Internet Exchanges :

Do we need a new Route Server?

Thomas Mangin
Exa Networks/**IXLeeds**/LINX

Euro-IX 24
16th/18th of March 2014

exa networks

Whatever a speaker is missing in depth he will compensate for in length

Montesquieu

Today's Route Servers Solution

Well known open source implementations of BGP

BIRD	http://bird.network.cz/
Quagga	http://www.quagga.net/
OpenBGPD	http://www.openbgpd.org/

And commercial

CISCO	http://www.cisco.com/

Problem solved !
… or is it really ? …

exa networks

A little learning is a dangerous thing
Alexander Pope

Perhaps
Perhaps not

Implementations have some shared design … :

designed to perform **best path selection**
to then program **ONE RIB**

Users are more demanding, they want :

shared fate between control and data path
more control for filtering or announcement
communities are not a programming language

The community may want to

easily experiment with new technology
such as **SDN** or **ADD-PATH**,…

I have always believed that to succeed in life, it is necessary to appear to be mad and to act wisely

Montesquieu

ADD PATH
Route Server

Possible use of ADD PATH on Route Servers

small IDR thread in November about use for EBGP ADD Path
Nick Hilliard and I expressed support for the idea

Last month: draft-francois-idr-rs-addpaths-00.txt

well received by the IETF / IDR community
one large network expressed to a vendor a need for the feature
many good discussions “around beers”

Still some work required

need to define “mixed environment” operations
could be fast .. in IETF time table terms …

ExaBGP
What is it used for today?

NOC usage ..

DDOS RTBH	prevents bad traffic from reaching its destination
Flow Spec	RTBH on steroid, firewall rules deployed using BGP
Interception	Legal requirements (IWF,…)
SDN	Many deployments

DevOps usage ..

Service IPs	servers mobility using extra/32 with BGP
Anycast	the same IP at different locations (CDN, DNS, ...)

IX usage ..

Collector	at IXLeeds

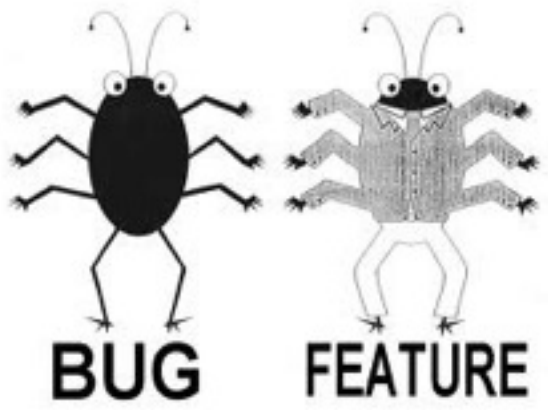
Be regular and orderly in your life, so that you may be violent and original in your work

Flaubert

ExaBGP
What is implemented today?

RFC (fully or mostly fully) implemented

- RFC 1997 - BGP Communities Attribute
- RFC 2385 - Protection of BGP Sessions via the TCP MD5 Signature (for OSes supporting TCP_MD5SIG)
- RFC 2545 - Use of BGP-4 Multiprotocol Extensions for IPv6 Inter-Domain Routing
- RFC 2918 - Route Refresh Capability for BGP-4
- RFC 3107 - Carrying Label Information in BGP-4
- RFC 3765 - NOPEER Community for Border Gateway Protocol (BGP) Route Scope Control
- RFC 4271 - A Border Gateway Protocol 4 (BGP-4), Obsoletes: 1771
- RFC 4360 - BGP Extended Communities Attribute
- RFC 4364 - Constrained Route Distribution for BGP/MPLS IP VPNs
- RFC 4456 - BGP Route Reflection: An Alternative to Full Mesh Internal BGP (IBGP)
- RFC 4659 - BGP-MPLS IP Virtual Private Network (VPN) Extension for IPv6 VPN
- RFC 4724 - Graceful Restart Mechanism for BGP
- RFC 4760 - Multiprotocol Extensions for BGP-4, Obsoletes: 2858
- RFC 4893 - BGP Support for Four-octet AS Number Space
- RFC 5492 - Capabilities Advertisement with BGP-4, Obsoletes 3392,2842
- RFC 5396 - Textual Representation of Autonomous System (AS) Numbers
- RFC 5492 - Capabilities Advertisement with BGP-4
- RFC 5575 - Dissemination of Flow Specification Rules
- RFC 6286 - Autonomous-System-Wide Unique BGP Identifier for BGP-4
- RFC 6608 - Subcodes for BGP Finite State Machine Error



exa networks

Nothing is more humiliating than to see idiots succeed in enterprises we have failed at

Flaubert

ExaBGP
Living on the edge …

- draft-scudder-bmp-01 - BGP Monitoring Protocol v1
- draft-ietf-idr-add-paths-08 - Advertisement of Multiple Paths in BGP
- draft-raszuk-idr-flow-spec-v6-03 - Dissemination of Flow Specification Rules for IPv6
- draft-ietf-idr-bgp-multisession-07 - Multisession BGP
- draft-ietf-idr-flowspec-redirect-ip-00 - BGP Flow-Spec Extended Community for Traffic Redirect to IP Next Hop
- draft-keyur-bgp-enhanced-route-refresh-00 - Enhanced Route Refresh Capability for BGP-4
- draft-ietf-idr-aigp-10 - The Accumulated IGP Metric Attribute for BGP

More about how to use it

http://thomas.mangin.com/data/pdf/

exa networks

I love fools' experiments. I am always making them.
Charles Darwin

ExaBGP
Is the code “production ready”?

Yes, it is used in production in many networks

Content	Facebook, Dailymotion
ISP	updating over 200k routes every 5 minutes ..
Vendor	BGP interoperability for latest drafts

Perfect? unfortunately no.

Latest feature always a bit “rough”

brave souls can use my tree (thomas-mangin)

It seems people like my email address

Please, please open bug reports on github

exa networks

Their is two rules for success in business, one do not tell all you know, …

Some bad joke site

ExaBGP design?

How was ExaBGP designed

- async IO, single threaded loop using windows 3.1 like multi-tasking
- programmed via forked program using PIPE
  - previously using “text interface”
  - introduced a JSON API

```
neighbor 127.0.0.1 {
        router-id 1.2.3.4;
        local-address 127.0.0.1;
        local-as 1;
        peer-as 1;
        graceful-restart;

        process announce-routes {
                run ./api-add-remove.run;
        }
```

```
> ./sbin/exabgp ./api-add-remove.conf
```

```
#!/usr/bin/env python

import sys, time

messages = [
'announce route 1.1.0.0/24 next-hop 101.1.101.1',
'announce route 1.1.0.0/25 next-hop 101.1.101.1',
'withdraw route 1.1.0.0/24 next-hop 101.1.101.1',
]

while messages:
        message = messages.pop(0)
        sys.stdout.write( message + '\n')
        sys.stdout.flush()
        time.sleep(1)

while True:
        time.sleep(1)
```

exa networks

Logic will get you from A to B. Imagination will take you everywhere

Albert Einstein

JSON messages

Text API

announce route 10.0.0.0/8 next-hop 1.1.1.1 as-path [1 2 3 4]
announce route 11.0.0.0/8 next-hop 1.1.1.1 as-path [1 2 3 4]
announce route 12.0.0.0/8 next-hop 1.1.1.1 as-path [1 2 3 4]

No JSON API yet ..


```
{
  "exabgp": "3.3.0",
  "time": 1388767944,
  "neighbor": {
    "ip": "127.0.0.1",
    "update": {
      "attribute": {
        "origin": "igp",
        "as-path": [ 1, 2, 3, 4 ]
      },
      "announce": {
        "1.1.1.1" : {
          "10.0.0.0/8": {  },
          "11.0.0.0/8": {  },
          "12.0.0.0/8": {  }
} } } } }
```

There is some "undocumented" features ..

Much of the configuration format ..
Much of the JSON format ..

There is some "secret" code .. (ie: unsupported yet)

To group announcement per attribute
to improve parsing speed
but it is unsupported ATM

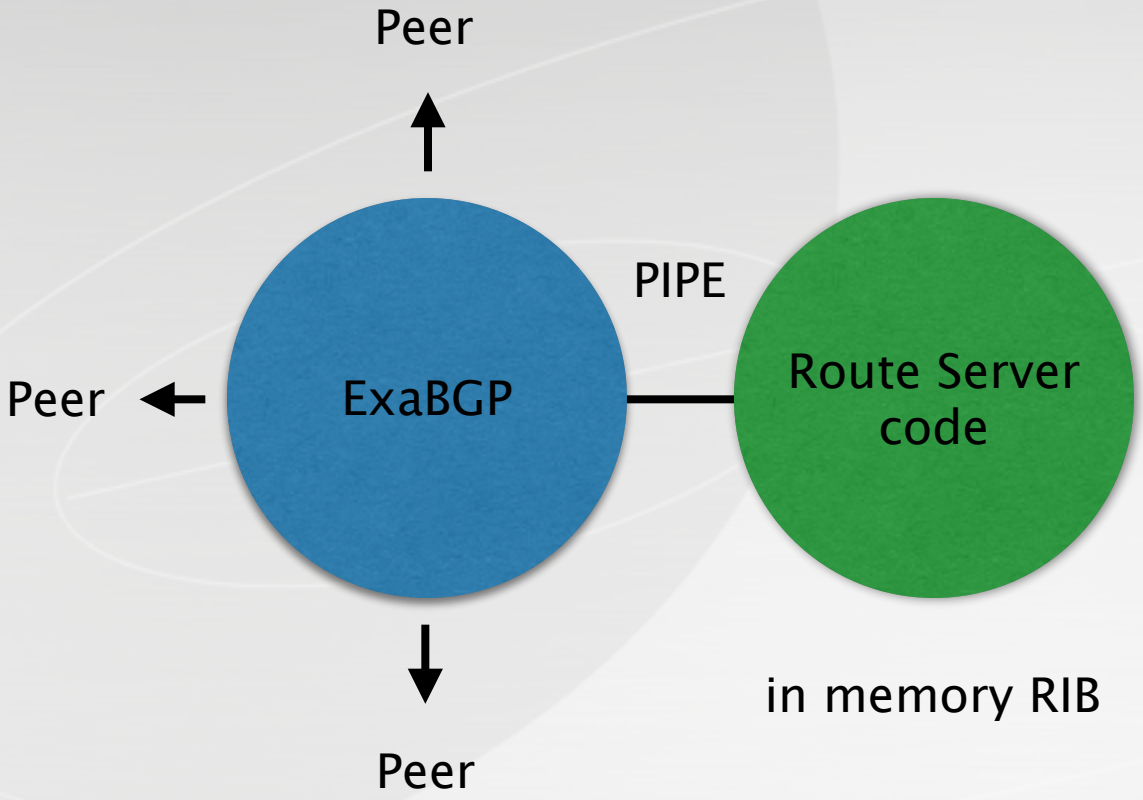
Since / Following Euro-IX 20

Support for incoming connections ..
Many improvements (including performances)

exa networks

Patience is bitter, but its fruit is sweet
Rousseau

“Route Server” backend



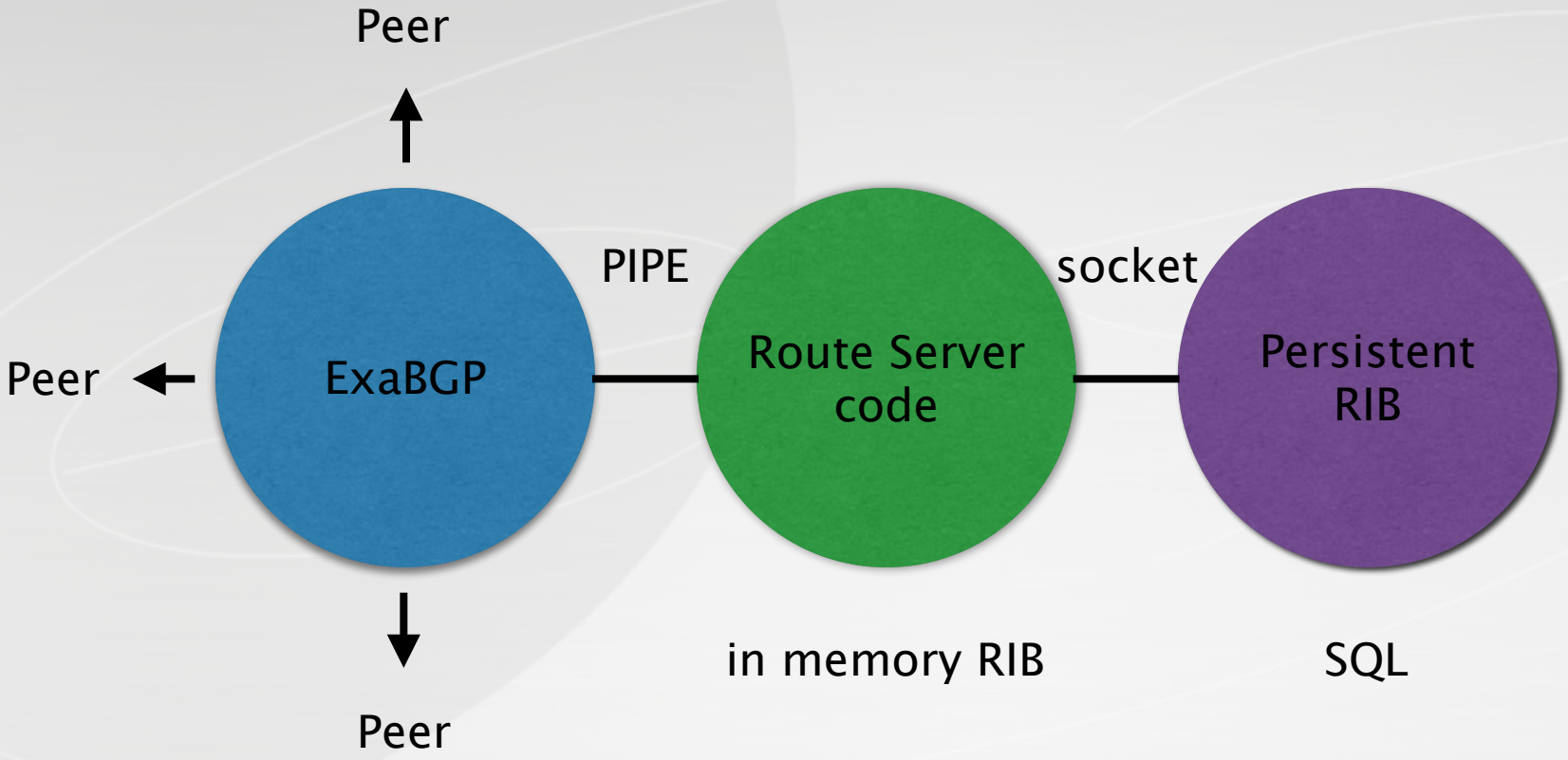
two processes -> two cores used
helps scalability

exa networks

Work delivers us from three great evils: boredom, vice and want.

Voltaire.

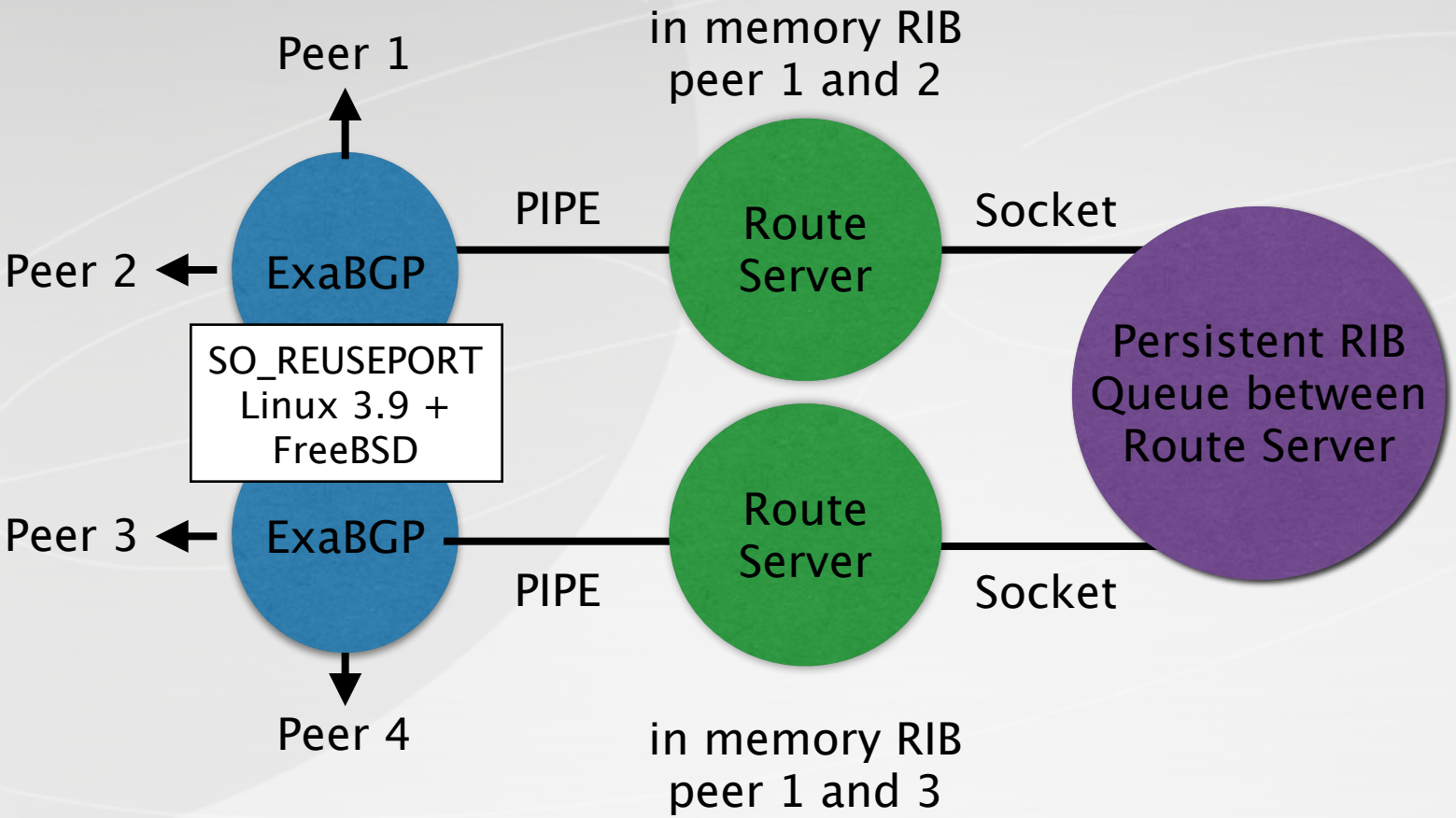
“Route Server” with persistent storage



It is dangerous to be right in matters on which the established authorities are wrong

Voltaire

“Route Server” load balanced



more processes, more cores used

exa networks

The secret of business is to know something that nobody else knows

Aristotle Onassis

Web inspired BGP

At this point this looks like a “LAMP” stack

Protocol Frontal	HTTP -> BGP
Application	APP -> RIB logic
Persistent Storage	(No)SQL -> (No)SQL
	PostgreSQL / Redis / RethinkDB

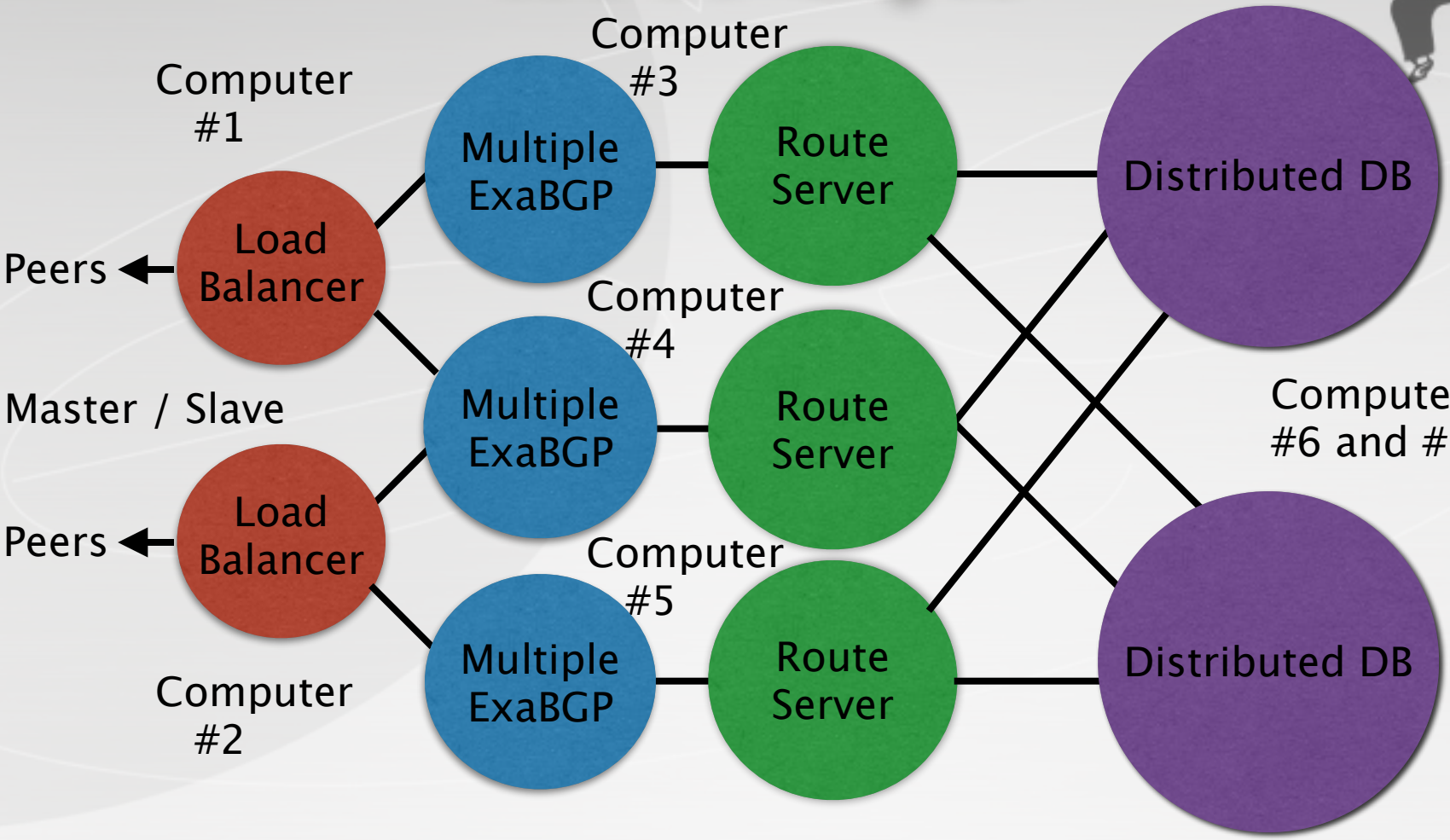
Could run on multiple machines

haproxy as BGP frontal for load balancing between hosts
one local route servers “per site” with distributed DB as RIB
cross monitoring of route servers, BFD at application layer

Work delivers us from three great evils: boredom, vice and want.

Voltaire.

Route Server HTTP Style

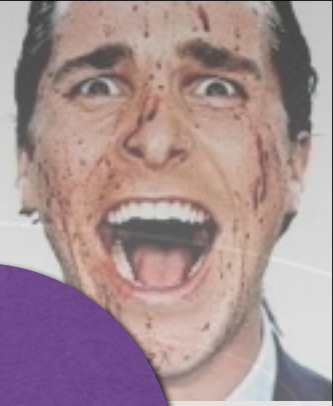


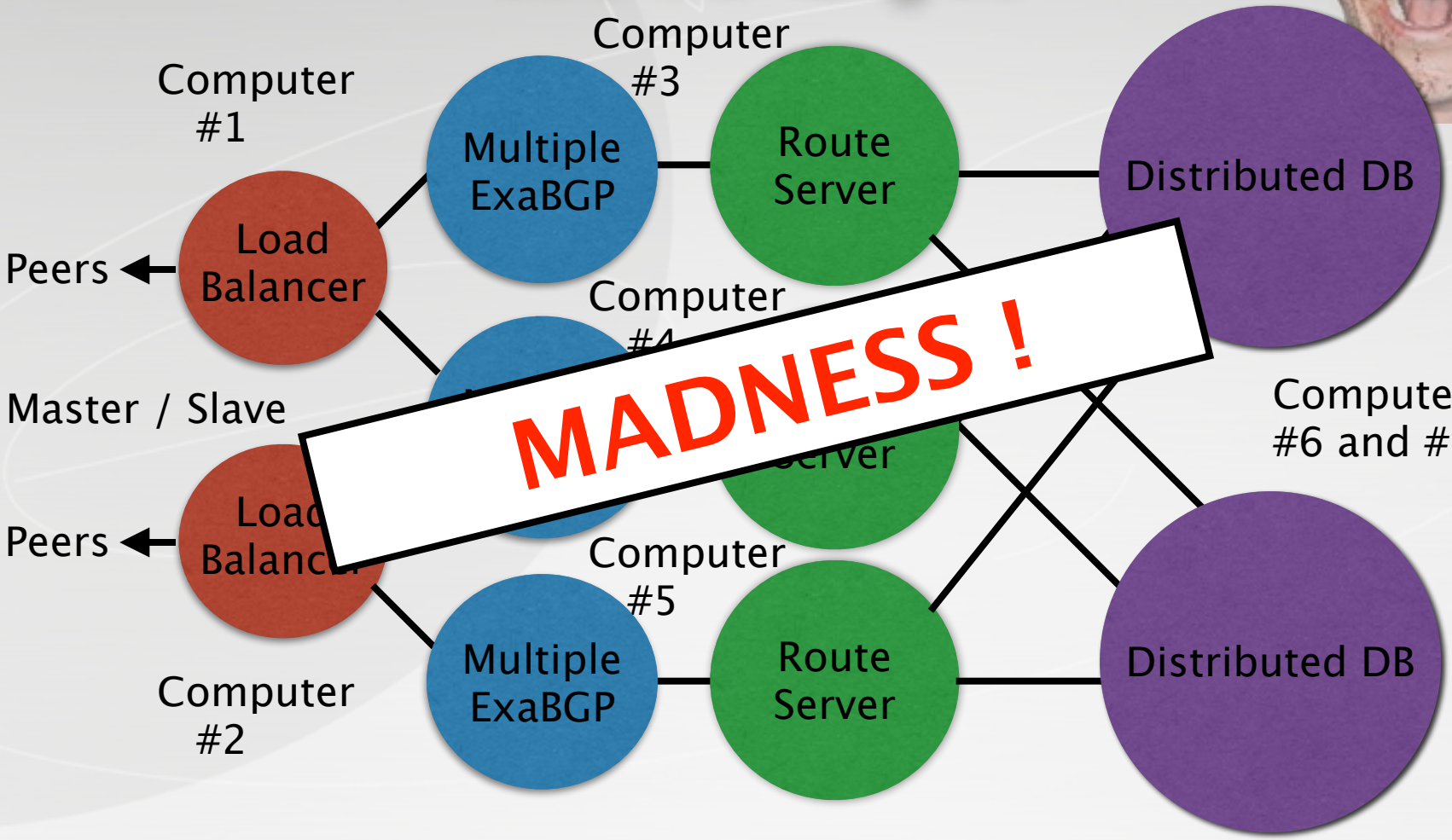
exa networks

The secret of business is to know something that nobody else knows

Aristotle Onassis

Route Server HTTP Style



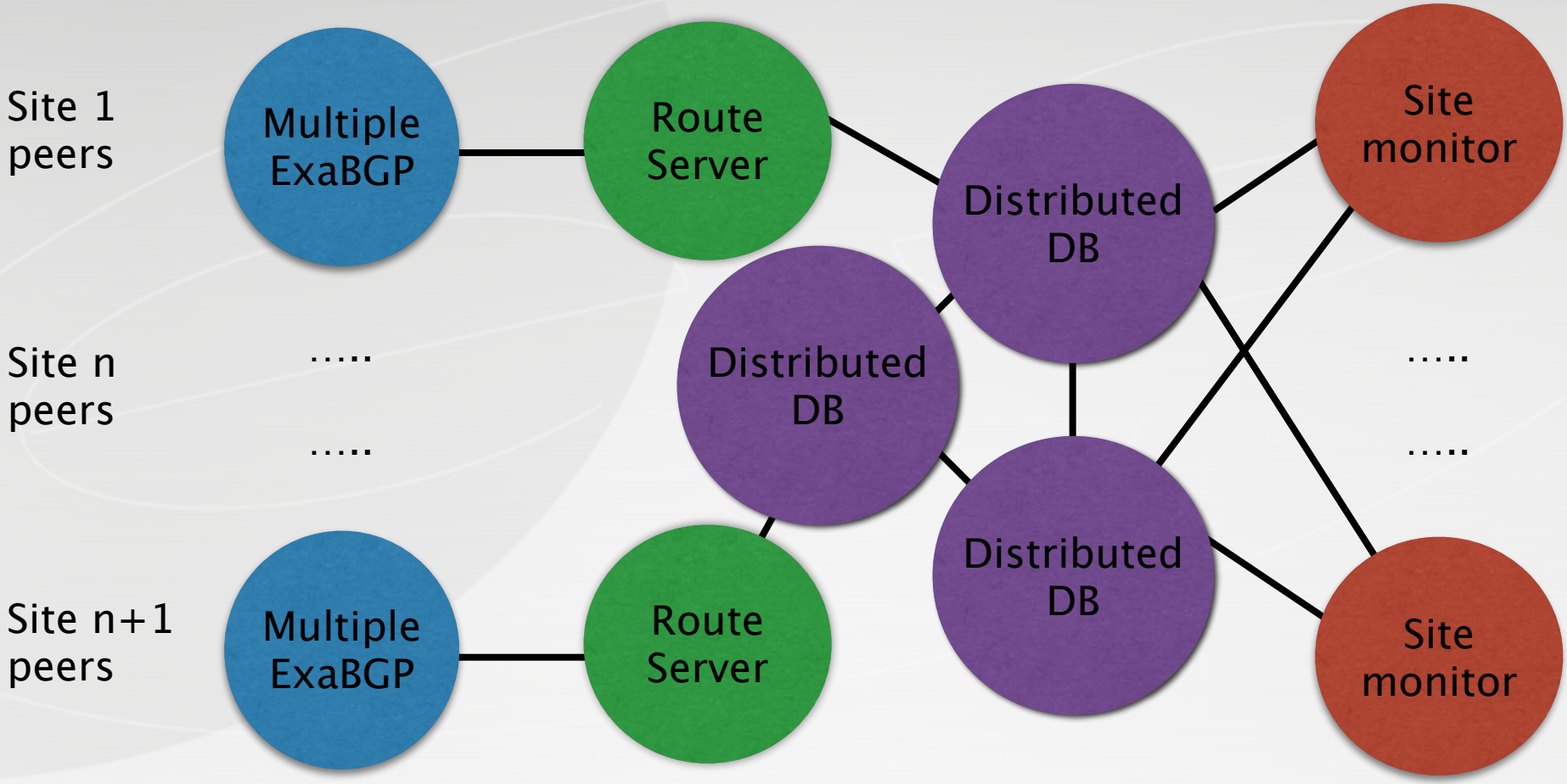


exa networks

The secret of business is to know something that nobody else knows

Aristotle Onassis

“Distributed” Route Server



exa networks

The secret of business is to know something that nobody else knows

Aristotle Onassis

Making ExaBGP Exa Route Server

A route server process

performing per peer **best path selection**
distributing prefix between peers registered

https://github.com/pcamarillo/exabgpRS/blob/master/etc/exabgp/processes/route-server.py

Improve ExaBGP

better control of ExaBGP adding a CLI
to **lookup** peers status
to **admin** peers down/up
the usual suspects …

Be regular and orderly in your life, so that you may be violent and original in your work
Flaubert

Would be nice

More advanced features – let your mind run wild !

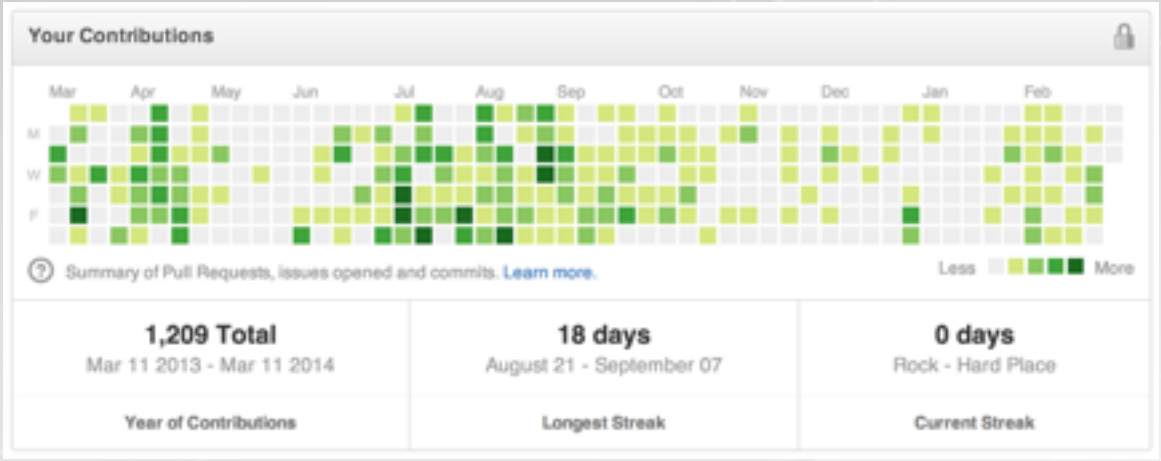
user controlled filtering
SDN with Route Servers

Improvement to ExaBGP and its API

JSON modification to reduce text parsing
JSON require the whole container to be parsed
Prepend JSON data with JSON header

named attributes for reduced ExaBGP chatter
Most NLRI share the same attributes
Name attributes and reference them

when for ?



Depends a lot on **YOU**

Would you **USE** it ?

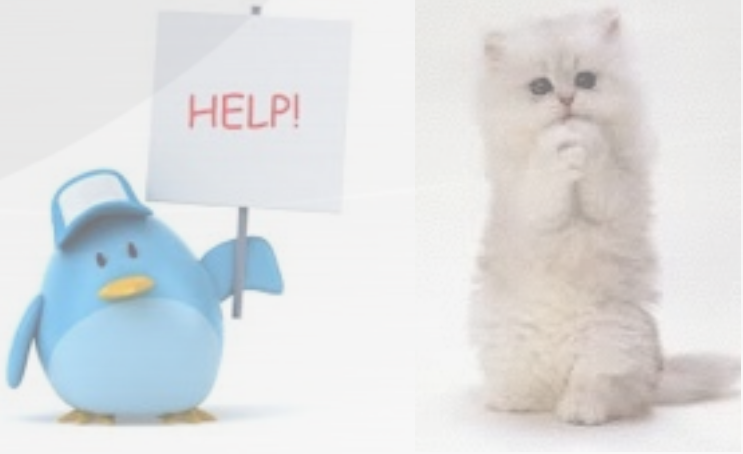
This is my “hobby” with

3 jobs (Exa, IXLeeds, LINX)
2 martial arts
1 “understanding” family
0 regular contributor

QED

I could do with **HELP**

(and on ExaDDOS too)



exa networks

It is easier to ask for forgiveness than permission

Stewart’s law of retraction

Last words… perhaps!

I could do with … many things …

help with the documentation

code testing

Please let me know if you use it for anything

any indication that '**it just work for me**' is always appreciated

my email is also my xmpp account

LINX agreed to let me use their IXIA to see how it performs

and see how it compares to BIRD

who would be interested in seeing the results?

Questions?

Thank you for your kindness on IRC ..

thomas.mangin@exa-networks.co.uk

https://github.com/thomas-mangin/exabgp/

exa networks

Judge a man by his questions rather than by his answers

Voltaire