\)](#) が提供するこのスクリプトの出力に関する詳細情報は、[SystemTap 関数 例 3.6 「thread\\_indent.stp」](#) を参照してください。

### 4.1.3. 着信 TCP 接続の監視

このセクションでは、着信 TCP 接続の監視方法を説明します。このタスクは、承認されていない、疑わしい、さもなければ望ましくないネットワークアクセス要求をリアルタイムで特定する場合に便利です。

#### tcp\_connections.stp

```

#!/usr/bin/env stap

probe begin {
    printf("%6s %16s %6s %6s %16s\n",
        "UID", "CMD", "PID", "PORT", "IP_SOURCE")
}

probe kernel.function("tcp_accept").return?,
    kernel.function("inet_csk_accept").return? {
    sock = $return
    if (sock != 0)
        printf("%6d %16s %6d %6d %16s\n", uid(), execname(), pid(),
            inet_get_local_port(sock), inet_get_ip_source(sock))
}

```

[tcp\\_connections.stp](#) の実行中は、システムが受け付けた着信 TCP 接続の以下の情報がリアルタイムでプリントアウトされます。

- 現在の **UID**
- **CMD** - 接続を受け付けるコマンド
- そのコマンドの **PID**
- 接続が使用するポート
- TCP 接続の発信元となる IP アドレス

#### 例4.3 [tcp\\_connections.stp](#) のサンプル出力

```

UID      CMD  PID  PORT  IP_SOURCE
0        sshd 3165  22   10.64.0.227
0        sshd 3165  22   10.64.0.227

```

#### 4.1.4. カーネルでのネットワークパケットドロップの監視

Linux のネットワークスタックは、様々な理由でパケットを破棄する場合があります。Linux カーネルにはトレースポイント `kernel.trace("kfree_skb")` を含むものもあり、これは簡単にパケットが破棄された場所を追跡します。`dropwatch.stp` では、`kernel.trace("kfree_skb")` を使用して、パケットの破棄を追跡します。このスクリプトは、パケットが破棄された場所を 5 秒ごとに要約します。

##### dropwatch.stp

```

#!/usr/bin/stap

#####
# Dropwatch.stp
# Author: Neil Horman <nhorman@redhat.com>
# An example script to mimic the behavior of the dropwatch utility
# http://fedorahosted.org/dropwatch
#####

# Array to hold the list of drop points we find
global locations

# Note when we turn the monitor on and off
probe begin { printf("Monitoring for dropped packets\n") }
probe end { printf("Stopping dropped packet monitor\n") }

# increment a drop counter for every location we drop at
probe kernel.trace("kfree_skb") { locations[$location] <<< 1 }

# Every 5 seconds report our drop locations
probe timer.sec(5)
{
  printf("\n")
  foreach (l in locations-) {
    printf("%d packets dropped at location %p\n",
      @count(locations[l]), l)
  }
  delete locations
}

```

`kernel.trace("kfree_skb")` は、カーネル内でネットワークパケットがドロップした場所を追跡します。`kernel.trace("kfree_skb")` には、解放されているバッファへのポインター (`$skb`) と、解放されているバッファのカーネルコード内での場所 (`$location`) という 2 つの引数があります。

`dropwatch.stp` スクリプトを 15 秒間実行すると、例 4.4 「`dropwatch.stp` のサンプル出力」のような結果が出力されます。ここでは、トレースポイントアドレスと実際のアドレスのミス数が記載されています。

#### 例4.4 dropwatch.stp のサンプル出力

```
Monitoring for dropped packets

51 packets dropped at location 0xffffffff8024cd0f
2 packets dropped at location 0xffffffff8044b472

51 packets dropped at location 0xffffffff8024cd0f
1 packets dropped at location 0xffffffff8044b472

97 packets dropped at location 0xffffffff8024cd0f
1 packets dropped at location 0xffffffff8044b472
Stopping dropped packet monitor
```

パケットドロップの場所をより意味のあるものにするには、`/boot/System.map-`uname -r`` ファイルを参照してください。このファイルには、各関数の開始アドレスが記載されており、例 4.4 「dropwatch.stp のサンプル出力」 の出力のアドレスを特定の関数名にマップできます。以下の `/boot/System.map-`uname -r`` ファイルの抜粋を使用して、アドレス `0xffffffff8024cd0f` を関数 `unix_stream_recvmsg` に、アドレス `0xffffffff8044b472` を関数 `arp_rcv` にマッピングします。

```
[...]
ffffff8024c5cd T unlock_new_inode
ffffff8024c5da t unix_stream_sendmsg
ffffff8024c920 t unix_stream_recvmsg
ffffff8024cea1 t udp_v4_lookup_longway
[...]
ffffff8044addc t arp_process
ffffff8044b360 t arp_rcv
ffffff8044b487 t parp_redo
ffffff8044b48c t arp_solicit
[...]
```

## 4.2. ディスク

以下のセクションでは、ディスクおよび I/O アクティビティを監視するスクリプトを説明します。

### 4.2.1. ディスク読み取り/書き込みトラフィックの要約

このセクションでは、どのプロセスが最も重いディスクの読み取り/書き込みをシステムに実行しているかを特定する方法を説明します。

#### disktop.stp

```
#!/usr/bin/stap
#
# Copyright (C) 2007 Oracle Corp.
#
# Get the status of reading/writing disk every 5 seconds,
# output top ten entries
#
# This is free software,GNU General Public License (GPL);
# either version 2, or (at your option) any later version.
```

```
#
# Usage:
# ./disktop.stp
#

global io_stat,device
global read_bytes,write_bytes

probe vfs.read.return {
  if ($return>0) {
    if (devname!="N/A") { /*skip read from cache*/
      io_stat[pid(),execname(),uid(),ppid(),"R"] += $return
      device[pid(),execname(),uid(),ppid(),"R"] = devname
      read_bytes += $return
    }
  }
}

probe vfs.write.return {
  if ($return>0) {
    if (devname!="N/A") { /*skip update cache*/
      io_stat[pid(),execname(),uid(),ppid(),"W"] += $return
      device[pid(),execname(),uid(),ppid(),"W"] = devname
      write_bytes += $return
    }
  }
}

probe timer.ms(5000) {
  /* skip non-read/write disk */
  if (read_bytes+write_bytes) {

    printf("\n%-25s, %-8s%4dKb/sec, %-7s%6dKb, %-7s%6dKb\n\n",
           ctime(gettimeofday_s()),
           "Average:", ((read_bytes+write_bytes)/1024)/5,
           "Read:",read_bytes/1024,
           "Write:",write_bytes/1024)

    /* print header */
    printf("%8s %8s %8s %25s %8s %4s %12s\n",
           "UID","PID","PPID","CMD","DEVICE","T","BYTES")
  }
  /* print top ten I/O */
  foreach ([process,cmd,userid,parent,action] in io_stat- limit 10)
    printf("%8d %8d %8d %25s %8s %4s %12d\n",
           userid,process,parent,cmd,
           device[process,cmd,userid,parent,action],
           action,io_stat[process,cmd,userid,parent,action])

  /* clear data */
  delete io_stat
  delete device
  read_bytes = 0
  write_bytes = 0
}
```

```

probe end{
  delete io_stat
  delete device
  delete read_bytes
  delete write_bytes
}

```

`disktop.stp` は、ディスクに重い読み取りまたは書き込みを行なっているプロセス上位 10 位を出力します。例4.5「`disktop.stp` のサンプル出力」はこのスクリプトのサンプル出力となり、記載されているプロセスについて以下のデータが含まれます。

- **UID** – ユーザー ID。ユーザー ID が **0** の場合は、root ユーザーを参照します。
- **PID** – プロセスの ID。
- **PPID** – プロセスの 親プロセスのプロセス ID。
- **CMD** – プロセスの名前。
- **DEVICE** – プロセスが読み取りまたは書き込みを行なっているストレージデバイス。
- **T** – プロセスが実行したアクションのタイプ。**W** は書き込みを、**R** は読み取りを指します。
- **BYTES** – ディスクから読み取った、またはディスクに書き込んだデータ量。

`disktop.stp` の出力の日時は関数 `ctime()` と関数 `gettimeofday_s()` によって返されます。`ctime()` は、Unix 時間 (1970 年 1 月 1 日) 以降の秒単位経過をカレンダー時刻で引き出します。`gettimeofday_s()` は、Unix 時間以降の **実際の** 秒数をカウントします。これは、出力についてかなり正確でヒューマンリーダブルなタイムスタンプを提供します。

このスクリプトでは、`$return` は各プロセスが仮想ファイルシステムから読み取ったまたは書き込んだ実際のバイト数を保存するローカル変数です。`$return` は `return` プローブ (`vfs.read.return` や `vfs.read.return` など) での使用のみが可能です。

#### 例4.5 `disktop.stp` のサンプル出力

```

[...]
Mon Sep 29 03:38:28 2008 , Average: 19Kb/sec, Read: 7Kb, Write: 89Kb

UID  PID  PPID          CMD DEVICE  T  BYTES
0 26319 26294      firefox sda5  W   90229
0  2758  2757  pam_timestamp_c sda5  R    8064
0  2885   1      cupsd sda5  W    1678

Mon Sep 29 03:38:38 2008 , Average: 1Kb/sec, Read: 7Kb, Write: 1Kb

UID  PID  PPID          CMD DEVICE  T  BYTES
0  2758  2757  pam_timestamp_c sda5  R    8064
0  2885   1      cupsd sda5  W    1678

```

#### 4.2.2. ファイル読み取り/書き込みの I/O 時間の追跡

このセクションでは、各プロセスのファイルの読み取りおよび書き込み時間を監視する方法を説明します。これは、あるシステム上でどのファイルの読み込みが遅いかということ判断する際に便利です。

### io.time.stp

```
global start
global entry_io
global fd_io
global time_io

function timestamp:long() {
    return gettimeofday_us() - start
}

function proc:string() {
    return sprintf("%d (%s)", pid(), execname())
}

probe begin {
    start = gettimeofday_us()
}

global filenames
global filehandles
global fileread
global filewrite

probe syscall.open {
    filenames[pid()] = user_string($filename)
}

probe syscall.open.return {
    if ($return != -1) {
        filehandles[pid(), $return] = filenames[pid()]
        fileread[pid(), $return] = 0
        filewrite[pid(), $return] = 0
    } else {
        printf("%d %s access %s fail\n", timestamp(), proc(), filenames[pid()])
    }
    delete filenames[pid()]
}

probe syscall.read {
    if ($count > 0) {
        fileread[pid(), $fd] += $count
    }
    t = gettimeofday_us(); p = pid()
    entry_io[p] = t
    fd_io[p] = $fd
}

probe syscall.read.return {
    t = gettimeofday_us(); p = pid()
    fd = fd_io[p]
    time_io[p,fd] <<< t - entry_io[p]
}
```

```

probe syscall.write {
  if ($count > 0) {
    filewrite[pid(), $fd] += $count
  }
  t = gettimeofday_us(); p = pid()
  entry_io[p] = t
  fd_io[p] = $fd
}

probe syscall.write.return {
  t = gettimeofday_us(); p = pid()
  fd = fd_io[p]
  time_io[p,fd] <<< t - entry_io[p]
}

probe syscall.close {
  if (filehandles[pid(), $fd] != "") {
    printf("%d %s access %s read: %d write: %d\n", timestamp(), proc(),
      filehandles[pid(), $fd], fileread[pid(), $fd], filewrite[pid(), $fd])
    if (@count(time_io[pid(), $fd]))
      printf("%d %s iotime %s time: %d\n", timestamp(), proc(),
        filehandles[pid(), $fd], @sum(time_io[pid(), $fd]))
  }
  delete fileread[pid(), $fd]
  delete filewrite[pid(), $fd]
  delete filehandles[pid(), $fd]
  delete fd_io[pid()]
  delete entry_io[pid()]
  delete time_io[pid(),$fd]
}

```

`iotime.stp` では、システムコールが開かれる、閉じる、ファイルから読み取る、およびファイルに書き込む際に、毎回これを追跡します。`iotime.stp` はシステムコールがアクセスする各ファイルについて、読み取りもしくは書き込みが終了するまでの時間をマイクロ秒単位でカウントし、読み取りもしくは書き込みされたデータ量をバイト単位で追跡します。

`iotime.stp` は、ローカル変数 `$count` を使用してシステムコールが読み取りまたは書き込みを **試みた** データ量 (バイト単位) も追跡します。(「[ディスク読み取り/書き込みトラフィックの要約](#)」の `disktop.stp` で使用されている) `$return` は、読み取り/書き込みされた **実際の** データ量を保存することに注意してください。`$count` を使用できるのは、(`syscall.read` および `syscall.write` など) データの読み取りや書き込みを追跡するプローブのみです。

#### 例4.6 `iotime.stp` のサンプル出力

```

[...]
825946 3364 (NetworkManager) access /sys/class/net/eth0/carrier read: 8190 write: 0
825955 3364 (NetworkManager) iotime /sys/class/net/eth0/carrier time: 9
[...]
117061 2460 (pcscd) access /dev/bus/usb/003/001 read: 43 write: 0
117065 2460 (pcscd) iotime /dev/bus/usb/003/001 time: 7
[...]
3973737 2886 (sendmail) access /proc/loadavg read: 4096 write: 0
3973744 2886 (sendmail) iotime /proc/loadavg time: 11
[...]

```

例4.6 「`iotime.stp` のサンプル出力」では、以下のデータがプリントアウトされます。

- タイムスタンプ (マイクロ秒単位)
- プロセス ID およびプロセス名
- **access** または **iotime** のフラグ
- アクセスされたファイル

プロセスがデータの読み取りまたは書き込みを実行すると、**access** と **iotime** の行のペアと一緒に表示されます。**access** 行のタイムスタンプは、そのプロセスがファイルにアクセスを開始した時間を指します。その行の最後には、読み取り/書き込みされたデータ量がバイト単位で表示されます。**iotime** の行では、読み取り/書き込みにプロセスが費やした時間がマイクロ秒単位で表示されます。

**access** 行の後に **iotime** 行が続いていない場合は、プロセスがデータの読み取りまたは書き込みを行わなかったことを意味します。

### 4.2.3. 累積 I/O の追跡

このセクションでは、システムへの累積 I/O の量を追跡する方法を説明します。

#### `traceio.stp`

```
#!/usr/bin/env stap
# traceio.stp
# Copyright (C) 2007 Red Hat, Inc., Eugene Teo <eteo@redhat.com>
# Copyright (C) 2009 Kai Meyer <kai@unixlords.com>
# Fixed a bug that allows this to run longer
# And added the humanreadable function
#
# This program is free software; you can redistribute it and/or modify
# it under the terms of the GNU General Public License version 2 as
# published by the Free Software Foundation.
#

global reads, writes, total_io

probe vfs.read.return {
  reads[pid(),execname()] += $return
  total_io[pid(),execname()] += $return
}

probe vfs.write.return {
  writes[pid(),execname()] += $return
  total_io[pid(),execname()] += $return
}

function humanreadable(bytes) {
  if (bytes > 1024*1024*1024) {
    return sprintf("%d GiB", bytes/1024/1024/1024)
  } else if (bytes > 1024*1024) {
    return sprintf("%d MiB", bytes/1024/1024)
  } else if (bytes > 1024) {

```



```

return sprintf("%d KiB", bytes/1024)
} else {
return sprintf("%d B", bytes)
}
}

probe timer.s(1) {
foreach([p,e] in total_io- limit 10)
printf("%8d %15s r: %12s w: %12s\n",
p, e, humanreadable(reads[p,e]),
humanreadable(writes[p,e]))
printf("\n")
# Note we don't zero out reads, writes and total_io,
# so the values are cumulative since the script started.
}

```

`traceio.stp` は、I/O トラフィックに応じた上位 10 位の実行可能ファイルをプリントします。さらに、これらの上位 10 位の実行可能ファイルによる I/O 読み取りおよび書き込みの累積分量も追跡します。この情報は降順で 1 秒ごとに追跡、プリントアウトされます。

`traceio.stp` では、「ディスク読み取り/書き込みトラフィックの要約」の `disktop.stp` でも使用されているローカル変数 `$return` が使用されていることに注意してください。

#### 例4.7 `traceio.stp` のサンプル出力

```

[...]
Xorg r: 583401 KiB w: 0 KiB
floaters r: 96 KiB w: 7130 KiB
multiloader-apple r: 538 KiB w: 537 KiB
sshd r: 71 KiB w: 72 KiB
pam_timestamp_c r: 138 KiB w: 0 KiB
staprun r: 51 KiB w: 51 KiB
snmpd r: 46 KiB w: 0 KiB
pcscd r: 28 KiB w: 0 KiB
irqbalance r: 27 KiB w: 4 KiB
cupsd r: 4 KiB w: 18 KiB

Xorg r: 588140 KiB w: 0 KiB
floaters r: 97 KiB w: 7143 KiB
multiloader-apple r: 543 KiB w: 542 KiB
sshd r: 72 KiB w: 72 KiB
pam_timestamp_c r: 138 KiB w: 0 KiB
staprun r: 51 KiB w: 51 KiB
snmpd r: 46 KiB w: 0 KiB
pcscd r: 28 KiB w: 0 KiB
irqbalance r: 27 KiB w: 4 KiB
cupsd r: 4 KiB w: 18 KiB

```

#### 4.2.4. I/O 監視 (デバイスごと)

このセクションでは、特定のデバイス上の I/O アクティビティを監視する方法を説明します。

`traceio2.stp`

```
#!/usr/bin/env stap

global device_of_interest

probe begin {
  /* The following is not the most efficient way to do this.
     One could directly put the result of usrdev2kerndev()
     into device_of_interest. However, want to test out
     the other device functions */
  dev = usrdev2kerndev($1)
  device_of_interest = MKDEV(MAJOR(dev), MINOR(dev))
}

probe vfs.write, vfs.read
{
  if (dev == device_of_interest)
    printf ("%s(%d) %s 0x%x\n",
            execname(), pid(), probefunc(), dev)
}
```

`traceio2.stp` は、全体デバイス番号という引数1つを取ります。この番号を取得するには、**stat -c "0x%D" directory** を使用します。ここでの **directory** は、監視するデバイス上にあるものにします。

`usrdev2kerndev()` 関数は、全体デバイス番号をカーネルが理解する書式に変換します。`usrdev2kerndev()` が生成した出力は、特定デバイスのメジャー番号とマイナー番号を決定するために **MKDEV()**、**MINOR()**、および **MAJOR()** 関数とともに使用されます。

`traceio2.stp` の出力には、読み取り/書き込みを実行しているプロセスの名前と ID、実行している関数 (**vfs\_read** または **vfs\_write**)、およびカーネルデバイス番号が含まれます。

以下の例は、**stap traceio2.stp 0x805** の出力の抜粋です。ここでの **0x805** は、**/home** の全体デバイス番号です。**/home** は **/dev/sda5** に存在しており、これが監視対象のデバイスになります。

#### 例4.8 `traceio2.stp` のサンプル出力

```
[...]
synergyc(3722) vfs_read 0x800005
synergyc(3722) vfs_read 0x800005
cupsd(2889) vfs_write 0x800005
cupsd(2889) vfs_write 0x800005
cupsd(2889) vfs_write 0x800005
[...]
```

#### 4.2.5. ファイルの読み取りおよび書き込みの監視

このセクションでは、ファイルの読み取りおよび書き込みをリアルタイムで監視する方法を説明します。

##### `inodewatch.stp`

```
#!/usr/bin/env stap
```

```

probe vfs.write, vfs.read
{
  # dev and ino are defined by vfs.write and vfs.read
  if (dev == MKDEV($1,$2) # major/minor device
      && ino == $3)
    printf ("%s(%d) %s 0x%x/%u\n",
            execname(), pid(), probefunc(), dev, ino)
}

```

`inodewatch.stp` は、ファイルの以下の情報をコマンドラインの引数として取ります。

- ファイルのメジャーデバイス番号
- ファイルのマイナーデバイス番号
- ファイルの **inode** 番号

この情報を取得するには、`stat -c '%D %i' filename` を使用します。ここで、**filename** は絶対パスになります。

たとえば、`/etc/crontab` を監視する場合は、最初に `stat -c '%D %i' /etc/crontab` を実行します。これにより、以下の出力が表示されます。

```
805 1078319
```

**805** は、ベース 16 (16 進数) のデバイス番号です。最後の 2 桁がマイナーデバイス番号で、その上の数字がメジャー番号です。**1078319** は **inode** 番号になります。`/etc/crontab` の監視を開始するには、`stap inodewatch.stp 0x8 0x05 1078319` を実行します (**0x** 接頭辞は、ベース 16 の値を示します)。

このコマンドの出力には、読み取り/書き込みを実行しているプロセスの名前と ID、実行している関数 (`vfs_read` または `vfs_write`)、デバイス番号 (16 進数形式)、および **inode** 番号が含まれます。例 4.9 「`inodewatch.stp` のサンプル出力」は、`stap inodewatch.stp 0x8 0x05 1078319` の出力になります (スクリプトの実行中に `cat /etc/crontab` を実行した場合)。

#### 例4.9 `inodewatch.stp` のサンプル出力

```

cat(16437) vfs_read 0x800005/1078319
cat(16437) vfs_read 0x800005/1078319

```

### 4.2.6. ファイル属性の変更の監視

このセクションでは、プロセスがターゲットファイルの属性を変更したかどうかをリアルタイムで監視する方法を説明します。

#### `inodewatch2-simple.stp`

```

global ATTR_MODE = 1

probe kernel.function("inode_setattr") {
  dev_nr = $inode->i_sb->s_dev
  inode_nr = $inode->i_ino
}

```

```

if (dev_nr == ($1 << 20 | $2) # major/minor device
    && inode_nr == $3
    && $attr->ia_valid & ATTR_MODE)
printf ("%s(%d) %s 0x%x/%u %o %d\n",
    execname(), pid(), probefunc(), dev_nr, inode_nr, $attr->ia_mode, uid())
}

```

「ファイルの読み取りおよび書き込みの監視」の `inodewatch.stp` のように、`inodewatch2-simple.stp` はターゲットファイルのデバイス番号 (整数形式) と **inode** 番号を引数として取ります。この情報の取得方法は、「ファイルの読み取りおよび書き込みの監視」を参照してください。

`inodewatch2-simple.stp` の出力は `inodewatch.stp` の出力に似ていますが、`inodewatch2-simple.stp` には監視対象ファイルの属性変更と、変更を行ったユーザー ID (`uid()`) が含まれている点で異なります。例4.10「`inodewatch2-simple.stp` のサンプル出力」は `inodewatch2-simple.stp` の出力になります。`/home/joe/bigfile` の監視中に、ユーザー `joe` が `chmod 777 /home/joe/bigfile` と `chmod 666 /home/joe/bigfile` を実行しています。

#### 例4.10 `inodewatch2-simple.stp` のサンプル出力

```

chmod(17448) inode_setattr 0x800005/6011835 100777 500
chmod(17449) inode_setattr 0x800005/6011835 100666 500

```

## 4.3. プロファイリング

以下のセクションでは、関数呼び出しを監視することでカーネルアクティビティのプロファイルを実行するスクリプトを説明します。

### 4.3.1. 関数呼び出しのカウント

このセクションでは、30 秒間のサンプルで特定のカーネル関数をシステムが呼び出した回数を特定する方法を説明します。ワイルドカードの使用によっては、このスクリプトを使用して複数のカーネル関数を対象とすることもできます。

#### `functioncallcount.stp`

```

#!/usr/bin/env stap
# The following line command will probe all the functions
# in kernel's memory management code:
#
# stap functioncallcount.stp "*"@mm/*.c"

probe kernel.function(@1).call { # probe functions listed on commandline
  called[probecfunc()] <<< 1 # add a count efficiently
}

global called

probe end {
  foreach (fn in called-) # Sort by call count (in decreasing order)
  # (fn+ in called) # Sort by function name
  printf("%s %d\n", fn, @count(called[fn]))
}

```

```

exit()
}

```

`functioncallcount.stp` は、ターゲットカーネルの関数を引数として取ります。この引数はワイルドカードに対応しているので、ある程度までの複数のカーネル関数をターゲットにできます。

`functioncallcount.stp` の出力には、サンプル時間中に呼び出された関数の名前とその回数がアルファベット順に表示されます。例4.11「`functioncallcount.stp` のサンプル出力」は、`stap functioncallcount.stp "**@mm/*.c"` の出力抜粋になります。

#### 例4.11 `functioncallcount.stp` のサンプル出力

```

[...]
__vma_link 97
__vma_link_file 66
__vma_link_list 97
__vma_link_rb 97
__xchg 103
add_page_to_active_list 102
add_page_to_inactive_list 19
add_to_page_cache 19
add_to_page_cache_lru 7
all_vm_events 6
alloc_pages_node 4630
alloc_slabmgmt 67
anon_vma_alloc 62
anon_vma_free 62
anon_vma_lock 66
anon_vma_prepare 98
anon_vma_unlink 97
anon_vma_unlock 66
arch_get_unmapped_area_topdown 94
arch_get_unmapped_exec_area 3
arch_unmap_area_topdown 97
atomic_add 2
atomic_add_negative 97
atomic_dec_and_test 5153
atomic_inc 470
atomic_inc_and_test 1
[...]

```

### 4.3.2. 呼び出し先の追跡

このセクションでは、着信および発信関数呼び出しを追跡する方法を説明します。

#### `para-callgraph.stp`

```

#!/usr/bin/env stap

function trace(entry_p, extra) {
    %($# > 1 %? if (tid() in trace) %)
    printf("%s%s%s %s\n",

```

```

    thread_indent (entry_p),
    (entry_p>0?"->":"<-"),
    probefunc (),
    extra)
}

%( $# > 1 %?
global trace
probe $2.call {
  trace[tid()] = 1
}
probe $2.return {
  delete trace[tid()]
}
%)

probe $1.call { trace(1, $$parms) }
probe $1.return { trace(-1, $$return) }

```

`para-callgraph.stp` は、以下の 2 つのコマンドライン引数を取っています。

- その開始または終了が追跡対象となっている関数 (**\$1**)。
- 2 つ目のオプションとなる *trigger function* (**\$2**)。これは、スレッドごとの追跡を有効、無効にします。trigger function が終了していなければ、各スレッドにおける追跡は継続されます。

`para-callgraph.stp` では `thread_indent()` を使用しているので、その出力には、**\$1** (追跡しているプロンプ関数) のタイムスタンプ、プロセス名、およびスレッド ID が含まれます。`thread_indent()` の詳細は、[SystemTap 関数](#) のエントリーを参照してください。

以下は、`stap para-callgraph.stp 'kernel.function("*@fs/*.c") 'kernel.function("sys_read")'` の出力抜粋例になります。

#### 例4.12 `para-callgraph.stp` のサンプル出力

```

[...]
267 gnome-terminal(2921): <-do_sync_read return=0xffffffffffff5
269 gnome-terminal(2921):<-vfs_read return=0xffffffffffff5
0 gnome-terminal(2921):->fput file=0xffff880111ebbc0
2 gnome-terminal(2921):<-fput
0 gnome-terminal(2921):->fget_light fd=0x3 fput_needed=0xffff88010544df54
3 gnome-terminal(2921):<-fget_light return=0xffff8801116ce980
0 gnome-terminal(2921):->vfs_read file=0xffff8801116ce980 buf=0xc86504 count=0x1000
pos=0xffff88010544df48
4 gnome-terminal(2921): ->rw_verify_area read_write=0x0 file=0xffff8801116ce980
ppos=0xffff88010544df48 count=0x1000
7 gnome-terminal(2921): <-rw_verify_area return=0x1000
12 gnome-terminal(2921): ->do_sync_read filp=0xffff8801116ce980 buf=0xc86504 len=0x1000
ppos=0xffff88010544df48
15 gnome-terminal(2921): <-do_sync_read return=0xffffffffffff5
18 gnome-terminal(2921):<-vfs_read return=0xffffffffffff5
0 gnome-terminal(2921):->fput file=0xffff8801116ce980

```

### 4.3.3. カーネルおよびユーザースペースで費やした時間の判定

このセクションでは、スレッドがカーネルまたはユーザースペースで費やした時間を判定する方法を説明します。

#### thread-times.stp

```
#!/usr/bin/env stap

probe perf.sw.cpu_clock!, timer.profile {
  // NB: To avoid contention on SMP machines, no global scalars/arrays used,
  // only contention-free statistics aggregates.
  tid=tid(); e=execname()
  if (!user_mode())
    kticks[e,tid] <<< 1
  else
    uticks[e,tid] <<< 1
  ticks <<< 1
  tids[e,tid] <<< 1
}

global uticks, kticks, ticks

global tids

probe timer.s(5), end {
  allticks = @count(ticks)
  printf ("%16s %5s %7s %7s (of %d ticks)\n",
    "comm", "tid", "%user", "%kernel", allticks)
  foreach ([e,tid] in tids- limit 20) {
    uscaled = @count(uticks[e,tid])*10000/allticks
    kscaled = @count(kticks[e,tid])*10000/allticks
    printf ("%16s %5d %3d.%02d%% %3d.%02d%%\n",
      e, tid, uscaled/100, uscaled%100, kscaled/100, kscaled%100)
  }
  printf("\n")

  delete uticks
  delete kticks
  delete ticks
  delete tids
}
```

`thread-times.stp` は、5 秒間のサンプル内で CPU 時間を占めている上位 20 位のプロセスと、CPU ティックの総数を一覧表示します。このスクリプトの出力は、各プロセスが使用した CPU 時間のパーセント表示と、その時間がカーネルスペースかユーザースペースで費やされたかも示します。

例4.13 「`thread-times.stp` のサンプル出力」 は、`thread-times.stp` の 5 秒間サンプルの出力です。

#### 例4.13 `thread-times.stp` のサンプル出力

```
tid  %user %kernel (of 20002 ticks)
0   0.00% 87.88%
32169 5.24% 0.03%
```

```

9815 3.33% 0.36%
9859 0.95% 0.00%
3611 0.56% 0.12%
9861 0.62% 0.01%
11106 0.37% 0.02%
32167 0.08% 0.08%
3897 0.01% 0.08%
3800 0.03% 0.00%
2886 0.02% 0.00%
3243 0.00% 0.01%
3862 0.01% 0.00%
3782 0.00% 0.00%
21767 0.00% 0.00%
2522 0.00% 0.00%
3883 0.00% 0.00%
3775 0.00% 0.00%
3943 0.00% 0.00%
3873 0.00% 0.00%

```

#### 4.3.4. ポーリングアプリケーションの監視

このセクションでは、どのアプリケーションがポーリングを行なっているかを特定、監視する方法を説明します。これを行うことで、不必要または過剰なポーリングが追跡でき、CPUの使用量を改善し、省電力効果を向上させるエリアを特定できるようになります。

##### timeout.stp

```

#!/usr/bin/env stap
# Copyright (C) 2009 Red Hat, Inc.
# Written by Ulrich Drepper <drepper@redhat.com>
# Modified by William Cohen <wcohen@redhat.com>

global process, timeout_count, to
global poll_timeout, epoll_timeout, select_timeout, itimer_timeout
global nanosleep_timeout, futex_timeout, signal_timeout

probe syscall.poll, syscall.epoll_wait {
  if (timeout) to[pid()]=timeout
}

probe syscall.poll.return {
  p = pid()
  if ($return == 0 && to[p] > 0 ) {
    poll_timeout[p]++
    timeout_count[p]++
    process[p] = execname()
    delete to[p]
  }
}

probe syscall.epoll_wait.return {
  p = pid()
  if ($return == 0 && to[p] > 0 ) {
    epoll_timeout[p]++

```



```
    timeout_count[p]++
    process[p] = execname()
    delete to[p]
  }
}

probe syscall.select.return {
  if ($return == 0) {
    p = pid()
    select_timeout[p]++
    timeout_count[p]++
    process[p] = execname()
  }
}

probe syscall.futex.return {
  if (errno_str($return) == "ETIMEDOUT") {
    p = pid()
    futex_timeout[p]++
    timeout_count[p]++
    process[p] = execname()
  }
}

probe syscall.nanosleep.return {
  if ($return == 0) {
    p = pid()
    nanosleep_timeout[p]++
    timeout_count[p]++
    process[p] = execname()
  }
}

probe kernel.function("it_real_fn") {
  p = pid()
  itimer_timeout[p]++
  timeout_count[p]++
  process[p] = execname()
}

probe syscall.rt_sigtimedwait.return {
  if (errno_str($return) == "EAGAIN") {
    p = pid()
    signal_timeout[p]++
    timeout_count[p]++
    process[p] = execname()
  }
}

probe syscall.exit {
  p = pid()
  if (p in process) {
    delete process[p]
    delete timeout_count[p]
    delete poll_timeout[p]
    delete epoll_timeout[p]
  }
}
```

```

delete select_timeout[p]
delete itimer_timeout[p]
delete futex_timeout[p]
delete nanosleep_timeout[p]
delete signal_timeout[p]
}
}

probe timer.s(1) {
ansi_clear_screen()
printf (" pid | poll select  epoll itimer  futex nanosle  signal| process\n")
foreach (p in timeout_count- limit 20) {
printf ("%5d |%7d %7d %7d %7d %7d %7d %7d | %-.38s\n", p,
poll_timeout[p], select_timeout[p],
epoll_timeout[p], itimer_timeout[p],
futex_timeout[p], nanosleep_timeout[p],
signal_timeout[p], process[p])
}
}
}

```

`timeout.stp` は、各アプリケーションが以下のシステムコールを使用した回数を追跡します。

- `poll`
- `select`
- `epoll`
- `itimer`
- `futex`
- `nanosleep`
- `signal`

いくつかのアプリケーションでは、これらのシステムコールは過剰に使用されています。このため、これらは通常、ポーリングアプリケーションの「有力な容疑者」として特定されます。ただし、アプリケーションが異なるシステムコールを使用して過剰にポーリングしている場合もあります。場合によっては、システムが使用している上位のシステムコールを見つけることが便利なこともあります(手順は「[最もよく使われるシステムコールの追跡](#)」を参照してください)。これを行うことで新たな容疑者を特定でき、`timeout.stp` に追加して追跡することができます。

#### 例4.14 `timeout.stp` のサンプル出力

```

uid | poll select  epoll itimer  futex nanosle  signal| process
28937 | 148793  0  0  4727  37288  0  0| firefox
22945 |  0 56949  0  1  0  0  0| scim-bridge
  0 |  0  0  0 36414  0  0  0| swapper
4275 | 23140  0  0  1  0  0  0| mixer_applet2
4191 |  0 14405  0  0  0  0  0| scim-launcher
22941 | 7908  1  0  62  0  0  0| gnome-terminal
4261 |  0  0  0  2  0  7622  0| escd
3695 |  0  0  0  0  0  7622  0| gdm-binary
3483 |  0 7206  0  0  0  0  0| dhcdbd

```

```

4189 | 6916  0  0  2  0  0  0| scim-panel-gtk
1863 | 5767  0  0  0  0  0  0| iscsid
2562 |  0 2881  0  1  0 1438  0| pcscd
4257 | 4255  0  0  1  0  0  0| gnome-power-man
4278 | 3876  0  0  60  0  0  0| multiload-apple
4083 |  0 1331  0 1728  0  0  0| Xorg
3921 | 1603  0  0  0  0  0  0| gam_server
4248 | 1591  0  0  0  0  0  0| nm-applet
3165 |  0 1441  0  0  0  0  0| xterm
29548|  0 1440  0  0  0  0  0| httpd
1862 |  0  0  0  0  0  0 1438  0| iscsid

```

2つ目のプローブのタイマー (`timer.s()`) を編集すると、サンプル時間を長くすることができます。`functioncallcount.stp` の出力には、上位 20 位のポーリングアプリケーションの名前と UID、各アプリケーションがシステムコールのポーリングを行った回数が含まれています。例4.14「`timeout.stp` のサンプル出力」はスクリプトの抜粋になります。

#### 4.3.5. 最もよく使われるシステムコールの追跡

「[ポーリングアプリケーションの監視](#)」の `timeout.stp` では、以下のシステムコールを最もよく使用したものを挙げることで、どのアプリケーションがポーリングを行なっているかを特定します。

- `poll`
- `select`
- `epoll`
- `itimer`
- `futex`
- `nanosleep`
- `signal`

ただしシステムによっては、別のシステムコールが過剰なポーリングを行なっている可能性もあります。ポーリングしているアプリケーションがポーリングに異なるシステムコールを使用していることが疑われる場合は、まずシステムが使用している上位のシステムコールを特定する必要があります。これには、`topsys.stp` を使用します。

##### `topsys.stp`

```

#!/usr/bin/env stap
#
# This script continuously lists the top 20 systemcalls in the interval
# 5 seconds
#
global syscalls_count

probe syscall.* {
  syscalls_count[name]++
}

```

```

function print_systop () {
  printf ("%25s %10s\n", "SYSCALL", "COUNT")
  foreach (syscall in syscalls_count- limit 20) {
    printf ("%25s %10d\n", syscall, syscalls_count[syscall])
  }
  delete syscalls_count
}

probe timer.s(5) {
  print_systop ()
  printf("-----\n")
}

```

`topsys.stp` は、5 秒ごとにシステムが使用している上位 20 位のシステムコールを一覧表示します。また、同期間に各システムコールが使用された回数も表示されます。例4.15「`topsys.stp` のサンプル出力」がサンプル出力になります。

#### 例4.15 `topsys.stp` のサンプル出力

```

-----
      SYSCALL   COUNT
gettimeofday   1857
      read     1821
      ioctl    1568
      poll     1033
      close    638
      open     503
      select   455
      write    391
      writev   335
      futex    303
      recvmsg  251
      socket   137
clock_gettime   124
rt_sigprocmask 121
      sendto   120
      setitimer 106
      stat     90
      time     81
      sigreturn 72
      fstat    66
-----

```

#### 4.3.6. プロセスごとのシステムコールボリュームの追跡

このセクションでは、最もボリュームの大きいシステムコールを実行しているプロセスを判定する方法を説明します。ここまでのセクションでは、システムが使用している上位のシステムコールを監視する方法を説明しました（「最もよく使われるシステムコールの追跡」）。また、どのアプリケーションが最も多く特定の「ポーリング容疑者」のシステムコールを使用しているかを特定する方法も説明しました（「ポーリングアプリケーションの監視」）。プロセスごとのシステムコールのボリュームを監視することで、ポーリングプロセスおよび他のリソースの大量消費についてシステムを調査する際により多くのデータが提供できます。

## syscalls\_by\_proc.stp

```

#!/usr/bin/env stap

# Copyright (C) 2006 IBM Corp.
#
# This file is part of systemtap, and is free software. You can
# redistribute it and/or modify it under the terms of the GNU General
# Public License (GPL); either version 2, or (at your option) any
# later version.

#
# Print the system call count by process name in descending order.
#

global syscalls

probe begin {
  print ("Collecting data... Type Ctrl-C to exit and display results\n")
}

probe syscall.* {
  syscalls[execname()]++
}

probe end {
  printf ("%s\n", "#SysCalls", "Process Name")
  foreach (proc in syscalls-)
    printf ("%d %s\n", syscalls[proc], proc)
}

```

`syscalls_by_proc.stp` は、システムコールを多く実行している上位 20 位のプロセスを一覧表示します。また、一定期間内に各プロセスが実行したシステムコールの数も表示されます。例4.16「`topsys.stp` のサンプル出力」がサンプル出力になります。

例4.16 `topsys.stp` のサンプル出力

```

Collecting data... Type Ctrl-C to exit and display results
#SysCalls Process Name
1577    multiload-apple
692     synergyc
408     pcscd
376     mixer_applet2
299     gnome-terminal
293     Xorg
206     scim-panel-gtk
95      gnome-power-man
90      artsd
85      dhcdbd
84      scim-bridge
78      gnome-screensav
66      scim-launcher
[...]

```

プロセス名の代わりにプロセス ID を表示するには、以下のスクリプトを使用します。

#### syscalls\_by\_pid.stp

```
#!/usr/bin/env stap

# Copyright (C) 2006 IBM Corp.
#
# This file is part of systemtap, and is free software. You can
# redistribute it and/or modify it under the terms of the GNU General
# Public License (GPL); either version 2, or (at your option) any
# later version.

#
# Print the system call count by process ID in descending order.
#

global syscalls

probe begin {
  print ("Collecting data... Type Ctrl-C to exit and display results\n")
}

probe syscall.* {
  syscalls[pid()]++
}

probe end {
  printf ("%s\n", "#SysCalls", "PID")
  foreach (pid in syscalls-)
    printf ("%d %d\n", syscalls[pid], pid)
}
```

出力で表示されているように、結果を表示するにはスクリプトを手動で終了する必要があります。どちらのスクリプトにも **timer.s()** プローブを追加して、一定時間後に終了させることができます。たとえば、5 秒後にスクリプトを終了させるには、以下のプローブをスクリプトに追加します。

```
probe timer.s(5)
{
  exit()
}
```

## 4.4. 競合ユーザースペースのロックの特定

このセクションでは、特定期間におけるシステムを通じた競合ユーザースペースのロックを特定する方法を説明します。競合ユーザースペースのロックの特定は、**futex** 競合が原因と疑われるハングの調査に役立ちます。

簡単に説明すると、複数のプロセスがメモリーの同じ領域にアクセスしようとする、**futex** 競合が発生します。場合によっては、これは競合しているプロセス間のデッドロックになり、アプリケーションがハングしているように見えます。

これを特定するために、**futexes.stp** では **futex** システムコールをプローブします。

## futexes.stp

```

#!/usr/bin/env stap

# This script tries to identify contended user-space locks by hooking
# into the futex system call.

global thread_thislock # short
global thread_blocktime #
global FUTEX_WAIT = 0 /*, FUTEX_WAKE = 1 */

global lock_waits # long-lived stats on (tid,lock) blockage elapsed time
global process_names # long-lived pid-to-execname mapping

probe syscall.futex {
  if (op != FUTEX_WAIT) next # don't care about WAKE event originator
  t = tid ()
  process_names[pid()] = execname()
  thread_thislock[t] = $uaddr
  thread_blocktime[t] = gettimeofday_us()
}

probe syscall.futex.return {
  t = tid()
  ts = thread_blocktime[t]
  if (ts) {
    elapsed = gettimeofday_us() - ts
    lock_waits[pid(), thread_thislock[t]] <<< elapsed
    delete thread_blocktime[t]
    delete thread_thislock[t]
  }
}

probe end {
  foreach ([pid+, lock] in lock_waits)
    printf ("%s[%d] lock %p contended %d times, %d avg us\n",
            process_names[pid], pid, lock, @count(lock_waits[pid,lock]),
            @avg(lock_waits[pid,lock]))
}

```

[futexes.stp](#) は手動で停止する必要があります。終了時には以下の情報が表示されます。

- 競合の原因となったプロセスの名前と ID
- 競合対象となったメモリー領域
- メモリー領域が競合された回数
- プローブ中の競合の平均時間

例4.17 「[futexes.stp のサンプル出力](#)」 は [futexes.stp](#) を終了した際 (約 20 秒後) の出力の抜粋です。

例4.17 [futexes.stp](#) のサンプル出力

[...]

automount[2825] lock 0x00bc7784 contended 18 times, 999931 avg us

synergyc[3686] lock 0x0861e96c contended 192 times, 101991 avg us

synergyc[3758] lock 0x08d98744 contended 192 times, 101990 avg us

synergyc[3938] lock 0x0982a8b4 contended 192 times, 101997 avg us

[...]



## 第5章 SYSTEMTAP のエラーを理解する

本章では、SystemTap の使用中に発生する可能性のある最も一般的なエラーについて説明しています。

### 5.1. 解析エラーとセマンティックエラー

このタイプのエラーは、スクリプトがカーネルモジュールに変換される前の、SystemTap により構文解析が行われ C に変換される際に発生します。たとえば、これらのタイプのエラーは、無効な値を変数やアレイに割り当てる演算から発生します。

#### parse error: expected *foo*, saw *bar*

スクリプトに文法的なエラーまたは誤字や脱字のエラーが含まれています。プローブのコンテキストから考えると、SystemTap は間違っただコンストラクトのタイプを検出しました。

以下の無効な SystemTap スクリプトにはプローブのハンドラーがありません。

```
probe vfs.read
probe vfs.write
```

その結果、パーサーは 2 行目のコラム 1 の **probe** キーワードには別の値を期待していたことを示す以下のエラーメッセージが表示されます。

```
parse error: expected one of ' , ( ? ! { = += '
saw: keyword at perror.stp:2:1
1 parse error(s).
```

#### parse error: embedded code in unprivileged script

スクリプトには危険な埋め込み C コードが含まれています (%{ %} で囲まれたコードブロック)。SystemTap では、スクリプトに C コードを埋め込むことができるため、目的にあった tapsets がない場合には便利です。ただし、埋め込みの C コンストラクトは安全ではありません。そのため、スクリプトにこのような構造を検出すると SystemTap はこのエラーで警告を發します。

スクリプト内の似たようなコンストラクトがすべて確実に安全であることが分かっていて、かつ **stapdev** グループのメンバーである (または root 特権がある) 場合には、オプション **-g (stap -g script)** を使用して「guru」モードでスクリプトを実行します。

#### semantic error: type mismatch for identifier *foo*' ... string vs. long

スクリプト内の **foo** 関数は、間違っただタイプ (%s または %d) を使用していました。このエラーを [例 5.1 \[error-variable.stp\]](#) に示します。関数 **execname()** が文字列を返すため、書式指定子は %d ではなく %s にしてください。

#### 例5.1 error-variable.stp

```
probe syscall.open
{
    printf ("%d(%d) open\n", execname(), pid())
}
```

#### semantic error: unresolved type for identifier *foo*'

識別子 (変数など) が使用されているがタイプ (整数または文字列) を特定できません。たとえば、`printf` ステートメントに変数を使用しているのに、スクリプトでは変数に値を割り当てていない場合などにこのエラーが発生します。

#### semantic error: Expecting symbol or array index expression

SystemTap ではアレイ内の場所や変数に値を割り当てることはできません。割り当ての目的地が有効な目的地になっていません。以下のようなサンプルコードを使用すると、このエラーが発生することになります。

```
probe begin { printf("x") = 1 }
```

#### semantic error: unresolved function call (変数N関数の検索中に発生するエラー)

スクリプト内の関数呼び出しまたはアレイインデックス式が無効な数の引数/パラメーターを使用しました。SystemTap では、*arity* (アリティ) はアレイのインデックス数または関数へのパラメーター数を参照することができます。

#### semantic error: array locals not supported, missing global declaration?

アレイをグローバル変数と宣言せずに、スクリプトがアレイ演算を使用しました (グローバル変数は、SystemTap スクリプトでの使用後に宣言されることができます)。アレイが一貫性のないアリティと使用されると、同様のメッセージが表示されます。

#### semantic error: variable 'foo' modified during 'foreach' iteration

アレイ **foo** がアクティブな **foreach** ループ内で修正 (割り当てまたは削除) されています。このエラーは、スクリプト内の演算が **foreach** ループ内で関数呼び出しを実行しても表示されます。

#### semantic error: probe point mismatch at positionN, while resolving probe pointfoo

SystemTap は、イベントまたは SystemTap 関数 **foo** が参照しているものを理解できませんでした。これは通常、SystemTap が tapset ライブラリー内で **foo** に一致するものを見つけられなかったことを意味します。ここでの *N* は、エラーの行とコラムを指します。

#### semantic error: no match for probe point, while resolving probe pointfoo

イベント/ハンドラー関数 **foo** は、さまざまな理由で解決できませんでした。このエラーは、スクリプトにイベント **kernel.function("blah")** が含まれ、**blah** が存在しない場合に発生します。このエラーはスクリプトに無効なカーネルファイル名やソース行数が含まれていることを意味する場合があります。

#### semantic error: unresolved target-symbol expression

スクリプトのハンドラーはターゲット変数を参照しますが、変数の値が解決できませんでした。このエラーは、ハンドラーがターゲット変数を参照した際にその変数がコンテキスト内で無効であることも意味します。これは、生成されたコードのコンパイラ最適化の結果である可能性もあります。

#### semantic error: libdwfl failure

デバッグ情報の処理に問題がありました。ほとんどの場合、このエラーは、バージョンがプロブされたカーネルに完全に一致しない **kernel-debuginfo** RPM のインストールにより発生します。インストールされた **kernel-debuginfo** RPM 自体に一貫性や正確性の問題がある可能性があります。

#### semantic error: cannot findfoo debuginfo

SystemTap は適切な **kernel-debuginfo** を見つけることができませんでした。

## 5.2. ランタイムエラーおよび警告

ランタイムエラーと警告は、SystemTap インストールメンテーションがインストールされ、システム上でデータを収集すると発生します。

**WARNING: Number of errors:  $N$ , skipped probes:  $M$**

この実行時にエラーが発生し、プローブが省略されました。 $N$ と $M$ の両方は、ある期間にイベントハンドラーを実行する時間が足りなかったためなどの理由で実行されなかったプローブ数です。

**division by 0**

スクリプトコードが無効な除算を実行しました。

**aggregate element not found**

値が累積されていない集合で、**@count** 以外の統計情報抽出関数が呼び出されました。これは、division by zero と類似したものです。

**aggregation overflow**

現時点で、集計値を含むアレイに明確な鍵のペアが過剰にあります。

**MAXNESTING exceeded**

関数呼び出しの入れ子の試行が多すぎます。許可されるデフォルトの関数呼び出しの入れ子は 10 です。

**MAXACTION exceeded**

プローブハンドラーによるステートメントの実行の試みが多すぎました。プローブハンドラー内で許可されるデフォルトのアクション数は 1000 です。

**kernel/user string copy fault atADDR**

プローブハンドラーが、無効なアドレス (*ADDR*) でカーネルまたはユーザースペースから文字列をコピーしようとして失敗しました。

**pointer dereference fault**

ターゲット変数の評価など、ポインターの逆参照演算中に障害が発生しました。

## 第6章 リファレンス

本章では、SystemTap についてさらに詳細な記載をしている他の参考文献を紹介しています。高度なプローブや tapset を記述する際の参考にしてください。

### SystemTap Wiki

『SystemTap Wiki』とは、SystemTap のデプロイメント、使用、および開発に関連するリンクや記事がまとめられているページです。 <http://sourceware.org/systemtap/wiki/HomePage> をご覧ください。

### SystemTap チュートリアル

本ガイドのコンテンツの多くは『SystemTap Tutorial』を元としています。『SystemTap Tutorial』は、C++ やカーネル開発などに関して中から高程度の知識を有するユーザー向けの内容になります。 <http://sourceware.org/systemtap/tutorial/> をご覧ください。

### man stapprobes

man ページの **stapprobes** では、SystemTap で対応している各種のプローブポイントおよび SystemTap tapset ライブラリーで定義している他のエイリアスを紹介しています。man ページの下部には、特定のシステムコンポーネントの同様のプローブポイントを列挙する他の man ページの一覧が記載されています (**stapprobes.scsi**、**stapprobes.kprocess**、**stapprobes.signal** など)。

### man stapfuncs

man ページの **stapfuncs** は、SystemTap tapset ライブラリーで対応している多数の関数および各関数に規定の構文を説明しています。ただし、対応している関数が **すべて** 記載されているわけではない点に注意してください。説明がない関数が他にもあります。

### SystemTap Language Reference (SystemTap 言語リファレンス)

本ガイドは、SystemTap の言語構成および構文がすべて記載されているリファレンスです。C++ および他の似たようなプログラミング言語の基本的な知識または中程度の知識をお持ちのユーザーを対象としています。『SystemTap Language Reference』は、すべてのユーザーを対象に <http://sourceware.org/systemtap/langref/> で公開されています。

### Tapset Developers Guide

SystemTap スクリプトの記述に熟達したら、独自のタップセットの記述に挑戦することができます。『Tapset Developers Guide』では、タップセットライブラリーに関数を追加する方法を説明しています。

### Test Suite

**systemtap-testsuite** パッケージを使用すると、ソースから構築しなくても SystemTap ツールチェーン全体の検証を行うことができます。また、研究や検証対象となる SystemTap スクリプトのサンプルも数多く収納されています。一部のサンプルスクリプトは、「[4章 便利な SystemTap スクリプト](#)」で説明しています。

デフォルトでは、**systemtap-testsuite** に含まれるサンプルスクリプトは、`/usr/share/systemtap/testsuite/systemtap.examples` にあります。

## 付録A 改訂履歴

改訂 6-9.2 6.9 GA リリース向けバージョン	Mon 3 April 2017	Robert Krátký
改訂 2-32 修正を含む非同期リリース。	Sun Jun 26 2016	Mirek Jahoda
改訂 2-30 6.8 GA リリース向けバージョン	Wed May 3 2016	Robert Krátký
改訂 2-23 6.4 GA リリース向けバージョン	Tue Feb 17 2013	Jacquelynn East
改訂 2-22 BZ#598956	Tue May 17 2011	Jacquelynn East
改訂 2-18 BZ#698956	Mon May 16 2011	Jacquelynn East
改訂 2-15 Initialized	Mon Nov 29 2010	Michael Hideo-Smith
改訂 2-14 若干の編集	Thu Nov 18 2010	Jacquelynn East
改訂 2-13 若干の編集	Thu Nov 18 2010	Jacquelynn East

## 索引

### シンボル

#### \$count

##### サンプル使用例

ローカル変数, [ファイル読み取り/書き込みの I/O 時間の追跡](#)

#### \$return

##### サンプル使用例

ローカル変数, [ディスク読み取り/書き込みトラフィックの要約](#), [累積 I/O の追跡](#)

#### 'foreach' 中に変数が修正される

##### 解析エラー/セマンティックエラー

SystemTap のエラーを理解する, [解析エラーとセマンティックエラー](#)

#### @avg (整数抽出)

##### 統計集計の計算

アレイ演算, [統計集計の計算](#)

#### @count (整数抽出)

##### 統計集計の計算

アレイ演算, [統計集計の計算](#)

#### @max (整数抽出)

##### 統計集計の計算

アレイ演算, [統計集計の計算](#)

#### @min (整数抽出)

##### 統計集計の計算

アレイ演算, [統計集計の計算](#)

#### @sum (整数抽出)

##### 統計集計の計算

アレイ演算, [統計集計の計算](#)

### はじめに

SystemTap の機能, [SystemTap の機能](#)

パフォーマンスの監視, [はじめに](#)

本ガイドの目的, [本ガイドの目的](#)

目的、ドキュメント, [本ガイドの目的](#)

---

## アグリゲーションのオーバーフロー

### ランタイムエラー/警告

SystemTap のエラーを理解する, [ランタイムエラーおよび警告](#)

## アレイ, [連想アレイ](#)

(参照 [連想アレイ](#))

## アレイ/アレイ要素の消去

アレイ演算, [アレイおよびアレイ要素の消去/削除](#)

Delete 演算子, [アレイおよびアレイ要素の消去/削除](#)

仮想ファイルシステムの読み込み (累積的以外)、集計, [アレイおよびアレイ要素の消去/削除](#)

同一プローブ内の複数のアレイ演算, [アレイおよびアレイ要素の消去/削除](#)

## アレイからの値の読み取り

アレイ演算, [アレイからの値の読み取り](#)

シンプルな計算でのアレイの使用, [アレイからの値の読み取り](#)

空の一意の鍵, [アレイからの値の読み取り](#)

## タイムスタンプ差分の計算

アレイ演算, [アレイからの値の読み取り](#)

## アレイを使った代数公式

アレイからの値の読み取り

アレイ演算, [アレイからの値の読み取り](#)

## アレイメンバーシップのテスト

条件付きステートメント、アレイの使用

アレイ演算, [条件付きステートメントにおけるアレイの使用](#)

## アレイローカルがサポートされない

解析エラー/セマンティックエラー

SystemTap のエラーを理解する, [解析エラーとセマンティックエラー](#)

## アレイ内での複数要素の処理

foreach

アレイ演算, [アレイ内での複数要素の処理](#)

foreach 出力の制限

アレイ演算, [アレイ内での複数要素の処理](#)

foreach 出力の順序

アレイ演算, [アレイ内での複数要素の処理](#)

アレイ演算, [アレイ内での複数要素の処理](#)  
累積的仮想ファイルシステムの読み込み、集計  
アレイ演算, [アレイ内での複数要素の処理](#)

アレイ内の複数要素

アレイ演算, [アレイ内での複数要素の処理](#)

アレイ演算

アレイ/アレイ要素の消去, [アレイおよびアレイ要素の消去/削除](#)  
Delete 演算子, [アレイおよびアレイ要素の消去/削除](#)  
仮想ファイルシステムの読み込み (累積的以外)、集計, [アレイおよびアレイ要素の消去/削除](#)  
同一プローブ内の複数のアレイ演算, [アレイおよびアレイ要素の消去/削除](#)

アレイおよびアレイ要素の削除, [アレイおよびアレイ要素の消去/削除](#)

アレイからの値の読み取り, [アレイからの値の読み取り](#)  
シンプルな計算でのアレイの使用, [アレイからの値の読み取り](#)  
タイムスタンプ差分の計算, [アレイからの値の読み取り](#)  
空の一意の鍵, [アレイからの値の読み取り](#)

アレイ内での複数要素の処理, [アレイ内での複数要素の処理](#)

foreach, [アレイ内での複数要素の処理](#)  
foreach 出力の制限, [アレイ内での複数要素の処理](#)  
foreach 出力の順序, [アレイ内での複数要素の処理](#)  
反復、アレイ内の要素の処理, [アレイ内での複数要素の処理](#)  
累積的仮想ファイルシステムの読み込み、集計, [アレイ内での複数要素の処理](#)

アレイ内の複数要素, [アレイ内での複数要素の処理](#)

条件付きステートメント、アレイの使用, [条件付きステートメントにおけるアレイの使用](#)  
アレイメンバーシップのテスト, [条件付きステートメントにおけるアレイの使用](#)

統計集計の計算, [統計集計の計算](#)

@avg (整数抽出), [統計集計の計算](#)  
@count (整数抽出), [統計集計の計算](#)  
@max (整数抽出), [統計集計の計算](#)  
@min (整数抽出), [統計集計の計算](#)  
@sum (整数抽出), [統計集計の計算](#)  
count (演算子), [統計集計の計算](#)  
統計集計が収集したデータの抽出, [統計集計の計算](#)  
統計集計への値の追加, [統計集計の計算](#)

連想アレイ, [SystemTap でのアレイ演算](#)

関連する値の割り当て, [関連する値の割り当て](#)  
タイムスタンプをプロセス名に関連付ける, [関連する値の割り当て](#)



関連する値の増加, [関連する値の増加](#)

仮想ファイルシステムの読み込みの集計 (VFS 読み込み), [関連する値の増加](#)

アーキテクチャー表記法、決定, [他のコンピューター用のインストルメンテーション生成](#)  
アーキテクチャー表記法の決定, [他のコンピューター用のインストルメンテーション生成](#)  
イベント

begin, [イベント](#)

end, [イベント](#)

kernel.function("function"), [イベント](#)

kernel.trace("tracepoint"), [イベント](#)

module("module"), [イベント](#)

syscall.system\_call, [イベント](#)

timer イベント, [イベント](#)

vfs.file\_operation, [イベント](#)

ワイルドカード, [イベント](#)

同期および非同期イベントの例, [イベント](#)

同期イベント, [イベント](#)

概要, [イベント](#)

非同期イベント, [イベント](#)

イベントおよびハンドラー, [SystemTap の作動方法](#)

イベントタイプ

SystemTap の作動方法, [SystemTap の作動方法](#)

イベントワイルドカード, [イベント](#)

イベント内のワイルドカード, [イベント](#)

インストルメンテーションモジュール

クロスインストルメンテーション, [他のコンピューター用のインストルメンテーション生成](#)

インストーラ

systemtap パッケージ, [SystemTap のインストール](#)

systemtap-runtime パッケージ, [SystemTap のインストール](#)

カーネルバージョン、判断, [必要なカーネル情報 RPM パッケージのインストール](#)

カーネル情報パッケージ, [必要なカーネル情報 RPM パッケージのインストール](#)

初期テスト, [初期テスト](#)

必要なパッケージ, [必要なカーネル情報 RPM パッケージのインストール](#)

設定とインストーラ, [インストールと設定](#)

インデックス式

概要

アレイ, [連想アレイ](#)

エラー

ランタイムエラー/警告, [ランタイムエラーおよび警告](#)

aggregate element not found, [ランタイムエラーおよび警告](#)

aggregation overflow, [ランタイムエラーおよび警告](#)

division by 0, [ランタイムエラーおよび警告](#)

MAXACTION exceeded, [ランタイムエラーおよび警告](#)

MAXNESTING exceeded, [ランタイムエラーおよび警告](#)

number of errors: N, skipped probes: M, [ランタイムエラーおよび警告](#)

pointer dereference fault, [ランタイムエラーおよび警告](#)

コピー障害, [ランタイムエラーおよび警告](#)

解析エラー/セマンティックエラー, [解析エラーとセマンティックエラー](#)

'foreach' 中に変数が修正される, [解析エラーとセマンティックエラー](#)

guru モード, [解析エラーとセマンティックエラー](#)

libdwfl 失敗, [解析エラーとセマンティックエラー](#)

スクリプトの文法的なエラー/誤字のエラー, [解析エラーとセマンティックエラー](#)

プローブの不一致, [解析エラーとセマンティックエラー](#)

プローブポイントに一致するものがない, [解析エラーとセマンティックエラー](#)

変数/アレイに無効な値, [解析エラーとセマンティックエラー](#)

特権のないスクリプトにコードが埋め込まれている, [解析エラーとセマンティックエラー](#)

解決できないターゲット記号式, [解析エラーとセマンティックエラー](#)

解決できない識別子のタイプ, [解析エラーとセマンティックエラー](#)

解決できない関数呼び出し, [解析エラーとセマンティックエラー](#)

記号やアレイに期待されているインデックス式, [解析エラーとセマンティックエラー](#)

識別子タイプの不一致, [解析エラーとセマンティックエラー](#)

非グローバルアレイ, [解析エラーとセマンティックエラー](#)

オプション、stap

用途, [SystemTap スクリプトの実行](#)

カーネル/ユーザースペースで費やした時間、判定

SystemTap スクリプトの例, [カーネルおよびユーザースペースで費やした時間の判定](#)

カーネルおよびユーザースペース、費やした時間の判定

SystemTap スクリプトの例, [カーネルおよびユーザースペースで費やした時間の判定](#)

カーネルおよびユーザースペースで費やした時間の判定

SystemTap スクリプトの例, [カーネルおよびユーザースペースで費やした時間の判定](#)

カーネルバージョン、判断, [必要なカーネル情報 RPM パッケージのインストール](#)

カーネルバージョンの判断, [必要なカーネル情報 RPM パッケージのインストール](#)

カーネル情報パッケージ, [必要なカーネル情報 RPM パッケージのインストール](#)

キーのペア

概要

## アレイ, [連想アレイ](#)

### クロスインストールメンテーション

SystemTap スクリプトからインストールメンテーションを生成する, [他のコンピューター用のインストールメンテーション生成](#)

SystemTap スクリプトからカーネルモジュールを構築する, [他のコンピューター用のインストールメンテーション生成](#)

インストールメンテーションモジュール, [他のコンピューター用のインストールメンテーション生成](#)

ターゲットカーネル, [他のコンピューター用のインストールメンテーション生成](#)

ターゲットシステム, [他のコンピューター用のインストールメンテーション生成](#)

ホストシステム, [他のコンピューター用のインストールメンテーション生成](#)

利点; [他のコンピューター用のインストールメンテーション生成](#)

#### 設定

ホストシステムおよびターゲットシステム, [他のコンピューター用のインストールメンテーション生成](#)

クロスインストールメンテーションの利点, [他のコンピューター用のインストールメンテーション生成](#)

### コピー障害

ランタイムエラー/警告

SystemTap のエラーを理解する, [ランタイムエラーおよび警告](#)

### コマンドラインの引数

SystemTap ハンドラーコンストラクト

ハンドラー, [コマンドラインの引数](#)

### コンポーネント

SystemTap スクリプト

概要, [SystemTap スクリプト](#)

### システムコール、監視

SystemTap スクリプトの例, [最もよく使われるシステムコールの追跡](#)

### システムコールの監視

SystemTap スクリプトの例, [最もよく使われるシステムコールの追跡](#)

### システムコールの監視 (プロセスごとのボリューム)

SystemTap スクリプトの例, [プロセスごとのシステムコールボリュームの追跡](#)

### システムコールボリューム (プロセスごと)、監視

SystemTap スクリプトの例, [プロセスごとのシステムコールボリュームの追跡](#)

### シンプルな計算でのアレイの使用

アレイからの値の読み取り

## アレイ演算, アレイからの値の読み取り

### スクリプト

概要, [SystemTap スクリプト](#)

イベントおよびハンドラー, [SystemTap スクリプト](#)

コンポーネント, [SystemTap スクリプト](#)

ステートメントブロック, [SystemTap スクリプト](#)

プローブ, [SystemTap スクリプト](#)

形式および構文, [SystemTap スクリプト](#)

関数, [SystemTap スクリプト](#)

### スクリプトの例

CPU ティック, [カーネルおよびユーザースペースで費やした時間の判定](#)

[ctime\(\)](#), [使用例](#), [ディスク読み取り/書き込みトラフィックの要約](#)

[futex システムコール](#), [競合ユーザースペースのロックの特定](#)

[I/O 時間の監視](#), [ファイル読み取り/書き込みの I/O 時間の追跡](#)

[if/else 条件](#), [代替の構文](#), [ネットワークのプロファイリング](#)

[inode 番号](#), [ファイルの読み取りおよび書き込みの監視](#)

[net/socket.c](#), [関数の追跡](#), [ネットワークソケットコードで呼び出された関数の追跡](#)

[stat -c](#), [ファイルデバイス番号の決定 \(整数形式\)](#), [ファイルの読み取りおよび書き込みの監視](#)

[stat -c](#), [全体デバイス番号の決定](#), [I/O 監視 \(デバイスごと\)](#)

[thread\\_indent\(\)](#), [サンプル使用例](#), [呼び出し先の追跡](#)

[timer.ms\(\)](#), [サンプル使用例](#), [関数呼び出しのカウンタ](#)

[timer.s\(\)](#), [サンプル使用例](#), [ポーリングアプリケーションの監視](#), [最もよく使われるシステムコールの追跡](#)

[trigger function](#), [呼び出し先の追跡](#)

[usrdev2kerndev\(\)](#), [I/O 監視 \(デバイスごと\)](#)

[カーネルおよびユーザースペースで費やした時間の判定](#), [カーネルおよびユーザースペースで費やした時間の判定](#)

[システムコールの監視](#), [最もよく使われるシステムコールの追跡](#)

[システムコールの監視 \(プロセスごとのボリューム\)](#), [プロセスごとのシステムコールボリュームの追跡](#)

[ディスク I/O トラフィックの要約](#), [ディスク読み取り/書き込みトラフィックの要約](#)

[デバイス I/O の監視](#), [I/O 監視 \(デバイスごと\)](#)

[ネットワークのプロファイリング](#), [ネットワークのプロファイリング](#), [カーネルでのネットワークパケットドロップの監視](#)

[ネットワークソケットコードで呼び出された関数の追跡](#), [ネットワークソケットコードで呼び出された関数の追跡](#)

[ファイルの読み取りおよび書き込みの監視](#), [ファイルの読み取りおよび書き込みの監視](#)

[ファイルデバイス番号 \(整数形式\)](#), [ファイルの読み取りおよび書き込みの監視](#)

[ファイル属性の変更の監視](#), [ファイル属性の変更の監視](#)

[プロセスのデッドロック \(futex 競合により発生\)](#), [競合ユーザースペースのロックの特定](#)

ポーリングアプリケーションの監視, [ポーリングアプリケーションの監視](#)  
全体デバイス番号 (コマンドライン引数として使用), [I/O 監視 \(デバイスごと\)](#)  
呼び出し先の追跡, [呼び出し先の追跡](#)  
着信 TCP 接続の監視, [着信 TCP 接続の監視](#)  
競合ユーザースペースのロックの特定 (futex 競合など), [競合ユーザースペースのロックの特定](#)  
累積 I/O の追跡, [累積 I/O の追跡](#)  
複数のコマンドライン引数、例, [呼び出し先の追跡](#)  
関数呼び出しの集計, [関数呼び出しのカウント](#)

スクリプトの文法的なエラー/誤字のエラー  
解析エラー/セマンティックエラー  
SystemTap のエラーを理解する, [解析エラーとセマンティックエラー](#)

スクリプトの誤字のエラー  
解析エラー/セマンティックエラー  
SystemTap のエラーを理解する, [解析エラーとセマンティックエラー](#)

ステートメントブロック  
SystemTap スクリプト  
概要, [SystemTap スクリプト](#)

セッション、SystemTap, [アーキテクチャー](#)  
タイムスタンプ、プロセス名との関連付け  
関連する値の割り当て  
アレイ演算, [関連する値の割り当て](#)

タイムスタンプをプロセス名に関連付ける  
関連する値の割り当て  
アレイ演算, [関連する値の割り当て](#)

タイムスタンプ差分、計算  
アレイからの値の読み取り  
アレイ演算, [アレイからの値の読み取り](#)

タイムスタンプ差分の計算  
アレイからの値の読み取り  
アレイ演算, [アレイからの値の読み取り](#)

ターゲットカーネル  
クロスインストールメンテーション, [他のコンピューター用のインストールメンテーション生成](#)

## ターゲットシステム

クロスインストルメンテーション, [他のコンピューター用のインストルメンテーション生成](#)

## ターゲットシステムおよびホストシステム

設定, [他のコンピューター用のインストルメンテーション生成](#)

## ターゲット記号式、解決できない

解析エラー/セマンティックエラー

SystemTap のエラーを理解する, [解析エラーとセマンティックエラー](#)

## テスト、初期, [初期テスト](#)

### ディスク I/O トラフィック、要約

スクリプトの例, [ディスク読み取り/書き込みトラフィックの要約](#)

### ディスク I/O トラフィックの要約

スクリプトの例, [ディスク読み取り/書き込みトラフィックの要約](#)

## デバイス I/O の監視

SystemTap スクリプトの例, [I/O 監視 \(デバイスごと\)](#)

## デバイス I/O、監視

SystemTap スクリプトの例, [I/O 監視 \(デバイスごと\)](#)

## ネットワークのプロファイリング

SystemTap スクリプトの例, [ネットワークのプロファイリング](#), [カーネルでのネットワークパケットドロップの監視](#)

## ネットワークソケットコード、呼び出された関数の追跡

SystemTap スクリプトの例, [ネットワークソケットコードで呼び出された関数の追跡](#)

## ネットワークソケットコードで呼び出された関数、追跡

SystemTap スクリプトの例, [ネットワークソケットコードで呼び出された関数の追跡](#)

## ネットワークソケットコードで呼び出された関数の追跡

SystemTap スクリプトの例, [ネットワークソケットコードで呼び出された関数の追跡](#)

## ネットワークトラフィック、監視

SystemTap スクリプトの例, [ネットワークのプロファイリング](#), [カーネルでのネットワークパケットドロップの監視](#)

## ハンドラー

SystemTap ハンドラーコンストラクト, [基本的な SystemTap ハンドラーコンストラクト](#)

global, [変数](#)

コマンドラインの引数, [コマンドラインの引数](#)

変数, [変数](#)

変数の表記法, [コマンドラインの引数](#)

---

構文および形式, [基本的な SystemTap ハンドラーコンストラクト](#)

条件付き (conditional) ステートメント, [条件付き \(conditional\) ステートメント](#)

For ループ, [条件付き \(conditional\) ステートメント](#)

if/else, [条件付き \(conditional\) ステートメント](#)

While ループ, [条件付き \(conditional\) ステートメント](#)

条件演算子, [条件付き \(conditional\) ステートメント](#)

概要, [Systemtap ハンドラー/ボディー](#)

ハンドラーおよびイベント, [SystemTap の作動方法](#)

SystemTap スクリプト

概要, [SystemTap スクリプト](#)

ハンドラー関数, [Systemtap ハンドラー/ボディー](#)

パフォーマンスの監視

はじめに, [はじめに](#)

ファイルのデバイス番号 (整数形式)

SystemTap スクリプトの例, [ファイルの読み取りおよび書き込みの監視](#)

ファイルの読み取り/書き込み、監視

SystemTap スクリプトの例, [ファイルの読み取りおよび書き込みの監視](#)

ファイルの読み取りおよび書き込みの監視

SystemTap スクリプトの例, [ファイルの読み取りおよび書き込みの監視](#)

ファイルへの書き込み/読み取り、監視

SystemTap スクリプトの例, [ファイルの読み取りおよび書き込みの監視](#)

ファイルデバイス番号 (整数形式)

SystemTap スクリプトの例, [ファイルの読み取りおよび書き込みの監視](#)

ファイル属性、変更の監視

SystemTap スクリプトの例, [ファイル属性の変更の監視](#)

ファイル属性の変更、監視

SystemTap スクリプトの例, [ファイル属性の変更の監視](#)

ファイル属性の変更の監視

SystemTap スクリプトの例, [ファイル属性の変更の監視](#)

フィードバック

本ガイドの連絡先情報, [ご意見をお寄せください](#)

フライトレコーダーモード, [SystemTap フライトレコーダーモード](#)

ファイルモード, [ファイルフライトレコーダー](#)

メモリー内モード, [メモリー内フライトレコーダー](#)

プロセスのデッドロック (futex 競合により発生)

SystemTap スクリプトの例, [競合ユーザースペースのロックの特定](#)

プローブ

SystemTap スクリプト

[概要](#), [SystemTap スクリプト](#)

プローブの不一致

[解析エラー/セマンティックエラー](#)

SystemTap のエラーを理解する, [解析エラーとセマンティックエラー](#)

プローブポイント (一致なし)

[解析エラー/セマンティックエラー](#)

SystemTap のエラーを理解する, [解析エラーとセマンティックエラー](#)

プローブポイントに一致するものがない

[解析エラー/セマンティックエラー](#)

SystemTap のエラーを理解する, [解析エラーとセマンティックエラー](#)

ヘルプ

[ヘルプの表示](#), [ヘルプが必要ですか?](#)

ホストシステム

[クロスインストールメンテーション](#), [他のコンピューター用のインストールメンテーション生成](#)

ホストシステムおよびターゲットシステム

[クロスインストールメンテーション](#)

[設定](#), [他のコンピューター用のインストールメンテーション生成](#)

ポーリングアプリケーション、監視

SystemTap スクリプトの例, [ポーリングアプリケーションの監視](#)

ポーリングアプリケーションの監視

SystemTap スクリプトの例, [ポーリングアプリケーションの監視](#)

メンバーシップ (アレイ内)、テスト

[条件付きステートメント](#)、アレイの使用

[アレイ演算](#), [条件付きステートメントにおけるアレイの使用](#)

ユーザーおよびカーネルスペース、費やした時間の判定

SystemTap スクリプトの例, [カーネルおよびユーザースペースで費やした時間の判定](#)



## ランタイムエラー/警告

SystemTap のエラーを理解する, [ランタイムエラーおよび警告](#)  
aggregate element not found, [ランタイムエラーおよび警告](#)  
aggregation overflow, [ランタイムエラーおよび警告](#)  
division by 0, [ランタイムエラーおよび警告](#)  
MAXACTION exceeded, [ランタイムエラーおよび警告](#)  
MAXNESTING exceeded, [ランタイムエラーおよび警告](#)  
number of errors: N, skipped probes: M, [ランタイムエラーおよび警告](#)  
pointer dereference fault, [ランタイムエラーおよび警告](#)  
コピー障害, [ランタイムエラーおよび警告](#)

## ローカル変数

name, [Systemtap ハンドラー/ボディー](#)  
サンプル使用例  
\$count, [ファイル読み取り/書き込みの I/O 時間の追跡](#)  
\$return, [ディスク読み取り/書き込みトラフィックの要約, 累積 I/O の追跡](#)

## 一意の鍵

概要  
[アレイ](#), [連想アレイ](#)

## 仮想ファイルシステムの読み込み (累積的)、集計

[アレイ内での複数要素の処理](#)  
[アレイ演算](#), [アレイ内での複数要素の処理](#)

## 仮想ファイルシステムの読み込み (累積的以外)、集計

[アレイ/アレイ要素の消去](#)  
[アレイ演算](#), [アレイおよびアレイ要素の消去/削除](#)

## 仮想ファイルシステムの読み込みの集計 (VFS 読み込み)

[関連する値の増加](#)  
[アレイ演算](#), [関連する値の増加](#)

## 便利な SystemTap スクリプトの例, [便利な SystemTap スクリプト](#)

[値、割り当て](#)  
[アレイ演算](#), [関連する値の割り当て](#)

## 全体デバイス番号 (コマンドライン引数として使用)

[SystemTap スクリプトの例](#), [I/O 監視 \(デバイスごと\)](#)

## 初期テスト, [初期テスト](#)

## 反復、アレイ内の要素の処理

アレイ内での複数要素の処理

アレイ演算, [アレイ内での複数要素の処理](#)

## 同一プローブ内の複数のアレイ演算

アレイ/アレイ要素の消去

アレイ演算, [アレイおよびアレイ要素の消去/削除](#)

## 同期および非同期イベントの例

イベント, [イベント](#)

## 同期イベント

イベント, [イベント](#)

## 呼び出し先の追跡

SystemTap スクリプトの例, [呼び出し先の追跡](#)

## 変数

SystemTap ハンドラーコンストラクト

ハンドラー, [変数](#)

## 変数 (ローカル)

name, [Systemtap ハンドラー/ボディー](#)

サンプル使用例

\$count, [ファイル読み取り/書き込みの I/O 時間の追跡](#)

\$return, [ディスク読み取り/書き込みトラフィックの要約, 累積 I/O の追跡](#)

## 変数/アレイに無効な値

解析エラー/セマンティックエラー

SystemTap のエラーを理解する, [解析エラーとセマンティックエラー](#)

## 変数の表記法

SystemTap ハンドラーコンストラクト

ハンドラー, [コマンドラインの引数](#)

## 形式および構文

printf(), [Systemtap ハンドラー/ボディー](#)

SystemTap スクリプト

概要, [SystemTap スクリプト](#)

SystemTap ハンドラーコンストラクト

ハンドラー, [変数](#)

必要なパッケージ, 必要なカーネル情報 RPM パッケージのインストール

整数抽出

統計集計の計算

アレイ演算, 統計集計の計算

書式指定子

printf(), Systemtap ハンドラー/ボディー

書式文字列

printf(), Systemtap ハンドラー/ボディー

最も大きいディスクの読み取り/書き込み、識別

スクリプトの例, ディスク読み取り/書き込みトラフィックの要約

最も大きいディスクの読み取り/書き込みの特定

スクリプトの例, ディスク読み取り/書き込みトラフィックの要約

本ガイドの目的

はじめに, 本ガイドの目的

条件付きステートメント、アレイの使用

アレイ演算, 条件付きステートメントにおけるアレイの使用

アレイメンバーシップのテスト, 条件付きステートメントにおけるアレイの使用

条件演算子

条件付き (conditional) ステートメント

ハンドラー, 条件付き (conditional) ステートメント

構文

概要

アレイ, 連想アレイ

構文および形式

printf(), Systemtap ハンドラー/ボディー

SystemTap スクリプト

概要, SystemTap スクリプト

SystemTap ハンドラーコンストラクト

ハンドラー, 基本的な SystemTap ハンドラーコンストラクト

標準入力から、スクリプトを実行

用途, SystemTap スクリプトの実行

標準入力からスクリプトを実行, SystemTap スクリプトの実行

## 無効な除算

### ランタイムエラー/警告

SystemTap のエラーを理解する, [ランタイムエラーおよび警告](#)

## 特権のないスクリプト、コードが埋め込まれている

### 解析エラー/セマンティックエラー

SystemTap のエラーを理解する, [解析エラーとセマンティックエラー](#)

## 特権のないスクリプトにコードが埋め込まれている

### 解析エラー/セマンティックエラー

SystemTap のエラーを理解する, [解析エラーとセマンティックエラー](#)

## 特権のないスクリプトに安全でないコードが埋め込まれている

### 解析エラー/セマンティックエラー

SystemTap のエラーを理解する, [解析エラーとセマンティックエラー](#)

## 用途

stap, [SystemTap スクリプトの実行](#)

stapdev, [SystemTap スクリプトの実行](#)

staprun, [SystemTap スクリプトの実行](#)

stapusr, [SystemTap スクリプトの実行](#)

SystemTap スクリプトの実行, [SystemTap スクリプトの実行](#)

オプション、stap, [SystemTap スクリプトの実行](#)

標準入力から、スクリプトを実行, [SystemTap スクリプトの実行](#)

## 目的、ドキュメント

はじめに, [本ガイドの目的](#)

## 着信 TCP 接続、監視

SystemTap スクリプトの例, [着信 TCP 接続の監視](#)

## 着信 TCP 接続の監視

SystemTap スクリプトの例, [着信 TCP 接続の監視](#)

## 着信/発信の関数呼び出し、追跡

SystemTap スクリプトの例, [呼び出し先の追跡](#)

## 着信/発信の関数呼び出しの追跡

SystemTap スクリプトの例, [呼び出し先の追跡](#)

## 空の一意の鍵

アレイからの値の読み取り

アレイ演算, [アレイからの値の読み取り](#)

競合ユーザースペースのロック (futex 競合)、特定

SystemTap スクリプトの例, [競合ユーザースペースのロックの特定](#)

競合ユーザースペースのロックの特定 (futex 競合など)

SystemTap スクリプトの例, [競合ユーザースペースのロックの特定](#)

累積 I/O の監視

SystemTap スクリプトの例, [累積 I/O の追跡](#)

累積 I/O の追跡

SystemTap スクリプトの例, [累積 I/O の追跡](#)

累積 I/O、追跡

SystemTap スクリプトの例, [累積 I/O の追跡](#)

累積的仮想ファイルシステムの読み込み、集計

アレイ内での複数要素の処理

[アレイ演算](#), [アレイ内での複数要素の処理](#)

統計集計

[アレイ演算](#), [統計集計の計算](#)

統計集計が収集したデータの抽出

[統計集計の計算](#)

[アレイ演算](#), [統計集計の計算](#)

統計集計の計算

[アレイ演算](#), [統計集計の計算](#)

[@avg \(整数抽出\)](#), [統計集計の計算](#)

[@count \(整数抽出\)](#), [統計集計の計算](#)

[@max \(整数抽出\)](#), [統計集計の計算](#)

[@min \(整数抽出\)](#), [統計集計の計算](#)

[@sum \(整数抽出\)](#), [統計集計の計算](#)

[count \(演算子\)](#), [統計集計の計算](#)

[統計集計が収集したデータの抽出](#), [統計集計の計算](#)

[統計集計への値の追加](#), [統計集計の計算](#)

統計集計への値の追加

[統計集計の計算](#)

[アレイ演算](#), [統計集計の計算](#)

複数のコマンドライン引数、例

SystemTap スクリプトの例, [呼び出し先の追跡](#)

## 複数のコマンドライン引数の例

SystemTap スクリプトの例, [呼び出し先の追跡](#)

## 解析エラー/セマンティックエラー

SystemTap のエラーを理解する, [解析エラーとセマンティックエラー](#)

'foreach' 中に変数が修正される, [解析エラーとセマンティックエラー](#)

guru モード, [解析エラーとセマンティックエラー](#)

libdwfl 失敗, [解析エラーとセマンティックエラー](#)

スクリプトの文法的なエラー/誤字のエラー, [解析エラーとセマンティックエラー](#)

プローブの不一致, [解析エラーとセマンティックエラー](#)

プローブポイントに一致するものがない, [解析エラーとセマンティックエラー](#)

変数/アレイに無効な値, [解析エラーとセマンティックエラー](#)

特権のないスクリプトにコードが埋め込まれている, [解析エラーとセマンティックエラー](#)

解決できないターゲット記号式, [解析エラーとセマンティックエラー](#)

解決できない識別子のタイプ, [解析エラーとセマンティックエラー](#)

解決できない関数呼び出し, [解析エラーとセマンティックエラー](#)

記号やアレイに期待されているインデックス式, [解析エラーとセマンティックエラー](#)

識別子タイプの不一致, [解析エラーとセマンティックエラー](#)

非グローバルアレイ, [解析エラーとセマンティックエラー](#)

## 解決できないターゲット記号式

解析エラー/セマンティックエラー

SystemTap のエラーを理解する, [解析エラーとセマンティックエラー](#)

## 解決できない識別子のタイプ

解析エラー/セマンティックエラー

SystemTap のエラーを理解する, [解析エラーとセマンティックエラー](#)

## 解決できない関数呼び出し

解析エラー/セマンティックエラー

SystemTap のエラーを理解する, [解析エラーとセマンティックエラー](#)

## 記号やアレイに期待されているインデックス式

解析エラー/セマンティックエラー

SystemTap のエラーを理解する, [解析エラーとセマンティックエラー](#)

## 設定とインストール, インストールと設定

### 識別子タイプの不一致

解析エラー/セマンティックエラー

SystemTap のエラーを理解する, [解析エラーとセマンティックエラー](#)

## 連想アレイ

概要, [連想アレイ](#)

example, [連想アレイ](#)

インデックス式, [連想アレイ](#)

キーのペア, [連想アレイ](#)

一意の鍵, [連想アレイ](#)

構文, [連想アレイ](#)

関連する値, [連想アレイ](#)

## 関数, [Systemtap ハンドラー/ボディー](#)

cpu(), [Systemtap ハンドラー/ボディー](#)

ctime(), [Systemtap ハンドラー/ボディー](#)

gettimeofday\_s(), [Systemtap ハンドラー/ボディー](#)

pp(), [Systemtap ハンドラー/ボディー](#)

SystemTap スクリプト

概要, [SystemTap スクリプト](#)

target(), [Systemtap ハンドラー/ボディー](#)

thread\_indent(), [Systemtap ハンドラー/ボディー](#)

tid(), [Systemtap ハンドラー/ボディー](#)

uid(), [Systemtap ハンドラー/ボディー](#)

## 関数 (ハンドラーで使用)

exit(), [Systemtap ハンドラー/ボディー](#)

## 関数呼び出し (着信/発信)、追跡

SystemTap スクリプトの例, [呼び出し先の追跡](#)

## 関数呼び出し (解決できない)

解析エラー/セマンティックエラー

SystemTap のエラーを理解する, [解析エラーとセマンティックエラー](#)

## 関数呼び出し、集計

SystemTap スクリプトの例, [関数呼び出しのカウンタ](#)

## 関数呼び出しのカウンタ

SystemTap スクリプトの例, [関数呼び出しのカウンタ](#)

## 関数呼び出しの集計

SystemTap スクリプトの例, [関数呼び出しのカウンタ](#)

## 関連する値

概要

アレイ, [連想アレイ](#)

## 関連する値の割り当て

アレイ演算, [関連する値の割り当て](#)

タイムスタンプをプロセス名に関連付ける, [関連する値の割り当て](#)

タイムスタンプをプロセス名に関連付ける

アレイ演算, [関連する値の割り当て](#)

## 関連する値の増加

アレイ演算, [関連する値の増加](#)

仮想ファイルシステムの読み込みの集計 (VFS 読み込み), [関連する値の増加](#)

## 集計 (統計)

アレイ演算, [統計集計の計算](#)

## 非グローバルアレイ

解析エラー/セマンティックエラー

SystemTap のエラーを理解する, [解析エラーとセマンティックエラー](#)

## 非同期イベント

イベント, [イベント](#)

## A

### aggregate element not found

ランタイムエラー/警告

SystemTap のエラーを理解する, [ランタイムエラーおよび警告](#)

### aggregation overflow

ランタイムエラー/警告

SystemTap のエラーを理解する, [ランタイムエラーおよび警告](#)

## B

### begin

イベント, [イベント](#)

## C

CONFIG\_HZ, computing for, [変数](#)

### count 演算子

統計集計の計算

array (演算子), [統計集計の計算](#)



---

## CPU ティック

SystemTap スクリプトの例, [カーネルおよびユーザースペースで費やした時間の判定](#)

## cpu()

関数, [Systemtap ハンドラー/ボディー](#)

cross-compiling, [他のコンピューター用のインストルメンテーション生成](#)

## ctime()

関数, [Systemtap ハンドラー/ボディー](#)

## ctime()、使用例

スクリプトの例, [ディスク読み取り/書き込みトラフィックの要約](#)

## D

### Delete 演算子

アレイ/アレイ要素の消去

アレイ演算, [アレイおよびアレイ要素の消去/削除](#)

### division by 0

ランタイムエラー/警告

SystemTap のエラーを理解する, [ランタイムエラーおよび警告](#)

## E

### end

イベント, [イベント](#)

### example

概要

アレイ, [連想アレイ](#)

### exceeded MAXACTION

ランタイムエラー/警告

SystemTap のエラーを理解する, [ランタイムエラーおよび警告](#)

### exceeded MAXNESTING

ランタイムエラー/警告

SystemTap のエラーを理解する, [ランタイムエラーおよび警告](#)

## exit()

関数, [Systemtap ハンドラー/ボディー](#)

## F

### For ループ

条件付き (conditional) ステートメント

ハンドラー, [条件付き \(conditional\) ステートメント](#)

### foreach

アレイ内での複数要素の処理

アレイ 演算, [アレイ内での複数要素の処理](#)

### foreach 出力の制限

アレイ内での複数要素の処理

アレイ 演算, [アレイ内での複数要素の処理](#)

### foreach 出力の順序

アレイ内での複数要素の処理

アレイ 演算, [アレイ内での複数要素の処理](#)

### format

概要

アレイ, [連想アレイ](#)

### futex システムコール

SystemTap スクリプトの例, [競合ユーザースペースのロックの特定](#)

### futex 競合、定義

SystemTap スクリプトの例, [競合ユーザースペースのロックの特定](#)

### futex 競合、特定

SystemTap スクリプトの例, [競合ユーザースペースのロックの特定](#)

## G

### gettimeofday\_s()

関数, [Systemtap ハンドラー/ボディー](#)

### global

SystemTap ハンドラーコンストラクト

ハンドラー, [変数](#)

### guru モード

解析エラー/セマンティックエラー

SystemTap のエラーを理解する, [解析エラーとセマンティックエラー](#)

## I

## I/O の時間

SystemTap スクリプトの例, [ファイル読み取り/書き込みの I/O 時間の追跡](#)

## I/O アクティビティのプリント (累積)

SystemTap スクリプトの例, [累積 I/O の追跡](#)

## I/O トラフィック、要約

スクリプトの例, [ディスク読み取り/書き込みトラフィックの要約](#)

## I/O 時間、監視

SystemTap スクリプトの例, [ファイル読み取り/書き込みの I/O 時間の追跡](#)

## I/O 時間の監視

SystemTap スクリプトの例, [ファイル読み取り/書き込みの I/O 時間の追跡](#)

## I/O 監視 (デバイスごと)

SystemTap スクリプトの例, [I/O 監視 \(デバイスごと\)](#)

## if/else

条件付き (conditional) ステートメント

ハンドラー, [条件付き \(conditional\) ステートメント](#)

## if/else ステートメント、アレイの使用

アレイ演算, [条件付きステートメントにおけるアレイの使用](#)

## if/else 条件、代替の構文

SystemTap スクリプトの例, [ネットワークのプロファイリング](#)

## inode 番号

SystemTap スクリプトの例, [ファイルの読み取りおよび書き込みの監視](#)

## K

## kernel.function("function")

イベント, [イベント](#)

## kernel.trace("tracepoint")

イベント, [イベント](#)

## L

## libdwfl 失敗

解析エラー/セマンティックエラー

SystemTap のエラーを理解する, [解析エラーとセマンティックエラー](#)

## M

## MAXACTION exceeded

ランタイムエラー/警告

SystemTap のエラーを理解する, [ランタイムエラーおよび警告](#)

## MAXNESTING exceeded

ランタイムエラー/警告

SystemTap のエラーを理解する, [ランタイムエラーおよび警告](#)

## module("module")

イベント, [イベント](#)

## N

## name

ローカル変数, [Systemtap ハンドラー/ボディー](#)

## net/socket.c、関数の追跡

SystemTap スクリプトの例, [ネットワークソケットコードで呼び出された関数の追跡](#)

## number of errors: N, skipped probes: M

ランタイムエラー/警告

SystemTap のエラーを理解する, [ランタイムエラーおよび警告](#)

## O

## operations

[アレイ/アレイ要素の消去](#), [アレイおよびアレイ要素の消去/削除](#)

Delete 演算子, [アレイおよびアレイ要素の消去/削除](#)

仮想ファイルシステムの読み込み (累積的以外)、集計, [アレイおよびアレイ要素の消去/削除](#)

同一プローブ内の複数のアレイ演算, [アレイおよびアレイ要素の消去/削除](#)

[アレイおよびアレイ要素の削除](#), [アレイおよびアレイ要素の消去/削除](#)

[アレイからの値の読み取り](#), [アレイからの値の読み取り](#)

シンプルな計算でのアレイの使用, [アレイからの値の読み取り](#)

タイムスタンプ差分の計算, [アレイからの値の読み取り](#)

空の一意の鍵, [アレイからの値の読み取り](#)

[アレイ内での複数要素の処理](#), [アレイ内での複数要素の処理](#)

foreach, [アレイ内での複数要素の処理](#)

foreach 出力の制限, [アレイ内での複数要素の処理](#)

foreach 出力の順序, [アレイ内での複数要素の処理](#)

反復、アレイ内の要素の処理, [アレイ内での複数要素の処理](#)

累積的仮想ファイルシステムの読み込み、集計、[アレイ内での複数要素の処理](#)

[アレイ内の複数要素](#), [アレイ内での複数要素の処理](#)

[条件付きステートメント](#)、[アレイの使用](#), [条件付きステートメントにおけるアレイの使用](#)  
[アレイメンバーシップのテスト](#), [条件付きステートメントにおけるアレイの使用](#)

[統計集計の計算](#), [統計集計の計算](#)

[@avg \(整数抽出\)](#), [統計集計の計算](#)

[@count \(整数抽出\)](#), [統計集計の計算](#)

[@max \(整数抽出\)](#), [統計集計の計算](#)

[@min \(整数抽出\)](#), [統計集計の計算](#)

[@sum \(整数抽出\)](#), [統計集計の計算](#)

[count \(演算子\)](#), [統計集計の計算](#)

[統計集計が収集したデータの抽出](#), [統計集計の計算](#)

[統計集計への値の追加](#), [統計集計の計算](#)

[連想アレイ](#), [SystemTap でのアレイ演算](#)

[関連する値の割り当て](#)

[タイムスタンプをプロセス名に関連付ける](#), [関連する値の割り当て](#)

[関連する値の増加](#), [関連する値の増加](#)

[仮想ファイルシステムの読み込みの集計 \(VFS 読み込み\)](#), [関連する値の増加](#)

## P

[pointer dereference fault](#)

[ランタイムエラー/警告](#)

[SystemTap のエラーを理解する](#), [ランタイムエラーおよび警告](#)

[pp\(\)](#)

[関数](#), [Systemtap ハンドラー/ボディー](#)

[printf\(\)](#)

[書式指定子](#), [Systemtap ハンドラー/ボディー](#)

[書式文字列](#), [Systemtap ハンドラー/ボディー](#)

[構文および形式](#), [Systemtap ハンドラー/ボディー](#)

## S

[stap](#)

[用途](#), [SystemTap スクリプトの実行](#)

[stap オプション](#), [SystemTap スクリプトの実行](#)

[stapdev](#)

用途, [SystemTap スクリプトの実行](#)

staprun

用途, [SystemTap スクリプトの実行](#)

stapusr

用途, [SystemTap スクリプトの実行](#)

stat -c、ファイルデバイス番号の決定 (整数形式)

SystemTap スクリプトの例, [ファイルの読み取りおよび書き込みの監視](#)

stat -c、全体デバイス番号の決定

SystemTap スクリプトの例, [I/O 監視 \(デバイスごと\)](#)

syscall.system\_call

イベント, [イベント](#)

SystemTap のアーキテクチャー, [アーキテクチャー](#)

SystemTap のエラーを理解する

ランタイムエラー/警告, [ランタイムエラーおよび警告](#)

aggregate element not found, [ランタイムエラーおよび警告](#)

aggregation overflow, [ランタイムエラーおよび警告](#)

division by 0, [ランタイムエラーおよび警告](#)

MAXACTION exceeded, [ランタイムエラーおよび警告](#)

MAXNESTING exceeded, [ランタイムエラーおよび警告](#)

number of errors: N, skipped probes: M, [ランタイムエラーおよび警告](#)

pointer dereference fault, [ランタイムエラーおよび警告](#)

コピー障害, [ランタイムエラーおよび警告](#)

解析エラー/セマンティックエラー, [解析エラーとセマンティックエラー](#)

'foreach' 中に変数が修正される, [解析エラーとセマンティックエラー](#)

guru モード, [解析エラーとセマンティックエラー](#)

libdwfl 失敗, [解析エラーとセマンティックエラー](#)

スクリプトの文法的なエラー/誤字のエラー, [解析エラーとセマンティックエラー](#)

プローブの不一致, [解析エラーとセマンティックエラー](#)

プローブポイントに一致するものがない, [解析エラーとセマンティックエラー](#)

変数/アレイに無効な値, [解析エラーとセマンティックエラー](#)

特権のないスクリプトにコードが埋め込まれている, [解析エラーとセマンティックエラー](#)

解決できないターゲット記号式, [解析エラーとセマンティックエラー](#)

解決できない識別子のタイプ, [解析エラーとセマンティックエラー](#)

解決できない関数呼び出し, [解析エラーとセマンティックエラー](#)

記号やアレイに期待されているインデックス式, [解析エラーとセマンティックエラー](#)

識別子タイプの不一致, [解析エラーとセマンティックエラー](#)

非グローバルアレイ, [解析エラーとセマンティックエラー](#)

SystemTap の作動方法, [SystemTap の作動方法](#)

SystemTap セッション, [アーキテクチャー](#)

[アーキテクチャー](#), [アーキテクチャー](#)

イベントおよびハンドラー, [SystemTap の作動方法](#)

イベントタイプ, [SystemTap の作動方法](#)

SystemTap の使用, [SystemTap の使用](#)

SystemTap の実行に必要な RPM パッケージ, [必要なカーネル情報 RPM パッケージのインストール](#)

SystemTap の実行に必要なパッケージ, [必要なカーネル情報 RPM パッケージのインストール](#)

SystemTap の機能

はじめに, [SystemTap の機能](#)

SystemTap アーキテクチャー, [アーキテクチャー](#)

SystemTap スクリプト

便利な例, [便利な SystemTap スクリプト](#)

概要, [SystemTap スクリプト](#)

イベントおよびハンドラー, [SystemTap スクリプト](#)

コンポーネント, [SystemTap スクリプト](#)

ステートメントブロック, [SystemTap スクリプト](#)

プローブ, [SystemTap スクリプト](#)

形式および構文, [SystemTap スクリプト](#)

関数, [SystemTap スクリプト](#)

SystemTap スクリプト、実行方法, [SystemTap スクリプトの実行](#)

SystemTap スクリプトからインストールメンテーション/カーネルモジュールをコンパイルする, [他のコンピューター用のインストールメンテーション生成](#)

SystemTap スクリプトからインストールメンテーションモジュール、構築する, [他のコンピューター用のインストールメンテーション生成](#)

SystemTap スクリプトからインストールメンテーションモジュールを構築する, [他のコンピューター用のインストールメンテーション生成](#)

SystemTap スクリプトからカーネルモジュール、構築する, [他のコンピューター用のインストールメンテーション生成](#)

SystemTap スクリプトからカーネルモジュールを構築する, [他のコンピューター用のインストールメンテーション生成](#)

SystemTap スクリプトの例, [便利な SystemTap スクリプト](#)

CPU ティック, [カーネルおよびユーザースペースで費やした時間の判定](#)

ctime(), [使用例](#), [ディスク読み取り/書き込みトラフィックの要約](#)

futex システムコール, [競合ユーザースペースのロックの特定](#)

I/O 時間の監視, [ファイル読み取り/書き込みの I/O 時間の追跡](#)

if/else 条件、代替の構文, [ネットワークのプロファイリング](#)

inode 番号, [ファイルの読み取りおよび書き込みの監視](#)

net/socket.c、関数の追跡, [ネットワークソケットコードで呼び出された関数の追跡](#)

stat -c、ファイルデバイス番号の決定 (整数形式), [ファイルの読み取りおよび書き込みの監視](#)

stat -c、全体デバイス番号の決定、[I/O 監視 \(デバイスごと\)](#)  
thread\_indent(), サンプル使用例、[呼び出し先の追跡](#)  
timer.ms(), サンプル使用例、[関数呼び出しのカウンタ](#)  
timer.s(), サンプル使用例、[ポーリングアプリケーションの監視、最もよく使われるシステムコールの追跡](#)  
trigger function、[呼び出し先の追跡](#)  
usrdev2kerndev(), [I/O 監視 \(デバイスごと\)](#)  
カーネルおよびユーザースペースで費やした時間の判定、[カーネルおよびユーザースペースで費やした時間の判定](#)  
システムコールの監視、[最もよく使われるシステムコールの追跡](#)  
システムコールの監視 (プロセスごとのボリューム)、[プロセスごとのシステムコールボリュームの追跡](#)  
ディスク I/O トラフィックの要約、[ディスク読み取り/書き込みトラフィックの要約](#)  
デバイス I/O の監視、[I/O 監視 \(デバイスごと\)](#)  
ネットワークのプロファイリング、[ネットワークのプロファイリング](#)、[カーネルでのネットワークパケットドロップの監視](#)  
ネットワークソケットコードで呼び出された関数の追跡、[ネットワークソケットコードで呼び出された関数の追跡](#)  
ファイルの読み取りおよび書き込みの監視、[ファイルの読み取りおよび書き込みの監視](#)  
ファイルデバイス番号 (整数形式)、[ファイルの読み取りおよび書き込みの監視](#)  
ファイル属性の変更の監視、[ファイル属性の変更の監視](#)  
プロセスのデッドロック (futex 競合により発生)、[競合ユーザースペースのロックの特定](#)  
ポーリングアプリケーションの監視、[ポーリングアプリケーションの監視](#)  
全体デバイス番号 (コマンドライン引数として使用)、[I/O 監視 \(デバイスごと\)](#)  
呼び出し先の追跡、[呼び出し先の追跡](#)  
着信 TCP 接続の監視、[着信 TCP 接続の監視](#)  
競合ユーザースペースのロックの特定 (futex 競合など)、[競合ユーザースペースのロックの特定](#)  
累積 I/O の追跡、[累積 I/O の追跡](#)  
複数のコマンドライン引数、例、[呼び出し先の追跡](#)  
関数呼び出しの集計、[関数呼び出しのカウンタ](#)

## SystemTap スクリプトの実行

用途、[SystemTap スクリプトの実行](#)

## SystemTap スクリプト関数、Systemtap ハンドラー/ボディー

### SystemTap ステートメント

#### SystemTap ハンドラーコンストラクト

global、[変数](#)

コマンドラインの引数、[コマンドラインの引数](#)

変数、[変数](#)

変数の表記法、[コマンドラインの引数](#)

条件付き (conditional) ステートメント、[条件付き \(conditional\) ステートメント](#)

For ループ、[条件付き \(conditional\) ステートメント](#)



if/else, [条件付き \(conditional\) ステートメント](#)  
While ループ, [条件付き \(conditional\) ステートメント](#)  
条件演算子, [条件付き \(conditional\) ステートメント](#)

SystemTap セッション, [アーキテクチャー](#)

SystemTap ハンドラー

SystemTap ハンドラーコンストラクト, [基本的な SystemTap ハンドラーコンストラクト](#)  
[構文および形式](#), [基本的な SystemTap ハンドラーコンストラクト](#)

systemtap パッケージ, [SystemTap のインストール](#)

systemtap-runtime パッケージ, [SystemTap のインストール](#)

systemtap-testsuite パッケージ

[サンプルスクリプト](#), [便利な SystemTap スクリプト](#)

## T

Tapsets

[定義](#), [Tapsets](#)

target()

[関数](#), [Systemtap ハンドラー/ボディー](#)

TCP 接続 (着信)、監視

[SystemTap スクリプトの例](#), [着信 TCP 接続の監視](#)

thread\_indent()

[関数](#), [Systemtap ハンドラー/ボディー](#)

thread\_indent(), サンプル使用例

[SystemTap スクリプトの例](#), [呼び出し先の追跡](#)

tid()

[関数](#), [Systemtap ハンドラー/ボディー](#)

timer イベント

[イベント](#), [イベント](#)

timer.ms(), サンプル使用例

[SystemTap スクリプトの例](#), [関数呼び出しのカウント](#)

timer.s(), サンプル使用例

[SystemTap スクリプトの例](#), [ポーリングアプリケーションの監視](#), [最もよく使われるシステムコールの追跡](#)

tracepoint, [イベント](#), [カーネルでのネットワークパケットドロップの監視](#)

trigger function

## SystemTap スクリプトの例, 呼び出し先の追跡

### U

uid()

関数, [Systemtap ハンドラー/ボディー](#)

uname -m, [他のコンピューター用のインストールメンテーション生成](#)

uname -r, [必要なカーネル情報 RPM パッケージのインストール](#)

usrdev2kerndev()

SystemTap スクリプトの例, [I/O 監視 \(デバイスごと\)](#)

### V

VF 読み取り、集計

関連する値の増加

アレイ演算, [関連する値の増加](#)

vfs.file\_operation

イベント, [イベント](#)

### W

While ループ

条件付き (conditional) ステートメント

ハンドラー, [条件付き \(conditional\) ステートメント](#)