SecHack365

ゼロから作ったソフトウェアルータ CURONOS

学習駆動コース 柚山大哉

What is “ルータ”?

“ルーターまたはルータ（英: router）は、コンピュータネットワークにおいて、データを2つ以上の異なるネットワーク間に中継する通信機器である。”

「ルーター」, 『フリー百科事典　ウィキペディア日本語版』, 2022年1月25日 12:50 JST, URL: https://ja.wikipedia.org

ソフトウェアルータとは, ルータの機能を専用のハードウェアを用いずに汎用的なサーバやパソコンで動作させることができるソフトウェアです. ネットワークの仮想化やSDNの普及が進んでいる現在, 非常に大切なものとなっています.

世界最速を目指そう!

CURONOSはOSから開発された普通のコンピュータ(x86_64のCPUを搭載したもの)上で動作するソフトウェアルータであり, 既存のOSをベースとしません. 世の中にあるソフトウェアルータはたいてい, LinuxやBSDをベースに作られており, 一つぐらいOSからまるごと開発されたソフトウェアルータがあっても良いのではないか,すべて作ってすべて理解したら大変面白いのではないかと考え, 開発を始めました. **OSから全部開発しているので, 当然世界最速のルータになる潜在性があるのです.**

堅牢性の確保

CURONOSはOSからすべて開発しているため, 大変細かな問題に注意して開発しなければ突然停止するなどの問題を引き起こす恐れがあります. そのため, 問題にいち早く気付く必要があり, そのための工夫を行っています.

ファジングによる通信試験

オープンソースなファジングツールであるISIC(IP Stack Integrity Checker)を利用し, 開発しているプロトコルスタックがシステムに問題を引き起こすことがないかを検証しています. この試験によって, **受信に一度エラーが発生した場合, ネットワークインターフェースの受信処理が停止する問題を発見できました.**

```
192.168.111.3   192.168.111.1   UDP       81 16655 → 48033 [BAD UDP LENGTH 12613 > IP PAYLOAD LENGTH] Len=12605
192.168.111.3   192.168.111.1   MIPv6    214 Unknown Mobility Header (159)[Malformed Packet]
192.168.111.3   192.168.111.1   IGMP    1138 Unknown Type:0x25
192.168.111.3   192.168.111.1   UDP     1272 41594 → 9795 [BAD UDP LENGTH 37683 > IP PAYLOAD LENGTH] Len=37675
192.168.111.3   192.168.111.1   IPv4    1203 Unassigned (218)
192.168.111.3   192.168.111.1   IPv4    1122 Sitara Networks (109)
192.168.111.3   192.168.111.1   IPv4     162 Unassigned (146)
192.168.111.3   192.168.111.1   IPv4     171 Unknown (244)
```

ツールが生成する多様な通信

徹底したメモリ管理

デバッグビルド時にはメモリのアロケーションをハッシュテーブルに記録し, 開放する際に削除する機構を実装. BGPなどで大量の経路情報の登録と取り消しを扱う際に, メモリの開放に不備があると大きな問題を引き起こします. この機構によって実際に**IP経路テーブルにメモリの二重開放の開放の問題があることを発見できました.**

```
[06:36:48][CPU0/WARNING] Debug memory map index 38912 is already 0, 0x1f29390 double free?
[06:36:48][CPU0/TRACE] Free address 0x1f29350
[06:36:48][CPU0/WARNING] Debug memory map index 79872 is already 0, 0x1f29350 double free?
[06:36:48][CPU0/TRACE] Free address 0x1f29310
[06:36:48][CPU0/WARNING] Debug memory map index 120832 is already 0, 0x1f29310 double free?
```

テーブルに記録されていないアドレスを解放した際の警告

トラフィックを遮断する機械語を動的に生成

JITパケットフィルタ

```
ip filter 192.0.2.0/24
ip filter 133.27.0.0/16
ip filter jit flash
```

コマンド入力

Translate

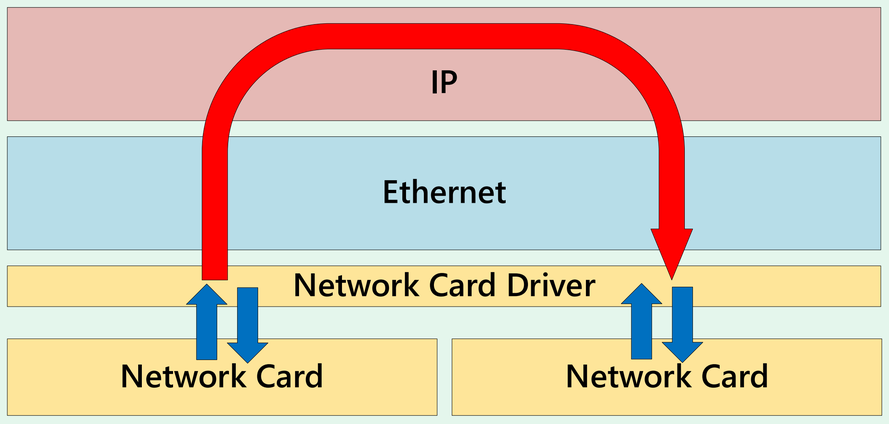
```
...
B9 00 00 00 00          ; mov ecx, 0
FF C1                   ; inc ecx
89 C2                   ; mov edx, eax
81 E2 00 FF FF FF       ; and eax, $00ffffff
81 FA 00 02 80 C0       ; cmp eax, $000280c0
74 17                   ; je 0x17
FF C1                   ; inc ecx
89 C2                   ; mov edx, eax
81 E2 00 00 FF FF       ; and eax, $0000ffff
81 FA 85 1B 00 00       ; cmp edx, $0x851b0000
74 05                   ; je $0x05
B9 00 00 00 00          ; mov ecx, 0
89 C8                   ; mov eax, ecx
...
```

生成された機械語
(eaxレジスタにIPアドレスが入っている)

IPフィルタリングを行う機械語をルータ起動後にも動的に生成できる機能を備え, コマンドを通してドライバ内にフィルタリングコードを配置できます. ドライバ内でドロップすることで, 無駄な処理を上位レイヤに送ることを防ぐことができます. さらに, フィルタの何番目のエントリに引っかかったかを記録しているため, それをもとにエントリの順番を入れ替えるなどの拡張ができます.
このような設定を機械語に変換して動作させるアプローチはパケットフィルタ以外の分野でも使用できるのではないかと考えています.

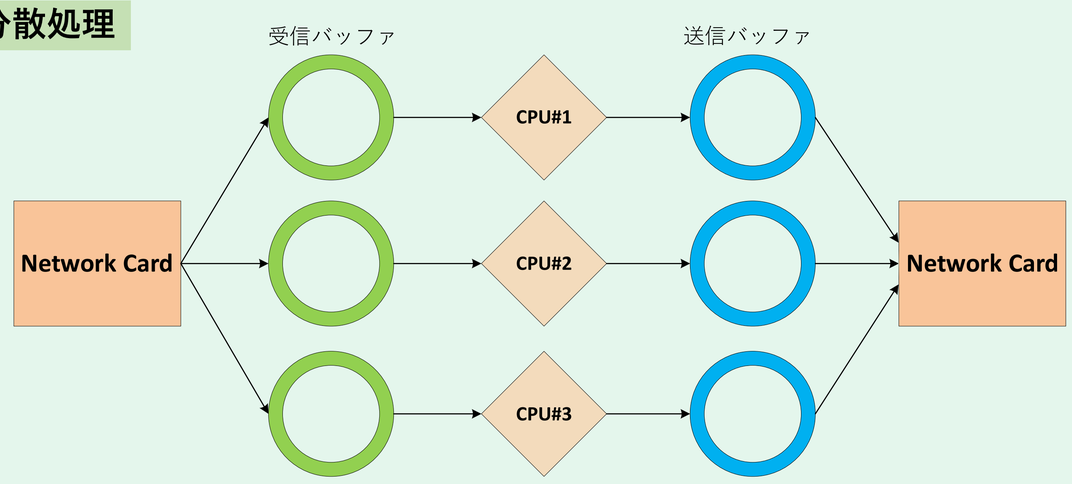
CURONOSのアーキテクチャ

IPフォワーディング



1CPUコンテクストで通信の受信からそのパケットの再送出までを行う. 受信処理は割り込みまたはPollingで行うことができる. バッファリングを行わないため, 低遅延な通信を実現しています.

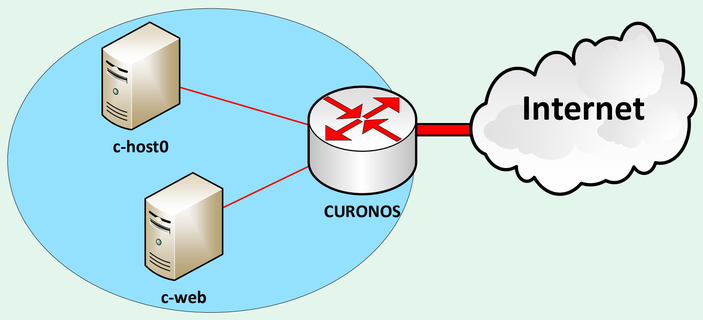
分散処理



マルチコアでの複数の送信/受信キューの使用もサポートしており, 10GbEやそれ以上に高速な通信行う準備が整っています.

現実世界・インターネットでの運用実験

ルータとはデータを2つ以上のネットワーク間に中継する機器であるので, ルータの機能を実験するにはどこかにネットワークを用意する必要があります. ローカルでの実験では面白さが足りないので, **大学よりグローバルアドレスのサブネットを借用し, インターネットにてグローバル通信の実験を行いました.**



大学内のデータセンタにあるVMware ESXi上に仮想マシンとしてCURONOSを仮想マシンとして動作させ, 借用したサブネットのゲートウェイとして稼働させました. CURONOSの配下に2台のホストを配置し, WebのやICMPの通信を実験しました.

```
                                My traceroute  [v0.93]

Keys:  Help   Display mode   Restart statistics   Order of fields   quit
                                                  Packets               Pings
 Host                                          Loss%   Snt   Last   Avg  Best  Wrst StDev
 1. ve-302.ncs1.sfc.wide.ad.jp                 0.0%    20    1.1   1.1   0.9   1.3   0.1
 2. 203.178.140.162                            0.0%    20    0.4   0.3   0.2   0.5   0.1
 3. 203.178.140.164                            0.0%    20    0.5   0.6   0.3   1.0   0.1
 4. fw1-sfc-wide-boundary-v4.sfc.keio.ac.jp    0.0%    20    0.7   1.9   0.3  25.8   5.6
 5. gw2-core-v4.sfc.keio.ac.jp                 0.0%    20    0.8   4.9   0.5  31.9   8.9
 6. gw-d-trust.sfc.keio.ac.jp                  0.0%    20    0.7   0.7   0.5   0.9   0.1
 7. curonos.d-trust.sfc.keio.ac.jp             0.0%    20    1.0   0.9   0.7   1.3   0.1
 8. c-host0.d-trust.sfc.keio.ac.jp             0.0%    20    1.3   1.2   1.0   1.6   0.1
```

インターネットからのCURONOSネットワーク内ホストの観測

その結果, 連続200時間以上, 総計500時間以上の稼働を記録し,　多くの方に通信を体験して楽しんでもらえました.

謝辞

本作品の開発で, SecHack365の皆様, 慶応義塾大学 環境情報学部 村井純合同研究室, WIDEプロジェクトの多くの方々に大変お世話になりました.ここにお礼申し上げます. ありがとうございました.

More Infomation => https://web.sfc.wide.ad.jp/~daiya/curonos

NICT 国立研究開発法人 情報通信研究機構

本作品に関する連絡先
daiya@sfc.wide.ad.jp

CURONOSを動かす場所を大募集!!独立行政法人やNICTの方など, ぜひどうぞ!