amsix
amsterdam internet exchange

Danger of Proxy ARP in IX environment

version 0.3

Maksym Tulyuk
maksym.tulyuk@ams-ix.net

RIPE 63. Vienna
03 November 2011

Content

1. The Outage
2. Analysis and improvements
3. Useful tools and procedures

Danger of Proxy ARP
in IX environment

amsix
amsterdam internet exchange

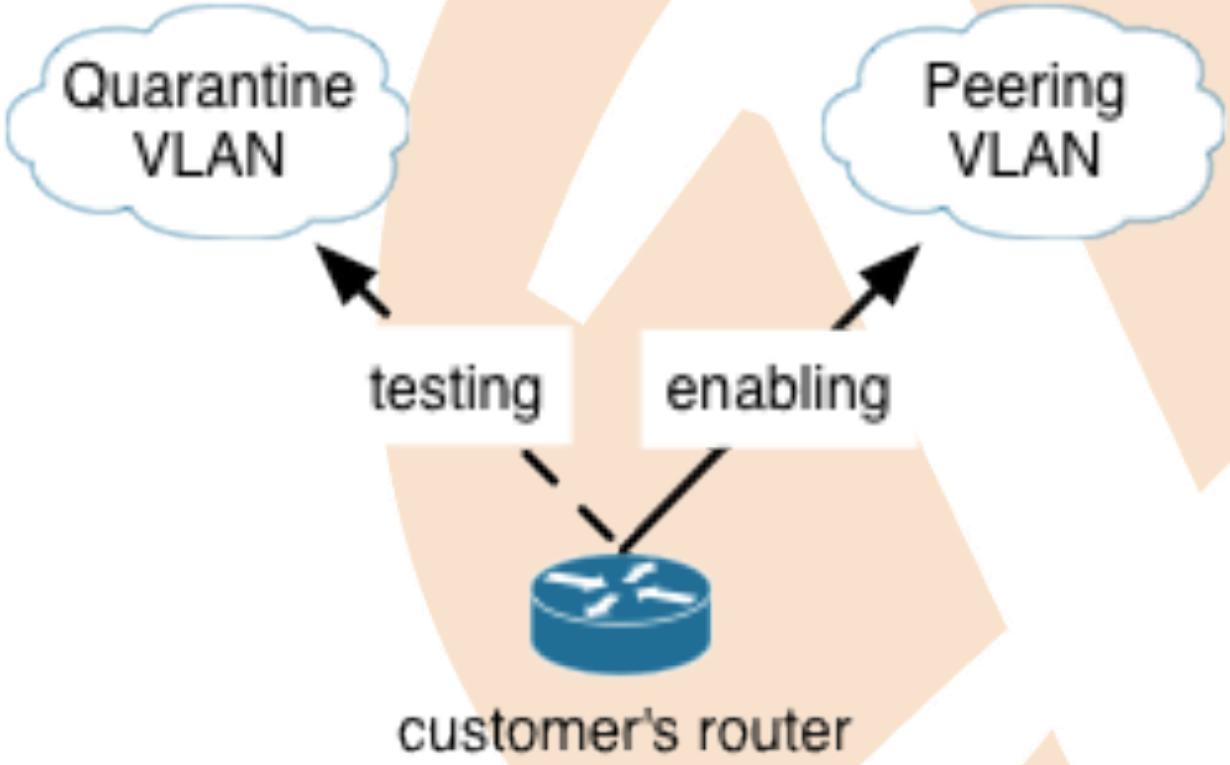
1. The Outage

Maksym Tulyuk
maksym.tulyuk@ams-ix.net

RIPE 63. Vienna
03 November 2011

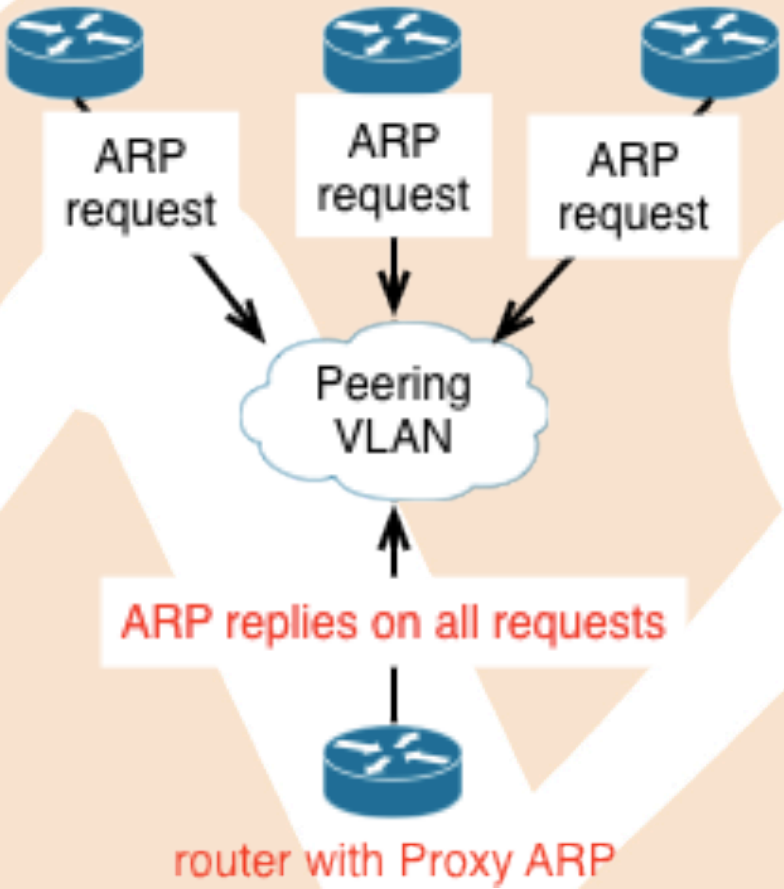
1.1. Customer in production

1. A customer's router was moved from the Quarantine network to the Production network
2. This is a standard procedure was done many times



1.2. Proxy ARP in action

Proxy ARP - router sends ARP replies from own mac address



Proxy ARP conditions

1. Proxy ARP was enabled on a customer's router
2. The router had no IP address from IX range
3. The default route was set on the router

1.3. BGP sessions down

Flapping of BGP sessions was only one visible thing of the accident

```
Aug 11 11:15:19 BGP: Peer 195.69.xxx.xxx DOWN (Rcv Notification:Hold Timer Expired)
Aug 11 11:15:42 BGP: Peer 195.69.xxx.xxx DOWN (Rcv Notification:Hold Timer Expired)
Aug 11 11:16:31 BGP: Peer 195.69.xxx.xxx DOWN (Rcv Notification:Hold Timer Expired)
Aug 11 11:16:58 BGP: Peer 195.69.xxx.xxx DOWN (Rcv Notification:Hold Timer Expired)
Aug 11 11:17:06 BGP: Peer 195.69.xxx.xxx DOWN (Rcv Notification:Hold Timer Expired)
Aug 11 11:19:19 BGP: Peer 195.69.xxx.xxx DOWN (Rcv Notification:Hold Timer Expired)
Aug 11 11:21:42 BGP: Peer 195.69.xxx.xxx DOWN (Rcv Notification:Hold Timer Expired)
Aug 11 11:23:54 BGP: Peer 195.69.xxx.xxx DOWN (Rcv Notification:Hold Timer Expired)
Aug 11 11:24:30 BGP: Peer 195.69.xxx.xxx DOWN (Rcv Notification:Hold Timer Expired)
Aug 11 11:26:30 BGP: Peer 195.69.xxx.xxx DOWN (Rcv Notification:Hold Timer Expired)
Aug 11 11:26:51 BGP: Peer 195.69.xxx.xxx DOWN (Rcv Notification:Hold Timer Expired)
Aug 11 11:27:48 BGP: Peer 195.69.xxx.xxx UP (ESTABLISHED)
Aug 11 11:28:02 BGP: Peer 195.69.xxx.xxx DOWN (Rcv Notification:Hold Timer Expired)
Aug 11 11:28:15 BGP: Peer 195.69.xxx.xxx DOWN (Rcv Notification:Hold Timer Expired)
Aug 11 11:28:36 BGP: Peer 195.69.xxx.xxx DOWN (Rcv Notification:Hold Timer Expired)
Aug 11 11:29:06 BGP: Peer 195.69.xxx.xxx DOWN (Rcv Notification:Hold Timer Expired)
Aug 11 11:29:12 BGP: Peer 195.69.xxx.xxx DOWN (Rcv Notification:Hold Timer Expired)
Aug 11 11:29:17 BGP: Peer 195.69.xxx.xxx UP (ESTABLISHED)
Aug 11 11:30:43 BGP: Peer 195.69.xxx.xxx DOWN (TCP Connection Closed by Remote)
Aug 11 11:31:02 BGP: Peer 195.69.xxx.xxx DOWN (Rcv Notification:Hold Timer Expired)
Aug 11 11:31:39 BGP: Peer 195.69.xxx.xxx DOWN (TCP Connection Closed by Remote)
Aug 11 11:31:39 BGP: Peer 195.69.xxx.xxx DOWN (TCP Connection Closed by Remote)
Aug 11 11:31:39 BGP: Peer 195.69.xxx.xxx DOWN (Rcv Notification:Hold Timer Expired)
Aug 11 11:31:39 BGP: Peer 195.69.xxx.xxx DOWN (TCP Connection Closed by Remote)
```

```
Aug 11 11:14:23 bgpd[6123]: neighbor 195.69.xxx.xxx: received notification: HoldTimer expired, unknown subcode 0
Aug 11 11:14:23 bgpd[6123]: neighbor 195.69.xxx.xxx: state change Established -> Idle, reason: NOTIFICATION received
Aug 11 11:14:24 bgpd[6123]: neighbor 195.69.xxx.xxx: received notification: HoldTimer expired, unknown subcode 0
Aug 11 11:14:24 bgpd[6123]: neighbor 195.69.xxx.xxx: state change Established -> Idle, reason: NOTIFICATION received
Aug 11 11:14:38 bgpd[12947]: neighbor 195.69.xxx.xxx: received notification: HoldTimer expired, unknown subcode 0
Aug 11 11:14:38 bgpd[12947]: neighbor 195.69.xxx.xxx: state change Established -> Idle, reason: NOTIFICATION received
Aug 11 11:14:45 bgpd[12947]: neighbor 195.69.xxx.xxx: received notification: HoldTimer expired, unknown subcode 0
Aug 11 11:14:45 bgpd[12947]: neighbor 195.69.xxx.xxx: state change Established -> Idle, reason: NOTIFICATION received
Aug 11 11:14:52 bgpd[6123]: neighbor 195.69.xxx.xxx: state change Idle -> Active, reason: Start
Aug 11 11:14:53 bgpd[6123]: neighbor 195.69.xxx.xxx: state change Idle -> Active, reason: Start
Aug 11 11:15:07 bgpd[12947]: neighbor 195.69.xxx.xxx: state change Idle -> Active, reason: Start
Aug 11 11:15:15 bgpd[12947]: neighbor 195.69.xxx.xxx: state change Idle -> Active, reason: Start
Aug 11 11:15:46 bgpd[6123]: neighbor 195.69.xxx.xxx: received notification: HoldTimer expired, unknown subcode 0
Aug 11 11:15:46 bgpd[6123]: neighbor 195.69.xxx.xxx: state change Established -> Idle, reason: NOTIFICATION received
Aug 11 11:16:06 bgpd[12947]: neighbor 195.69.xxx.xxx: received notification: HoldTimer expired, unknown subcode 0
Aug 11 11:16:06 bgpd[12947]: neighbor 195.69.xxx.xxx: state change Established -> Idle, reason: NOTIFICATION received
Aug 11 11:16:06 bgpd[12947]: neighbor 195.69.xxx.xxx: received notification: HoldTimer expired, unknown subcode 0
Aug 11 11:16:06 bgpd[12947]: neighbor 195.69.xxx.xxx: state change Established -> Idle, reason: NOTIFICATION received
Aug 11 11:16:16 bgpd[6123]: neighbor 195.69.xxx.xxx: received notification: HoldTimer expired, unknown subcode 0
Aug 11 11:16:16 bgpd[6123]: neighbor 195.69.xxx.xxx: state change Established -> Idle, reason: NOTIFICATION received
Aug 11 11:16:16 bgpd[6123]: neighbor 195.69.xxx.xxx: state change Idle -> Active, reason: Start
Aug 11 11:16:17 bgpd[6123]: neighbor 195.69.xxx.xxx: received notification: HoldTimer expired, unknown subcode 0
Aug 11 11:16:17 bgpd[6123]: neighbor 195.69.xxx.xxx: state change Established -> Idle, reason: NOTIFICATION received
Aug 11 11:16:28 bgpd[12947]: neighbor 195.69.xxx.xxx: received notification: HoldTimer expired, unknown subcode 0
```

1.4. Troubleshooting

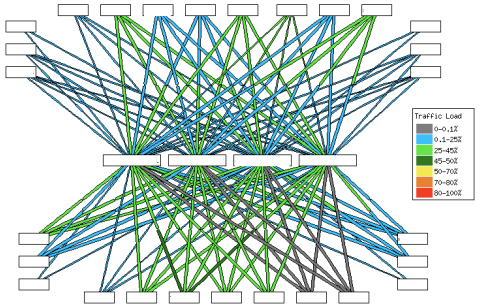
amsix
amsterdam internet exchange

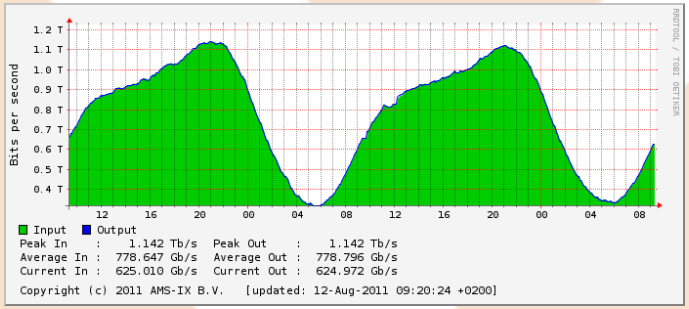
Hypotheses during outage

1. Platform issue?
2. BGP issue?
3. Cisco issue?
4. Route-server issue?

1.4.1. Platform issue?

1. No interfaces down – it's not Physical layer
2. No packet losses – it's not Data Link layer
3. Small drops on graph – it's not Network layer
4. Only BGP sessions down – something on Transport or BGP layer?





1.4.2. BGP issue?

1. At least, 111 BGP peers were down – available statistics only from our equipment (route-servers, routers)
2. 88 peers with route-servers were down
3. 57 peers with our routers were down
4. Most of mac addresses are … **Cisco**

1.4.3. Cisco issue?

1. **90%** affected routers were Cisco
2. Cisco's "bad fame" – Monday's presentation
3. RIPE "experiment":
   https://labs.ripe.net/Members/erik/ripe-ncc-and-duke-university-bgp-experiment
4. 10% routers of **other** vendors

1.4.4. Route-server issue?

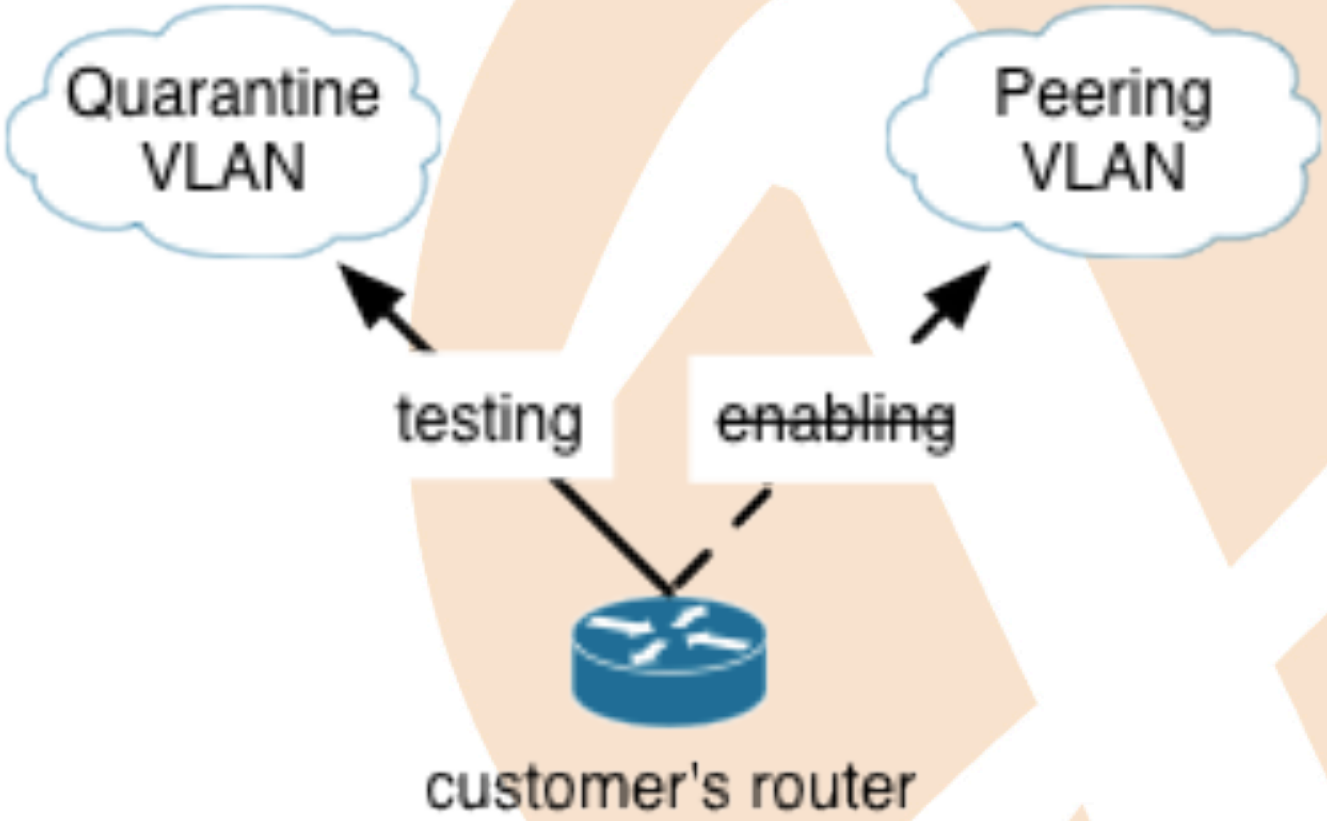
1. Our route-servers **were** unstable
2. /var/log/messages: “arp info overwritten for [IP] by [MAC] on [interface]” – mac address is **the same** all the time
3. Call from a customer: “router with [MAC] **is hijacking** IP addresses!”

```
#sh arp mac-address yyyy.yyyy.yyyy
     IP Address          MAC Address        Type        Age       Port

1     195.69.xxx.xxx      yyyy.yyyy.yyyy      Dynamic     4         1/2
2     195.69.xxx.xxx      yyyy.yyyy.yyyy      Dynamic     5         1/2
3     195.69.xxx.xxx      yyyy.yyyy.yyyy      Dynamic     24        1/2
4     195.69.xxx.xxx      yyyy.yyyy.yyyy      Dynamic     5         1/2
5     195.69.xxx.xxx      yyyy.yyyy.yyyy      Dynamic     0         1/2
6     195.69.xxx.xxx      yyyy.yyyy.yyyy      Dynamic     24        1/2
7     195.69.xxx.xxx      yyyy.yyyy.yyyy      Dynamic     4         1/2
8     195.69.xxx.xxx      yyyy.yyyy.yyyy      Dynamic     5         1/2
9     195.69.xxx.xxx      yyyy.yyyy.yyyy      Dynamic     13        1/2
10    195.69.xxx.xxx      yyyy.yyyy.yyyy      Dynamic     19        1/2
11    195.69.xxx.xxx      yyyy.yyyy.yyyy      Dynamic     8         1/2
12    195.69.xxx.xxx      yyyy.yyyy.yyyy      Dynamic     5         1/2
```

1.5. Disable port

1. The port with the broken router was disconnected
2. It was disconnected 3 (**three**) times – good team work ☺



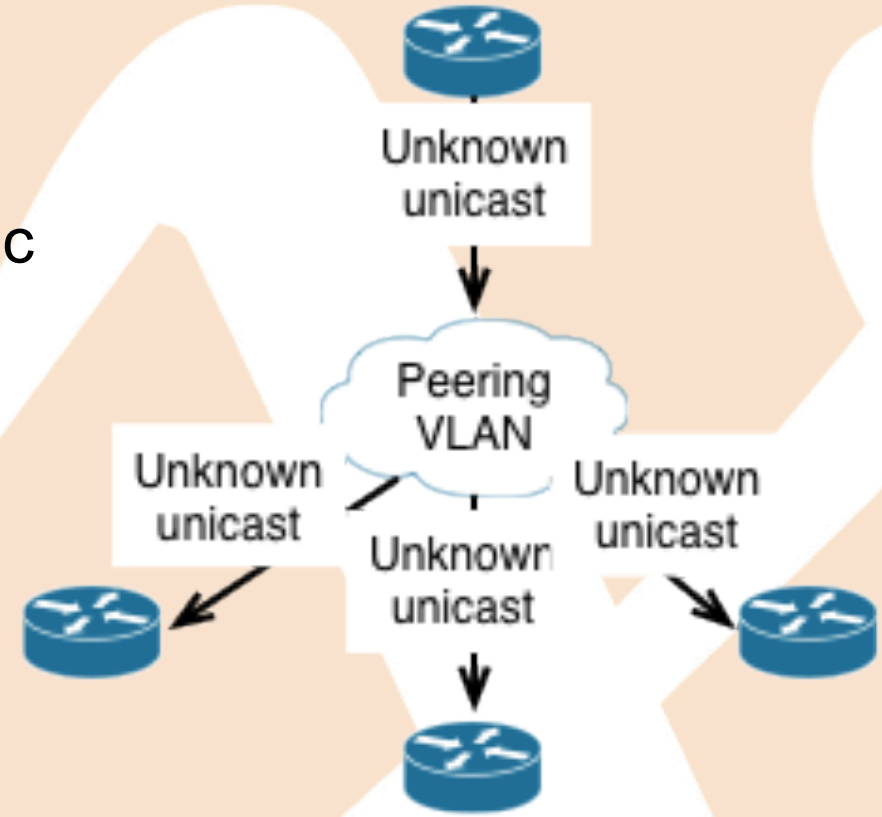
1.6. Calls from customers

1. 10 calls from customers (1 call per 4 min)
2. No network tickets were sent ☹
3. Confirmation that the right router was disconnected

1.7. Unknown unicast flood

Disabling the broken router caused “unknown unicast” flood

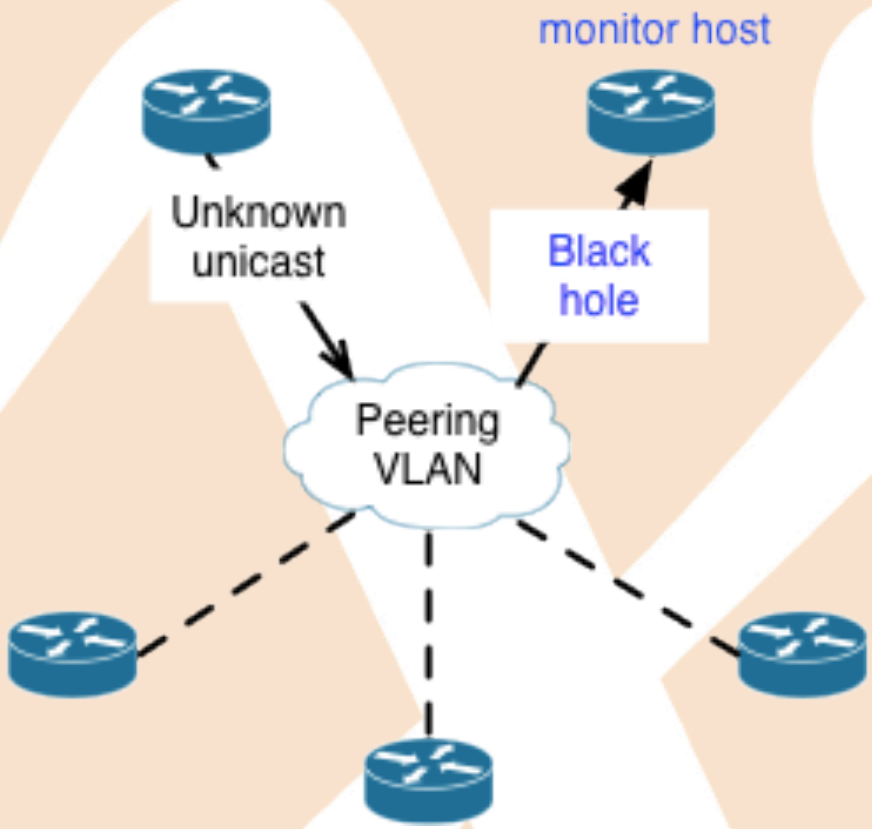
1. Routers still had wrong entries in ARP table
2. Routers were sending traffic to disconnected router



1.7. Unknown unicast flood

Solved by adding wrong mac-address on a monitor host

1. It was quickly detected
2. Didn't affect any customer's routers
3. Made black hole



1.8. Update mac tables

Despite disconnecting the broken router, all affected routers had wrong mac addresses in their ARP tables

Two ways to fix: **1) Reactive approach**

1. Do nothing – TCP/IP already has mechanism to fix
2. When timer “ARP cache timeout” is expired, ARP table will be automatically updated
3. AMS-IX recommends to set **4 hours** for “ARP cache timeout”
4. In the worst case, it would take 4 hours until full restoration

1.8. Update mac tables

Despite disconnecting the broken router, all affected routers had wrong mac addresses in their ARP tables

Two ways to fix: **2) Proactive approach**

1. Send spoofed ARP request (unsolicited ARP reply, gratuitous ARP, ARP request)
2. Proper mac addresses were got from ARP sponge (part of our monitoring system)

1. The Outage

1. Customer in production
2. Proxy ARP in action
3. BGP sessions are down
4. Troubleshooting
5. Disable port
6. Calls from customers
7. Unknown unicast flood
8. Update mac tables

1. The Outage

1. Customer in production
2. Proxy ARP in action
3. BGP sessions are down
4. Troubleshooting
5. Disable port
6. Calls from customers
7. Unknown unicast flood
8. Update mac tables

Downtime:

40 min

Why so long?

amsix
amsterdam internet exchange

2. Analysis and improvements: Root causes

Maksym Tulyuk
maksym.tulyuk@ams-ix.net

RIPE 63. Vienna
03 November 2011

2. Root causes of outage

1. **No tests** for Proxy ARP were made
2. **No confirmation** permanent IP address was set

The issue happened because 3 conditions had been met:

1. **Proxy ARP on customer's router**
2. **No IP address from IX range**
3. **Customer's router had default router**

If any of them wasn't met, the issue wouldn't have happened

2.1. Proxy ARP tests

Issue: the new router wasn't checked for Proxy ARP

Solution 1: Test all new connected routers if Proxy ARP is disabled on them.

Fix: proxy ARP **test was added** to internal procedure; all engineers were informed

Solution 2: Periodic check all connected routers if Proxy ARP is disabled on them

Fix: tool checks connected routers **once per day**

2.2. Confirmation

Issue: no confirmation permanent IP address was set

Idea:

1) Customer sets up an IP address from IX range
2) Engineer checks if the IP address was set
3) The router is moved to the production network

Discussion…

amsix
amsterdam internet exchange

2. Analysis and improvements: Lengthy troubleshooting

Maksym Tulyuk
maksym.tulyuk@ams-ix.net

RIPE 63. Vienna
03 November 2011

2. Lengthy troubleshooting

3. Duty engineer didn't have enough information
4. Customers weren't informed on time
5. No roles between engineers
6. No logs from kernel

In short

1. Internal procedures were not clear enough
2. Tools didn't provide enough functionality

2.3. Full information

Issue: duty engineer didn't have enough information because enabling the new link was done by engineer from **the previous shift**

Fix: everyone **has to inform duty engineer** what and when he is going to do

Discussion: tool to deal with planned works

2.4. On time information

Issue: customers weren't informed on time

Solution: a tool allows to open and send a network ticket quickly and easily

Development:

1. Option in the old interface to send network ticket **quickly**
2. New web interface – under testing

2.5. Outage procedure

Issue: no roles between engineers caused unnecessary delays

Development and discussion…

2.6. Enable logs

Issue: no logs from kernel on syslog server:

/bsd: arp info overwritten for [IP] by [MAC] on [interface]

Fix: route-servers **export to the syslog server** messages from BGP daemon **and** ARP messages

2. Analysis and improvements

Root causes

1. No tests for proxy ARP were made
2. No confirmation permanent IP address was set

Lengthy troubleshooting

3. Duty engineer didn’t have enough information
4. Customers weren’t informed on time
5. No roles between engineers
6. No logs from kernel

2. Analysis of the outage

1. No proxy ARP tests were made
2. No final confirmation from customer
3. Duty engineer didn’t have enough information
4. Customers weren’t informed on time
5. No roles between engineers
6. No logs from kernel

Why **only** 40 min?

amsix
amsterdam internet exchange

3. Tools and procedures

Maksym Tulyuk
maksym.tulyuk@ams-ix.net

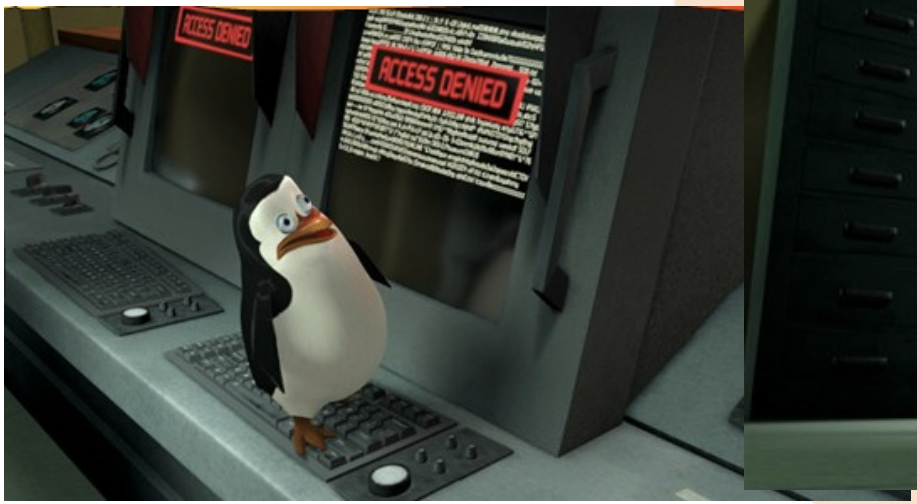
RIPE 63. Vienna
03 November 2011

3.1. Team work

“Never debug alone”:

1. 15 min – 3 engineers were involved
2. 25 min – 5 engineers were involved

Advice: if the issue is going more that 15-30 min, escalate to your colleague/2nd line

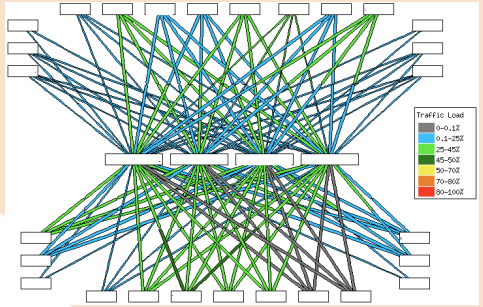


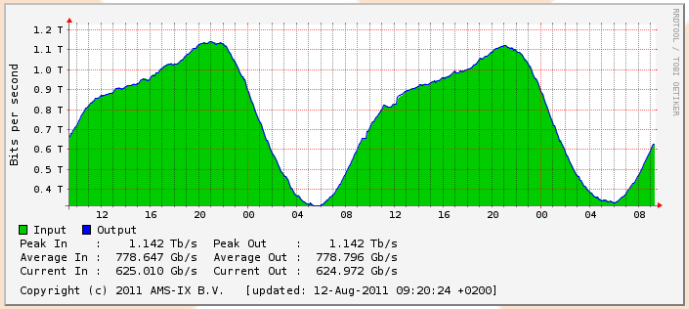
3.2. Central log server

1. Logs from all devices are being sent to a log server
2. The server is available during outage
3. Analysing tools (sed/grep/awk/etc) are installed on server
4. Engineers know how to use them

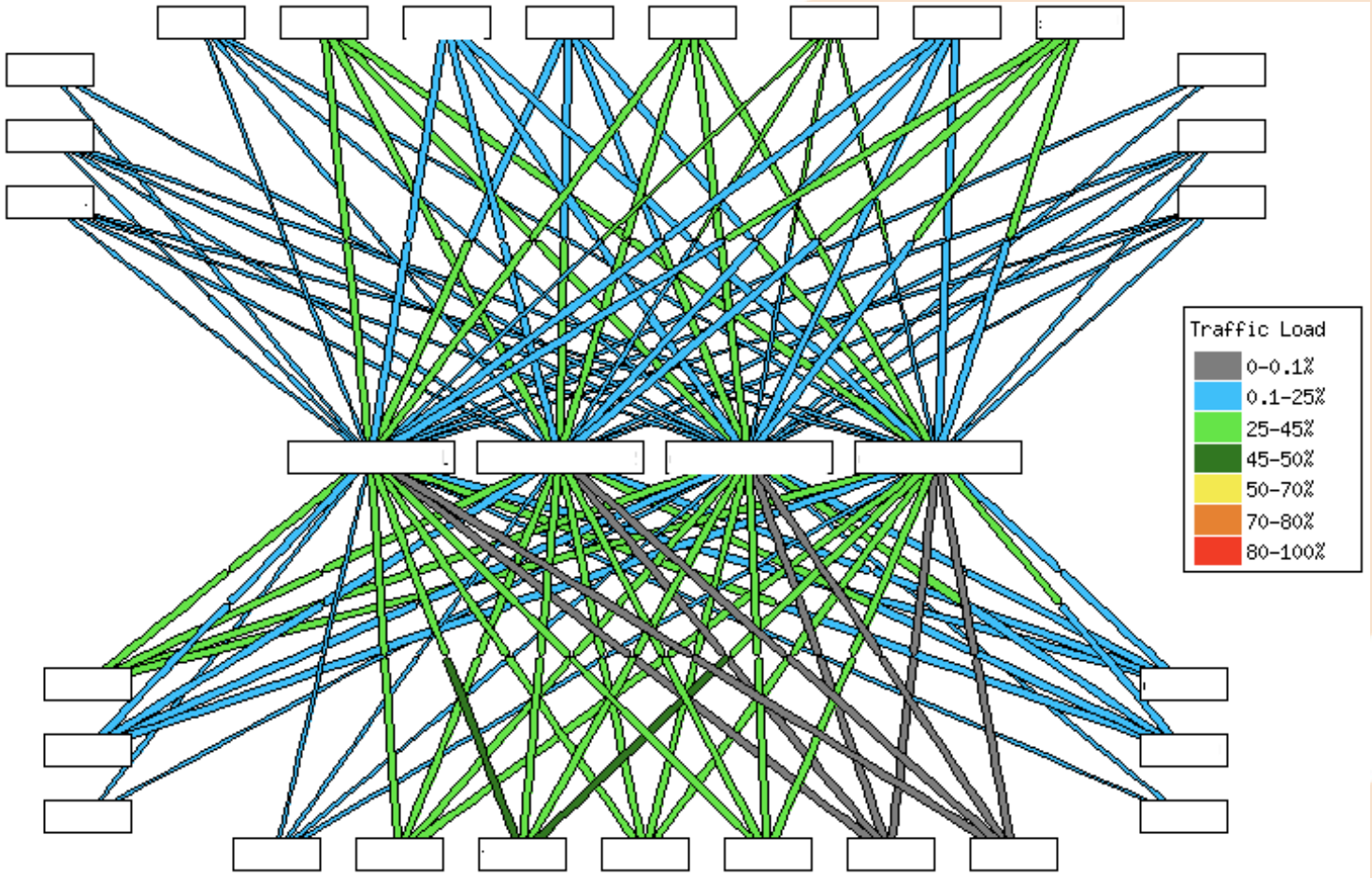
3.3. Visualisation

1. WeatherMap
2. Connectivity matrix
3. MRTG





3.3.1. Weather Map



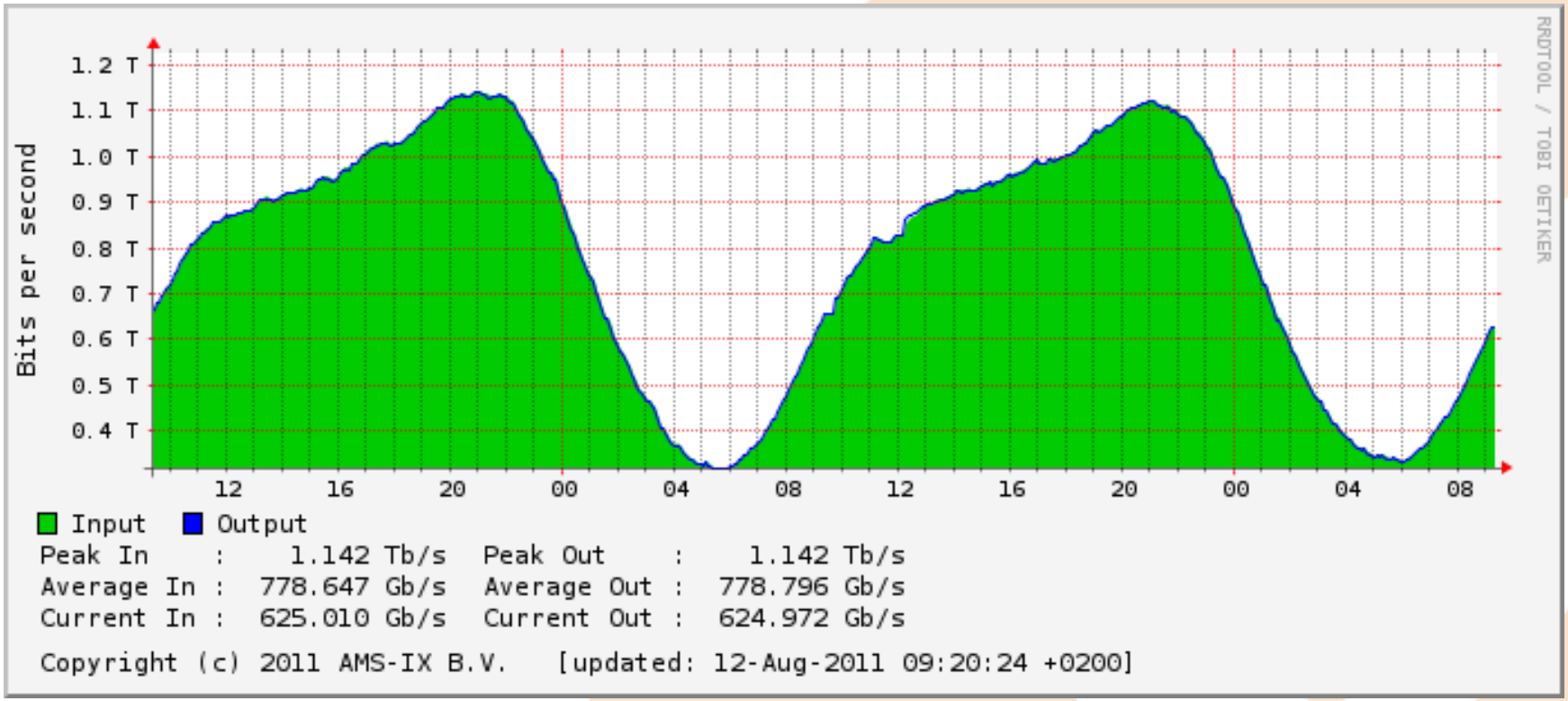
http://www.network-weathermap.com/

3.3.2. Real time matrix

Router-to-router performance: delay, jitter, frame loss

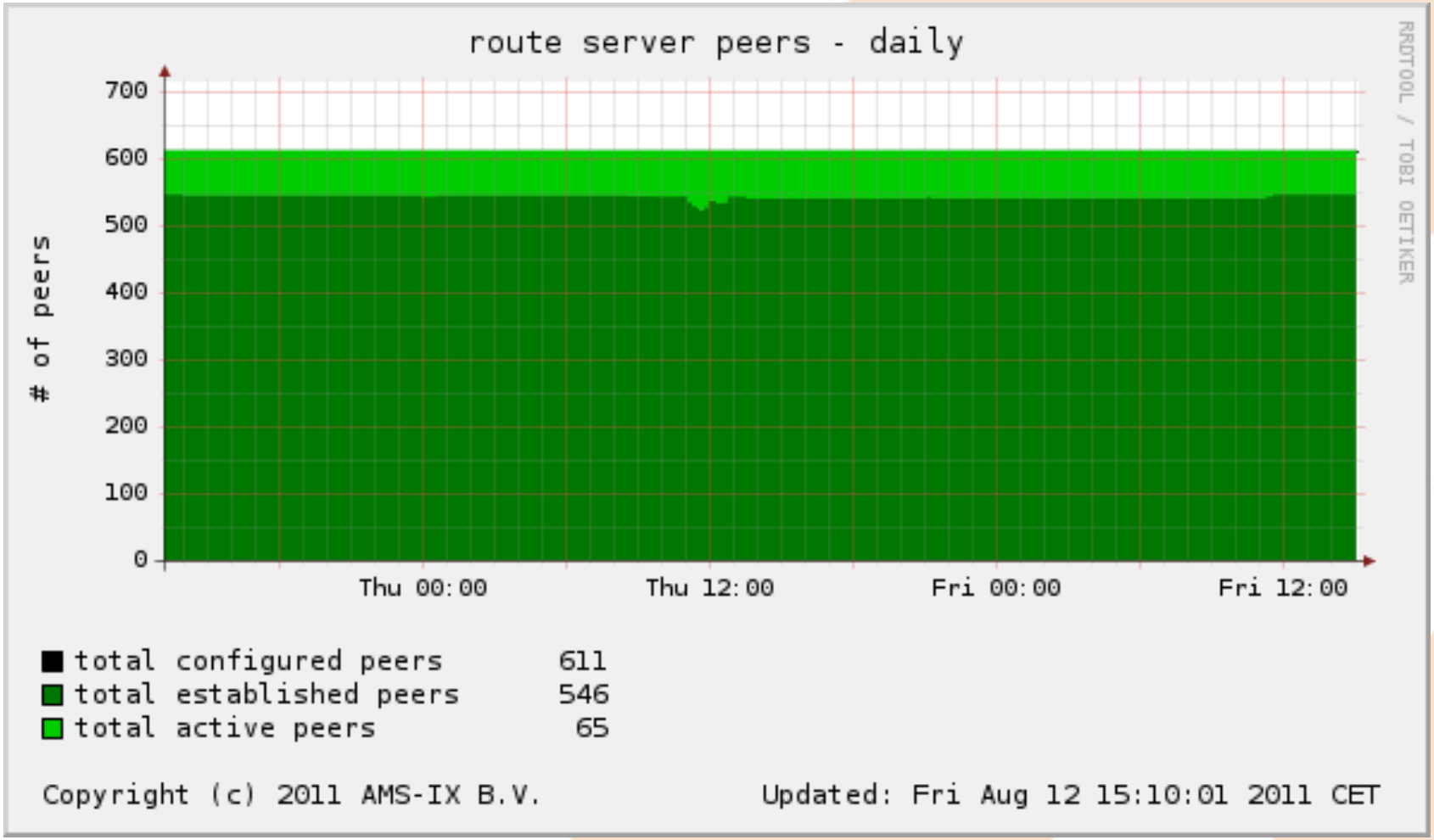
https://www.ams-ix.net/real-time-stats/

3.3.3. MRTG



Daily graph: https://www.ams-ix.net/statistics/

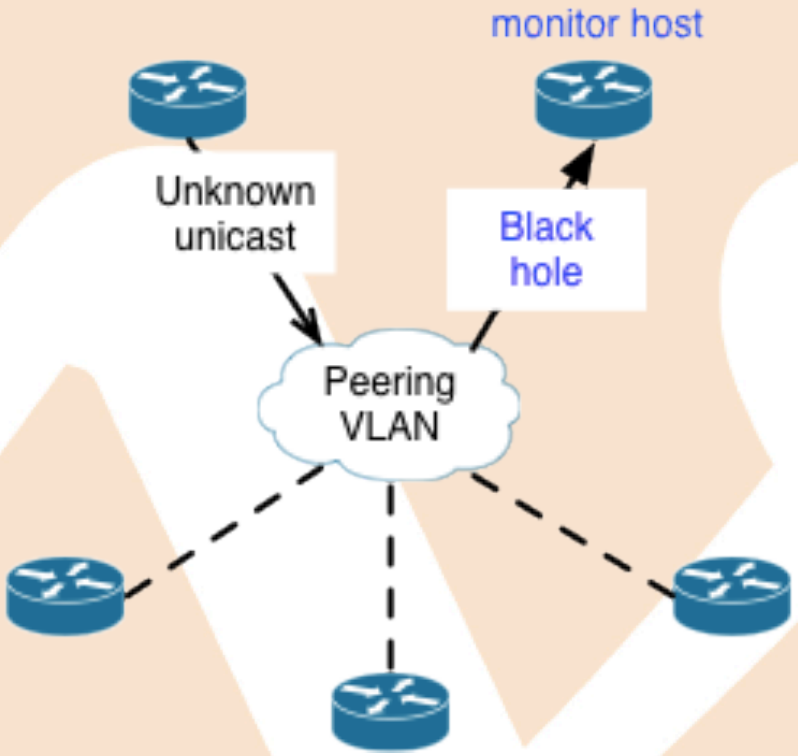
3.3.3. MRTG



Route servers: https://www.ams-ix.net/rs-stats/

3.4. Traffic snooping

1. Host connected to the production network with legal IX address stores all incoming traffic (broadcast, multicast, unknown unicast)
2. Already installed tools: tcpdump, tshark, etc
3. Black hole was made on it
4. Extra: proactive traffic monitoring (STP, CDP, IGMP) through traffic analysis



3.5. ARP sponge

1. ARP sponge – tool to reduce ARP traffic in case of “host is down/flapping”
2. If the sponge sees too many ARP requests it declares the host is down and starts replying with own address
3. ARP sponge monitors all ARP traffic, so it **knows** all mac addresses

More details about ARP sponge:
http://staff.science.uva.nl/~delaat/rp/2008-2009/p23/report.pdf
Source code:
http://www.ams-ix.net/downloads/arpsponge/

3.6. ARP cache update

1. The router with proxy ARP was disconnected but other routers still have wrong ARP records
2. To update their ARP tables (proactive approach) ARP spoof request was sent
3. Another option: Gratuitous ARP or Unsolicited ARP Reply
4. **Addresses were get** from ARP sponge
5. Spoofed ARP packets **were injected** by ARP sponge **also**

More details:
http://www.ams-ix.net/assets/Presentations/Euro-IX-19-ARP-Hijacking-Mitigation.pdf

3. Tools and procedures

Things helps to debug and solve the issue:

1. Team work
2. Central log server
3. Visualisation
4. Traffic snooping
5. ARP sponge
6. ARP cache update

amsix
amsterdam internet exchange

Questions?

Maksym Tulyuk
maksym.tulyuk@ams-ix.net