FreeBSD 12.0 と jail と VNET

seirios / HEO SeonMeyong

FreeBSD 12.0 と jail

- FreeBSD 12.0 Release で、GENERIC kernel に VIMAGE が入った
- jail/vnetでもpfできるらしい
- うちの環境、そろそろまともに組み直さないと....

Jailデビュ-のChance到来

喜び勇んでJail

- **Webサーバーとか**
  - PHPとRubyとPerlとStaticをバラバラにVMで持つのはそろそろ嫌
  - でも単純に全部まとめるのはもっと嫌
  - **Jailでしょ！**
- **Mailサーバーとか**
  - MTA(Postfix)とMDA(Dovecot)を (ry
  - **Jailでしょ！**
- **DNSとNTPとLDAPとか**
  - そんなに性能いらないのにいちいち(ry
  - でも単純に(ry
  - **Jailでしょ！**

と思ったんだよ、うん。

おおもとの設定 -0-

- XCP-ng配下で実行。なお、VMwareでもKVMでもBhyveでも同じはず
  - Jailの元となるJailerを作成(以下JL-1)
  - JL-1をcloneしてJL-2を作成
  - JL-1にJail(なんでもいい・以下j-1)を作成
  - JL-2にJail(なんでもいい・以下j-2)を作成

```
mweb01-php {
  $id               = "1";
  $ip4_xn1          = "10.1.105.103/24";
  $ip6_xn1          = "fe80::1:101/64";
  $ip4_xn2          = "10.1.106.103/24";
  $ip6_xn2          = "fe80::2:101/64";
  $route4           = "10.1.105.1";
  $route6           = "::1 -blackhole";
  vnet;

  exec.prestart     = "/sbin/ifconfig epair${id}1 create up";
  exec.prestart    += "/sbin/ifconfig bridge1 addm epair${id}1a";
  exec.prestart    += "/sbin/ifconfig epair${id}2 create up";
  exec.prestart    += "/sbin/ifconfig bridge2 addm epair${id}2a";

  exec.created      = "/sbin/ifconfig epair${id}1b vnet ${name}";
  exec.created     += "/sbin/ifconfig epair${id}2b vnet ${name}";

  exec.start        = "/sbin/ifconfig epair${id}1b inet  ${ip4_xn1} up";
  exec.start       += "/sbin/ifconfig epair${id}1b inet6 ${ip6_xn1} up";
  exec.start       += "/sbin/ifconfig epair${id}2b inet  ${ip4_xn2} up";
  exec.start       += "/sbin/ifconfig epair${id}2b inet6 ${ip6_xn2} up";
  exec.start       += "/sbin/route add -inet  default -gateway ${route4}";
  exec.start       += "/sbin/route add -inet6 default -gateway ${route6}";
  exec.start       += "/bin/sh /etc/rc";

  exec.stop         = "/bin/sh /etc/rc.shutdown";

  exec.poststop     = "ifconfig epair${id}1a destroy";
  exec.poststop    += "ifconfig epair${id}2a destroy";

  persist;
}
```

ハマった(お約束) -1-

1. vnet.interfaceに複数のepairがかけない...
   - exec.prestartで、epairをcreateする
   - exec.createdで、epairをattachする
2. exec.startでroute add -inet6 default ::1 -blackholeしたらエラー
   - 結果jailが起動しない
   - route add -inet default xxx.xxx.xxx.xxxは動作した

- 内藤さんのおかげで、1番は解決

```
vnet.interface      = "epair${id}1b";
vnet.interface     += "epair${id}2b";
```

- しかし、2番は未解決
  - 色々試したんだが...
  - 起動後にrouteコマンドを実行するとちゃんと動くんだけど...なぜ？

こうなった -1-

```
mweb01-php {
  $id               = "1";
  $ip4_xn1           = "10.1.105.103/24";
  $ip6_xn1           = "fe80::1:101/64";
  $ip4_xn2           = "10.1.106.103/24";
  $ip6_xn2           = "fe80::2:101/64";
  $route4           = "10.1.105.1";
  $route6           = “::1 -blackhole";
  vnet;
  vnet.interface     = "epair${id}1b";
  vnet.interface    += "epair${id}2b";

  exec.prestart      = "/sbin/ifconfig epair${id}1 create up";
  exec.prestart     += "/sbin/ifconfig bridge1     addm epair${id}1a";
  exec.prestart     += "/sbin/ifconfig epair${id}2 create up";
  exec.prestart     += "/sbin/ifconfig bridge2     addm epair${id}2a";

  exec.start         = "/sbin/ifconfig epair${id}1b inet  ${ip4_xn1} up";
  exec.start        += "/sbin/ifconfig epair${id}1b inet6 ${ip6_xn1} up";
  exec.start        += "/sbin/ifconfig epair${id}2b inet  ${ip4_xn2} up";
  exec.start        += "/sbin/ifconfig epair${id}2b inet6 ${ip6_xn2} up";
  exec.start        += "/sbin/route add -inet  default -gateway $
{route4}";
  exec.start        += "/bin/sh /etc/rc";

  exec.stop          = "/bin/sh /etc/rc.shutdown";

  exec.poststop      = "ifconfig epair${id}1a destroy";
  exec.poststop     += "ifconfig epair${id}2a destroy";
  persist;
}
```

これで Jail が起動するようになった。

やれやれ...

甘かった

ハマった(甘かった) -その2-

- “service jail start j-1” は成功するんだが...
  “service jail stop j-1” でkernel Panicを起こすんだ...

- なんで？
  - epairをdestroyするタイミングでkernel panicする「ことがある」事が判明
  - epairはbridgeにattachしている...
  - もしかして、bridgeからepairを外せば... → Bingo!

こうなった -2-

```
mweb01-php {
  $id               = "1";
  $ip4_xn1           = "10.1.105.103/24";
  $ip6_xn1           = "fe80::1:101/64";
  $ip4_xn2           = "10.1.106.103/24";
  $ip6_xn2           = "fe80::2:101/64";
  $route4           = "10.1.105.1";
  $route6           = “::1 -blackhole";
  vnet;
  vnet.interface     = "epair${id}1b";
  vnet.interface    += "epair${id}2b";

  exec.prestart      = "/sbin/ifconfig epair${id}1 create up";
  exec.prestart     += "/sbin/ifconfig bridge1     addm epair${id}1a";
  exec.prestart     += "/sbin/ifconfig epair${id}2 create up";
  exec.prestart     += "/sbin/ifconfig bridge2     addm epair${id}2a";

  exec.start         = "/sbin/ifconfig epair${id}1b inet  ${ip4_xn1} up";
  exec.start        += "/sbin/ifconfig epair${id}1b inet6 ${ip6_xn1} up";
  exec.start        += "/sbin/ifconfig epair${id}2b inet  ${ip4_xn2} up";
  exec.start        += "/sbin/ifconfig epair${id}2b inet6 ${ip6_xn2} up";
  exec.start        += "/sbin/route add -inet  default -gateway $
{route4}";
  exec.start        += "/bin/sh /etc/rc";

  exec.stop          = "/bin/sh /etc/rc.shutdown";

  exec.poststop      = "ifconfig bridge1 delete epair${id}1a";
  exec.poststop     += "ifconfig bridge2 delete epair${id}2a";
  exec.poststop     += "ifconfig epair${id}1a destroy";
  exec.poststop     += "ifconfig epair${id}2a destroy";

  persist;
}
```

これだけ踏んだら、まぁ、
もう大丈夫だろう。
やれやれ...

伊勢名物赤福餅
よりも
甘かった!

ハマった(甘かった) -その3-

JL-1のj-1とJL-2のj-2の間で通信ができない

❌ j-1:epair1b <-> j-2 epair1b

❌ j-1:epair2b <-> j-2 epair2b

⭕ JL1-xn1 <-> JL2-xn1

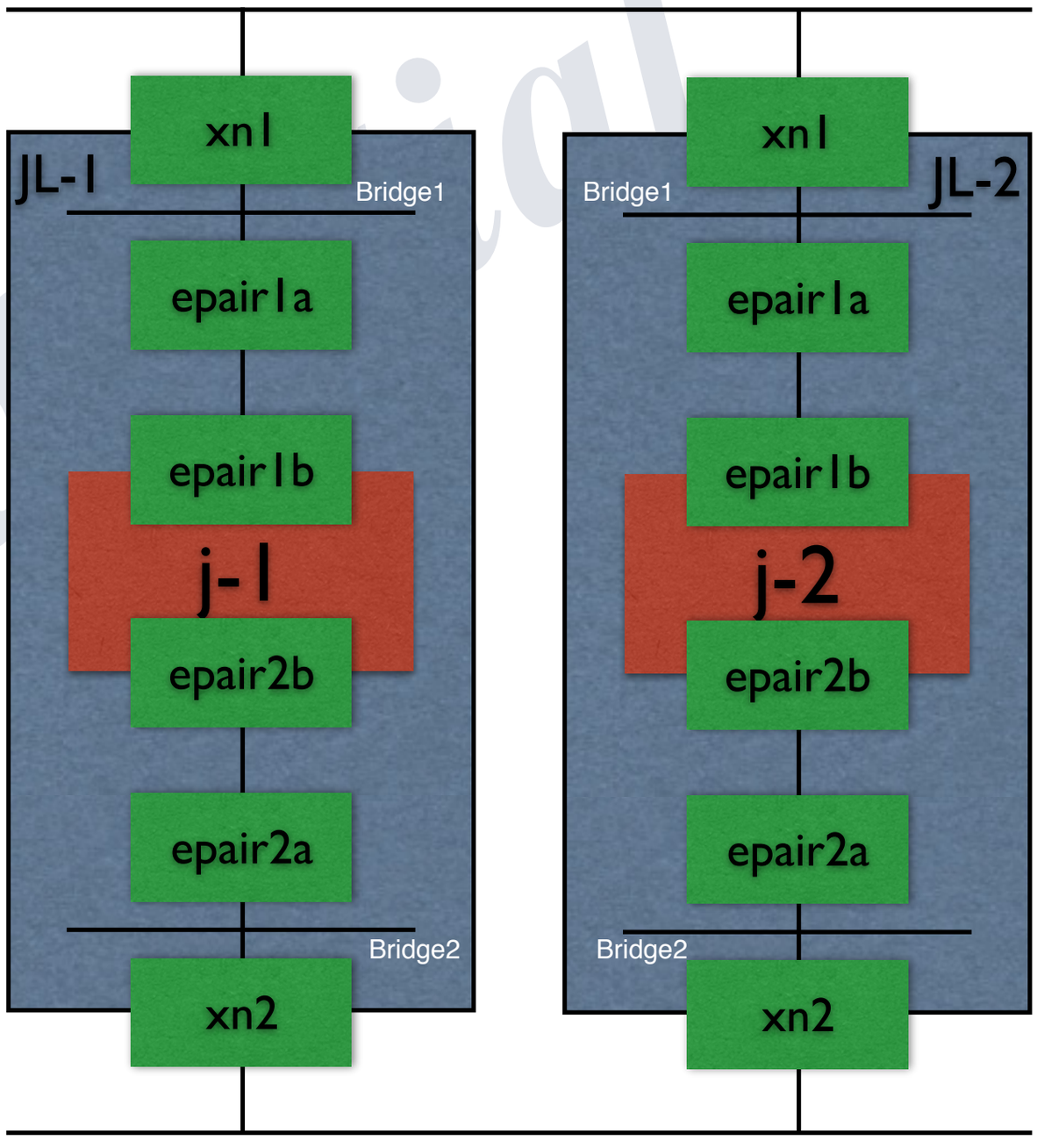
⭕ JL1-xn2 <-> JL2-xn2

❌ JL1-xn1 <-> j2-epair1b

❌ JL1-xn2 <-> j2-epair2b

❌ JL2-xn1 <-> j1-epair1b

❌ JL2-xn2 <-> j1-epair2b



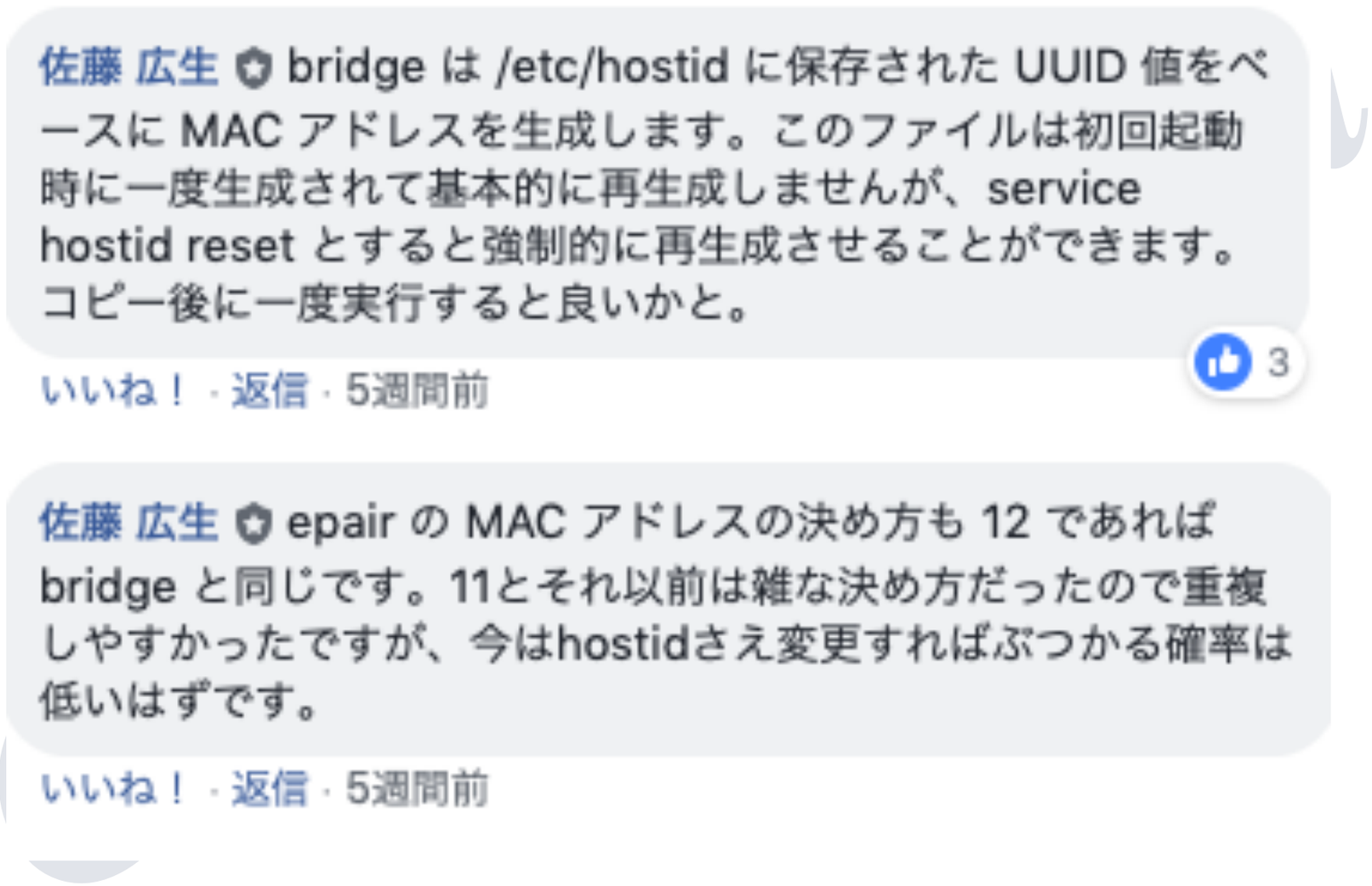
原因がわからない -3-

- 色々調べる
  - j-1のepair1a/1b/2a/2bのMacAddressとj-2のepair1a/1b/2a/2bのMacAddressが同じ？
  - JL-1のbridge11/12とJL-2のbridge11/12のMacAddressが同じ？
  - なぜ？
- Facebookで質問してみる
  - 佐藤先生、内藤さん、浅川さん、小野さん、ありがとうございました
  - 質問の傍、if_bridge.cを見てみる
    - **hostid? なにそれ？どこで定義されてるの？**
- JL-1をCloneしてJL-2を作成したのが祟ってそうだというところまでは解ったけど...
- ifconfig bridge? link xx:xx:xx:xx:xx:xx とかifconfig epair? link xx:xx:xx:xx:xx:xx とかしないとダメなん？**愚か過ぎない**？

わかった! -3-

- /entropy書き換えたらどう？（by 内藤さん）
  - JL-1/JL-2で /entropy の内容は異なっていたので、容疑者から外れる
  - beidge/epairのMacAddressはサイコロ振っているっぽいことはわかった
- とにかく、無理やりMacAddress書き換えたら通信できるのか？
  - できた！＞つまり、これさえ解決すれば.....なんとかなるはず？
- kern.hostidを書き換えたらbridgeのMacAddressが変わるよ
  service hostid resetで書き変わればラッキーかも（by 内藤さん）
  - えっ！マジ？
  - そういえば、if_bridge.cにhostidってあったよなぁ
  - あ、JL-1とJL-2でhostidが等しい！**これだ!!!!**
- やってみたら「**bridgeもepairもMacAddressが衝突しなくなった**」

佐藤先生のサマリー



ああ、やっと理解できた

結論

- Jailを使うときには以下の点に注意しましょう
  - jail.confである変数に複数の値を入れたければ、**”+=“ を利用**する。
    決して ” “ とか ”,” で区切るという考えを持ってはいけない!
  - bridgeを利用してepairとPhys NICを繋ぐなら、jailを停止するときに「明示的に」
    bridgeからepairを外し、それからepairをdestroyするんだ!
  - Jailを収容するマシンを作成するときには、とにかく**hostid**だけは気をつけろ。
    よくわからなかったら**”service hostid reset” を実行**するんだ!
  - どうしてもハマったらFaceBookのFreeBSD研究部に質問するといいかも
  - なお、**bridge に attach されている device を、bridge から detach する前に destroy** したら、ある程度の確率で kernel panic するという疑惑があります。
- (記事かworkshopで話すという約束はこれで果たした...よね？)