The benefits of BGP for every service provider

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

Montesquieu
I WILL NOT COVER
How to configure a BGP router for general purpose
(But you can grab me after the talk)
What is an IGP (Internal Gateway Protocol)

I ASSUME THAT ...
You have basic networking knowledge (connected, static routes)
Your organisation use some routers you can break
You know what IPs, netmasks, gateways are

I WILL COVER AS MUCH AS I CAN
What is BGP, the Border Gateway Protocol
Why BGP is a great protocol for sysadmins

Truth is more valuable if it takes you a few years to find it.

Renard
A Protocol to share routing information between ISPs

Many RFCs (main one being 4271), many optional features
http://www.bgp4.as/

Open Source implementation in BIRD, Quagga, OpenBGPD

To use it, you do **NOT** need to:

- ✓ be connected to the internet
- ✓ have real world IPs
- ✓ be or ask an ISP anything (but it can be useful)

Use TCP with its own failure detection mechanism.
→ minimum 3s for failure detection

BGP only has one active route for a prefix at a time but the IGP may use multiple paths to get to the next-hop.

There are many true statements about complex topics that are too long to fit on a PowerPoint slide.
Autonomous System Numbers

Unique Network identifier

30740 Exa Networks http://as30740.peeringdb.com/
2856 BT UK http://as2856.peeringdb.com/

initially 16bits, now extended to 32 bits (RFC 4893)
32 bits usage is a negotiated feature

Like RFC 1918, its reserves some IPs

Some ASNs are reserved for documentation (like the 192.0.2.0/24 range)
The range 64496–64511

Some ASNs are reserved for private use
The range 64512–65532

Given to LIR (LOCAL INTERNET REGISTRY)
In the UK, this means RIPE members
does not mean ISP only

A little learning is a dangerous thing
Alexander Pope
BGP transmits Routes

What makes a route

A PREFIX (a block of IP) – the “destination IP regex”
A DESTINATION (called next-hop)
with many optional information (called ATTRIBUTES)
use to select one route over another

The next-hop is a machine that should know how to contact any IP in the prefix, it does not have to be locally connected but just “known”.

Some of the attributes are

LOCAL PREFERENCE, a value to distinguish two 'identical routes'
AS PATH, the chain of ISP who have seen and transmitted the route

BGP will make sure

that the data is always sent to a machine nearer to the end point than itself
that the decision process between multiple routes does not cause loops

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

Albert Einstein
Options for service resilience?

**HSRP, VRRP**
resilience for the gateway, not the host

**Linux-HA solutions** *(Heartbeat, Pacemaker, Wackamole,..)*
Need both machine in the same Layer 2
Lack of IPv6 support!

ARP (relation MAC/IP) expiry 4 to 6 hours ..
MAC (relation ARP/Port) expiry 5 minutes
some kit only allow configuration per interface, not VLAN
enabling gratuitous ARP is a security risk

**Yahoo! L3DSR load balancing solution**
Layer 3 Load Balancing, encoding the destination IP in the DSCP field

**BGP ....**

Be regular and orderly in your life, so that you may be violent and original in your work
— Flaubert
Where does BGP fit?

External BGP: connecting to other networks
- protection from ISP outages

EBGP or IBGP
- Anycast: announce the same IP at different location (CDN, DNS, ...)
- DDOS "mitigation": prevent bad traffic to reach servers
- Flow Routes (firewall rules deployment using BGP)

Internal BGP: fully controlled BGP
- block/redirect some traffic (customers, countries, organisations, ...)
- Servers announcing some Service IPs

I love fools’ experiments. I am always making them.

Charles Darwin
RIPE Membership
Become your own ISP
IPv4 – running out!
do not wait too long if you want to do it!

Provider Aggregate versus Provider Independant
PA: a block of IP owned by the LIR (often the ISP)
changing ISP forces you to renumber
PI: a block of IP owned by the end users
changing ISP is a routing change

Announce your network to the world via BGP
Not as hard as it sounds
Ask your ISP

I have always believed that to succeed in life, it is necessary to appear to be mad and to act wisely
Montesquieu
Split personality ..
   Announcing the same IP with BGP in different location
   Another RFC (4786)
   The network finds the nearest server
   Not best suited for long lived TCP connections
       routing can change

On the internet used by
   Root servers (UDP mainly)

Within a networks
   caching DNS (UDP)
   CDN local DNS (UDP)
   Proxies (TDP, near DSL exit points, very stable routing)
Tell your provider to stop sending you traffic for some IPs

Announce some more specific routes (/32, ...) part of your network and TAG the route with communities so it can be filtered (dropped by the router)

Most useful when you have a public ASN and buy transit Traffic is dropped before it is billed

Many Talks (NANOG, APRICOT, ...) on the topic and an RFC (5635)

> google RTBH or REMOTELY TRIGGERED BLACKHOLE

The goal is to skip the transit provider NOC and NOC response time in time of emergency.

Each ISP implements it differently ..

level3 > whois -h whois.ripe.net AS3356 | grep -B1 -A15 Blackhole

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

Voltaire
Flow Routes

Use BGP to transmit firewall like rules
RFC 5575, Juniper routers only (atm)
Can be used to transproxy in the core things like ... spammers

Match possible components making the flow
Prefix (source and destination)
IP Protocol (list of <action, value>)
Port (source, destination, either)
ICMP (type, code)
TCP flag
Packet Len
DSCP value
Fragment (don't, is, first, last)

Then take action
Drop, Rate-limit, Redirect

exabpg is the only OSS application to support Flow Routes

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

Aristotle Onassis
Block / Redirect traffic

Intercept some traffic injecting BGP routes
the route must be more specific or have an higher LOCAL PREF

Your own IPs
  Move a machine to another geographical location
  connected traffic always preferred to a gateway
  Intercept traffic
  web server (using another server with destination NAT)

Another network IPs
  Block bad sources of traffic: spammers, proxies, TCP scanners, ...
  You are affecting the return packets
  it will not stop a UDP, SYN flood attack
  will prevent TCP 3 way handshake (block the SYN–ACK)

Force outgoing traffic to use one upstream over another
  even if default routes and do not use BGP today

Success is a result, not a goal
Flaubert
Use BGP to announce service IP

An extra IP added to a server for the purpose of providing a public service (ie: pop, imap, web, reverse proxy, vpn IP, ...)

provide IP stability, not physically bound to a location/machine

people SHOULD use DNS entries ... but don’t firewall configuration, etc ...

Have servers announcing their own service IP
Server outage means the IP stops to be routed

Or provision service IPs from a centralised location

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

Montesquieu
Service IPs announcement

Single server
Use **GRACEFUL RESTART** so the router does not forget the route for a programmed number of seconds when BGP goes down unexpectedly.

Active / Passive
Use **LOCAL PREFERENCE** (BGP route preference)
Use `ipvsadm` on the active to still balance traffic

Active/Active
For machine within the same Layer 2, look at using **OSPF**
Otherwise **ANycast** (if suitable)

---

In revolution there are only two sorts of men, those who cause them and those who profit by them

*Napoleon Bonaparte*
Configure IP /32 on the loopback interface, linux (debian/Ubuntu)

/etc/network/interfaces
  auto lo:SERVICE
  iface lo:SERVICE inet static
    address 192.0.2.1
    netmask 255.255.255.255
    network 192.0.2.1
    broadcast 192.0.2.1

Control ARP broadcast (as more than one machine has one IP on its loopback) and RPF check

/etc/sysctl.conf
  net.ipv4.conf.all.arp_filter = 1
  net.ipv4.conf.all.arp_ignore = 1
  net.ipv4.conf.eth0.arp_ignore = 1
  net.ipv4.conf.all.arp_announce = 2
  net.ipv4.conf.eth0.arp_announce = 2
Active / Passive Scenario

Active Server: an exabgp configuration (version 1.2.0+)

GROUP ANNOUNCE-MY-SERVICE-IP-OF-192.0.2.1 {
  # ETH0 10.0.0.1/24 GATEWAY 10.0.0.254 (HSRP/VRRP)
  LOCAL-ADDRESS 10.0.0.1;

  # WE SETUP AN IBGP CONNECTION
  LOCAL-AS 64520;
  PEER-AS 64520;

  STATIC {
    # 150 IS A BETTER LOCAL-PREFERENCE VALUE THAN 100 (DEFAULT VALUE)
    ROUTE 192.0.2.1/32 NEXT-HOP 10.0.0.1 LOCAL-PREFERENCE 150;
  }
  NEIGHBOR 172.16.0.1 {
    DESCRIPTION "BGP ROUTER 1 RUNNING HSRP/VRRP";
  }
  NEIGHBOR 172.16.0.2 {
    DESCRIPTION "BGP ROUTER 2 RUNNING HSRP/VRRP";
  }
}
Active / Passive Scenario

Passive Server: an exabgp configuration (version 1.2.0 +)

GROUP ANNOUNCE-MY-SERVICE-IP-OF-192.0.2.1 {
    # ETH0 10.0.0.2/24 GATEWAY 10.0.0.254 (HSRP/VRRP)
    local-address 10.0.0.2;

    # WE SETUP AN IBGP CONNECTION
    local-as 64520;
    peer-as 64520;

    STATIC {
        # 100 (DEFAULT VALUE) IS A WORSE LOCAL-PREFERENCE VALUE THAN 150
        route 192.0.2.1/32 next-hop 10.0.0.1 LOCAL-PREFERENCE 100;
    }
    NEIGHBOR 172.16.0.1 {
        description "BGP ROUTER 1 RUNNING HSRP/VRRP";
    }
    NEIGHBOR 172.16.0.2 {
        description "BGP ROUTER 2 RUNNING HSRP/VRRP";
    }
}

exa networks
**Active / Passive Scenario**

**Router**: Router 1 (cisco) BGP configuration example

```plaintext
BGP 64520
   NO SYNCHRONIZATION
   BGP ROUTER-ID 172.16.0.1

NEIGHBOR SERVICE-IP PEER-GROUP
NEIGHBOR SERVICE-IP REMOTE-AS 64520
NEIGHBOR SERVICE-IP DESCRIPTION SERVICE IPS
NEIGHBOR SERVICE-IP EBGP-MULTIHOP 5
NEIGHBOR SERVICE-IP UPDATE-SOURCE LOOPBACK1
NEIGHBOR SERVICE-IP DEFAULT-ORIGinate
NEIGHBOR SERVICE-IP ROUTE-MAP BGP-SERVICE-IP IN
NEIGHBOR SERVICE-IP ROUTE-MAP DENY-ANY OUT

NEIGHBOR 10.0.0.1 PEER-GROUP SERVICE-IP
NEIGHBOR 10.0.0.2 PEER-GROUP SERVICE-IP

NO AUTO-SUMMARY
```
Active / Passive Scenario

Router: Router 1 (cisco) BGP configuration example

```console
interface Loopback1
    description BGP
    ip address 172.16.0.1 255.255.255.255

    ip prefix-list service-ip seq 10 permit 192.0.2.1/32
    ip prefix-list service-ip seq 99999 deny 0.0.0.0/0 le 32

    ip access-list standard match-any
        permit any

    route-map bgp-service-ip permit 10
        match ip address prefix-list service-ip
        set community no-export additive

    route-map deny-any deny 10
        match ip address match-any
```

exa networks
Resilience with IPv6

2x Router Advertisement
   -> two default routes

BGP (over an IPv4 or IPv6 TCP connection)
   -> announce the IPv6 service IP

It is easier to ask for forgiveness than permission
   Stewart’s law of retraction
Questions?

Thank you for coming and listening.

Judge a man by his questions rather than by his answers

Voltaire

thomas.mangin@exa-networks.co.uk
http://code.google.com/p/exabgp/