Anycast on a shoestring

Nat Morris - @natmorris

Tuesday 4th November 2014 - RIPE69, London
About me

• Consultant, **Esgob Ltd**
• Team leader, **Cumulus**
• Board member, **UKNOF**
• Based in West Wales
• Aspiring lighthouse keeper
• Simple concept:
  – Announce the same address space from multiple locations
  – Multiple paths in BGP, best one selected based on policy

• Benefits:
  – Increased reliability
  – Load balancing
  – Improved performance
  – Localized impact of DoS attacks
Summary

• Fancied deploying a DNS Anycast service

• Motivated by:
  – Bill Woodcock, PCH
    • SANOG8: “Best Practices in DNS Anycast Service-Provision”
  – Dave Knight, ICANN (now Dyn)
    • RIPE64: “Dense Anycast Deployment of DNS Authority Servers”

• Gain more experience automating distributed environments.
• Can I do it without spending too much?

• What to offer?
  – Initially only secondary DNS
  – IPv4 + IPv6
  – Free service - no SLA, no Revenue

• Does Claire need to find out?
  Must be sub $1000/yr running cost
Requirements

• Separation from existing management network - AS30746
• Had spare PI /24 + /48
• Applied for new ASN - got AS60564
• Need highly automated framework
• Wanted to play with new tools
• Open source everything
  – https://github.com/esgob
Network

- Each anycast node:
  - Announce /24 + /48 via BGP
  - Static default route to the provider (accept no routes)
• All management traffic inside OpenVPN
  – Additional RFC1918 loopback per node
  – No need for TSIG from AXFR to anycast nodes
  – Protects beanstalk traffic, API, collectd, zone transfers
DNS zone transfers

- Master servers operated by others
- Anycast nodes
  - ns1.esgob.com
- Zone transfer
- Supports notifies from masters

AS30746

ns0.esgob.co.uk (only my zones)

axfr.esgob.com

AS60564
Store everything in **RethinkDB**
- Distributed JSON document database
- Free form, easy to add attributes

```json
{
  "city": "Edinburgh",
  "country": "Scotland",
  "countryiso": "gb",
  "flag": "scotland",
  "fqdn": "alban.esgob.com",
  "geolat": 55.929959,
  "geolng": -3.294011,
  "hostdesc": "VPS at Fluency",
  "id": "alban",
  "livedisplay": "25th Oct 2013",
  "locationdisplay": "Scotland, Edinburgh",
  "mgmtip4": "46.226.2.182",
  "mgmtip6": "2a00:a600:0:6::2",
  "peers4": [
    {
      "account": "esgob",
      "added": 1391807801,
      "id": "esgob.co.uk",
      "masterip": "195.177.253.166",
      "type": "slave"
    },
    {
      "account": "mike",
      "added": 1392340310,
      "id": "grepular.com",
      "masterip": "2001:470:1f15:d7::2",
      "type": "slave"
    }
  ]
}
Application design

- REST API - Python + Flask framework
- cronjob + rsync free zone
- Beanstalk - Message queue
  - Add/change/removal of zones via API triggers job to be added to queue
  - Daemons on AXFR + Anycast nodes wait for jobs
  - New zones live on all nodes with 5 seconds
Where to host the anycast nodes?

- Can’t cost too much each month
  - Ruled out renting co-lo / dedicated servers

- Need a wide reach

- Virtual machines
  - Swap with friends?

- Low cost hardware
Mythic Beasts

• UK clueful hosting company
• £7/month = £84/yr = $141/yr
  – 512mb RAM, 1 CPU, 10gb HDD, 75gb BW
• Bought a VM, opened a support ticket
  – BGP not listed on the website.
  – Me: “I have a /24 + /48 of PI, can you set me up a BGP session? see AS60564 + AS-ESGOB-ANYCAST”
  – Them: “Peer with these IPs, sessions are ready!”
• Too easy!
Quest to find more friendly VM hosts

- LowEndBox.com
  - Blog / adverts for VM providers, all < $10 month
  - Usually OpenVZ based
  - Need KVM/XEN/VMWare to support Quagga/BIRD
- LowEndTalk.com
  - Message board, various small scale VM hosts
- Google
  - “vps bgp session”
Dear Nat Morris (ESGOB LTD),

We have received your order and will be processing it shortly. The details of the order are below:

Order Number: 6758645372

Product/Service: Detroit KVM - DK-512
Domain: esgob

First Payment Amount: $80.00 USD
Recurring Amount: $0.00 USD
Billing Cycle: Annually

Total Due Today: $80.00 USD
• LeapSwitch
  – New
• $120/yr

Dear Nat Morris (Esgob Ltd),

This is a notice that an invoice has been generated on 29/09/2013.

Your payment method is: PayPal

Invoice #44717
Amount Due: $48.00
Due Date: 29/09/2013

Invoice Items

Control Panel: NONE (Default)
FTP Backup: No Backup
IP Address: 0 x IP Addresses (In addition to default 1 IP) $20.00
cPanel Addon: None
Monitoring: No Monitoring

Operating System: Debian 7.0.0 64bit $120.00
Promotional Code: INKVMVPS512MB48 - $72.00 Recurring Discount $-72.00

Sub Total: $48.00
Credit: $0.00
Total: $48.00

You can login to your client area to view and pay the invoice at https://service.leapswitch.com
• Up to 4 VMs
  – London A (on existing KVM server)
  – London B
  – Detroit
  – India
• Spinning up VM taking about 10 minutes
  – Add JSON entry into RethinkDB
  – Install Debian
  – Install Puppet (Collectd, BIND, Quagga, custom daemons etc)
  – Done
  – Pub!
• Offers to host VMs for free from:
  – Edinburgh @ Fluency
  – Boston @ TorwardEx
• One swap:
  – Bremen, Germany @ Fremaks
• Quickly up to 7 nodes after 1 month

<table>
<thead>
<tr>
<th>#</th>
<th>Location</th>
<th>State</th>
<th>Protocols</th>
<th>In service</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>United States, Detroit</td>
<td>Live</td>
<td>IPv4, IPv6</td>
<td>9th Oct 2013</td>
</tr>
<tr>
<td>4</td>
<td>United States, Boston</td>
<td>Live</td>
<td>IPv4, IPv6</td>
<td>13th Oct 2013</td>
</tr>
<tr>
<td>5</td>
<td>India, Pune</td>
<td>Live</td>
<td>IPv4, IPv6</td>
<td>14th Oct 2013</td>
</tr>
<tr>
<td>6</td>
<td>Germany, Bremen</td>
<td>Live</td>
<td>IPv4</td>
<td>23rd Oct 2013</td>
</tr>
</tbody>
</table>
• “We can host something, but not a VM”

• Various issues:
  – Network engineers didn’t have access to VM hosts
  – No VM infrastructure at all
  – All VMs routed by hypervisor, can’t bridge to BGP routers

• Solution...
  – Raspberry PI, $35
  – 512mb / 16gb SD Card
  – 800mhz (overclocked)
• Sent RPI to Belfast – David Farrell @ Tibus
• Ran NSD3 great, ~200qps
• Adding / removing zones with NSD3 required service restart 😞
• Swapped to PowerDNS, testing went ok
• pdns_control segfault’ing on ARM
• Gave in and moved to BIND
Time to peer - Jan 2014

- Offer of a node at SFMIX from Matt Peterson
- San Francisco based IXP
- Couldn’t turn this down.
- Zero U install
- Needed 2 NICs, IX + OOB
  - FitPC2i - perfect 😊
  - Atom 1.6ghz, 1gb RAM, 16gb SSD
  - $150 on eBay
- Peering with:
  - HE.net, ISC, Unwired, PCH
  - Layer42, Lookout, DigitalOcean
Pi problems

- Puppet slow + loading zones taking too long
- More offers to host h/w nodes
  - Manchester, Andy Davidson @ Allegro
- Gigabyte BRIX
  - Quad Core, 1.8ghz, 4gb RAM, 30gb MSATA
- Swap out Belfast Pi
Still growing today

- 12 nodes live, 6 in build

As of 3rd Nov 2014
Fun along the way

• First BGP customer for some of the VM hosts:
  – Assisted educating them...
  – Prefer transit routes over customers, eek
  – No BGP filters / route-maps or prefix lists - Ahhhhh!

• Some hosts don’t have communities
  – Helping people get those implemented

• RIPE Atlas
  – Scheduled measurements to look at latency + CHAOS id.server

• RIPEstat
  – Using the REST API to query visibility of a /48 anchor from each instance, 15mins to get working.
Market for VMs with BGP sessions

You decide what you wish to pay for, no terrifying invoices each month. You configure your server with the specifications you desire, adding the extra addons and then easily pay in advance whenever you are.

**Mini**  
12€/month  
- CPU: 1 CORE  
- RAM: 512 MB  
- Storage (SAN): 50GB  
- Bandwidth: 1000GB

**Medi**  
72€/month  
- CPU: 4 CORES  
- RAM: 4096 MB  
- Storage (SAN): 50GB  
- Bandwidth: 1000GB

### Storage (SAN)
Up to 200GB per drive (multiple hard drives per VPS)

### Offsite Backup
Up to 100GB

### Bandwidth
Up to 100TB

### Private Network
10€/month

### BGP session
10€/month

### Additional IPv4
1€/month for each additional IP

---

**BGP Service**

If you want to assign your own IP ranges to a ZettaGrid Cloud Server then you will need this option. Please note that you will require your own AS (Autonomous System) Number and IP Range.

**Virtual Server**

A dedicated single virtual machine for any purpose, on demand, ready in minutes. Read More.

Processor (CPU)  0.4  
Memory (RAM)  0.25 GB  
Storage  Basic Server - e.g. Web Server (Max 250 IOPS)  100 GB  
Operating System  CentOS Server 6  (0.00)  
Bandwidth  20 GB  
Location  Melbourne  Perth  Sydney

**Summary**

<table>
<thead>
<tr>
<th>Item</th>
<th>Quantity</th>
<th>Unit</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor (CPU)</td>
<td>0.4</td>
<td>ea</td>
<td>0.00</td>
</tr>
<tr>
<td>Memory (RAM)</td>
<td>0.25 GB</td>
<td>ea</td>
<td>0.00</td>
</tr>
<tr>
<td>Storage</td>
<td>100 GB</td>
<td>ea</td>
<td>0.00</td>
</tr>
<tr>
<td>Operating System</td>
<td>CentOS Server 6 (0.00)</td>
<td>ea</td>
<td>0.00</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>20 GB</td>
<td>ea</td>
<td>0.00</td>
</tr>
</tbody>
</table>

**Total Cost**  
$0.00
Discoveries

• Easy to find budget anycasters
  – Webhost forums / LowEndTalk / bgp.he.net

• “Interesting” deployment methods
  – Shared /24
    • Dedicated IP, custom fwd/rev DNS, slaved zones
  – Hosted /24
    • Same as above but customer provides /24 to announce
  – Shared /24, /32 tunneled
    • Dedicated IP, tunneled by anycaster to customer via GRE
  – Single /24
    • Anycaster only has one /24, website + mail + mgmt in same space. DNS answered at edge, other IPs tunneled to another VM/dedicated box.
Discoveries

- Not all budget DNS hosting companies host every zone at the edge
  - Some host zones centrally and cache at the edge
- Some tunnel all DNS traffic back to a location:
  - Looks like anycast
  - Poor DNS performance
  - GRE from edge to a single auth server
  - Selling /32 GRE tunnels to customers!
Not just DNS?

• Simple HTTP/HTTPS service to assist debugging
  – local.esgob.com
  – local4.esgob.com / local6.esgob.com
• Returns JSON via lighttpd on each node

```bash
nat@enw:~ $ curl http://local.esgob.com
{
    "city": "London",
    "country": "England",
    "countryiso": "gb",
    "flag": "england",
    "locationdisplay": "England, London, B",
    "ref": "q17f823b"
}
```
What next?

• Keep honest – host every zone at the edge
• Finish web interface
• Make the Github repos consumable:
  – Produce install instructions
  – Getting started guide
• Support: Multi master, TSIG inbound
• Mix of routing and DNS daemons
  – configurable per node via JSON in RethinkDB
  – BIND, NSD4, KNOT, Quagga, BIRD + ExaBGP
• Looking for friendly hosts:
  – Europe, Africa, India, Asia, South America, anywhere!
As of today...

<table>
<thead>
<tr>
<th>#</th>
<th>Location</th>
<th>State</th>
<th>Protocols</th>
<th>In service</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Los Angeles, US</td>
<td>Maintenance</td>
<td>IPv4, IPv6</td>
<td>18th Jun 2014</td>
<td>Hardware issue being investigated</td>
</tr>
<tr>
<td>4</td>
<td>United States, Detroit</td>
<td>Live</td>
<td>IPv4, IPv6</td>
<td>9th Oct 2013</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>United States, Boston</td>
<td>Live</td>
<td>IPv4, IPv6</td>
<td>13th Oct 2013</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>India, Pune</td>
<td>Live</td>
<td>IPv4, IPv6</td>
<td>14th Oct 2013</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Germany, Bremen</td>
<td>Live</td>
<td>IPv4, IPv6</td>
<td>23rd Oct 2013</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Scotland, Edinburgh</td>
<td>Live</td>
<td>IPv4, IPv6</td>
<td>25th Oct 2013</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Northern Ireland, Belfast</td>
<td>Live</td>
<td>IPv4, IPv6</td>
<td>28th Oct 2013</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>United States, San Francisco</td>
<td>Live</td>
<td>IPv4, IPv6</td>
<td>12th Jan 2014</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>England, Studley</td>
<td>Live</td>
<td>IPv4, IPv6</td>
<td>25th May 2014</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>England, Hemel Hempstead</td>
<td>Live</td>
<td>IPv4, IPv6</td>
<td>26th Sep 2014</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Reykjavik, Iceland</td>
<td>Live</td>
<td>IPv4, IPv6</td>
<td>16th Sep 2014</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>England, Manchester</td>
<td>In Progress</td>
<td>IPv4</td>
<td>16th Sep 2014</td>
<td>Partner to fix BGP filters</td>
</tr>
<tr>
<td>16</td>
<td>Manama, Kingdom of Bahrain</td>
<td>In Progress</td>
<td>IPv4, IPv6</td>
<td></td>
<td>Hardware to be delivered</td>
</tr>
<tr>
<td>17</td>
<td>Vienna, Austria</td>
<td>In Progress</td>
<td>IPv4, IPv6</td>
<td></td>
<td>Commissioning to be delivered</td>
</tr>
<tr>
<td>18</td>
<td>Singapore</td>
<td>In Progress</td>
<td>IPv4, IPv6</td>
<td></td>
<td>Hardware to be delivered</td>
</tr>
<tr>
<td>19</td>
<td>Sydney, Australia</td>
<td>In Progress</td>
<td>IPv4, IPv6</td>
<td></td>
<td>Commissioning underway</td>
</tr>
<tr>
<td>20</td>
<td>England, Fareham</td>
<td>In Progress</td>
<td>IPv4, IPv6</td>
<td></td>
<td>Commissioning underway</td>
</tr>
</tbody>
</table>

Currently costing just under $480/yr - Claire compliant!
**Take aways**

- Low barrier to entry - VM with BGP full table $40/yr

- Don’t become part of the problem
  - Ensure your customers aren’t hijacking prefixes
  - Always use IRR prefix lists on cust BGP sessions

- Using anycast can improve service delivery to your customers

- Automate all things!

- Have fun and share your experiences
Questions?

https://noc.esgob.com

@esgobltd

https://nat.ms

@natmorris