NOKIA

BGP Monitoring Protocol

NLNOG Day 2018

Henk Smit (Nokia) & Paolo Lucente (NTT Communications)

7 September 2018

Agenda

- Original problem
- Solution: a new protocol: BMP, RFC 7854
- We want moar
- Two proposed drafts from 2017
- A proposal for route-monitoring messages with an extensible message-format

NOKIA

Original problem

- Providers want to know what their network is doing
- In this particular case: what BGP is doing
- Most interesting: received routes and peer state
- Snmp is useless for these large amounts of data
- Netconf could be used, but still it's a lot of routes
  - Netconf wasn't widely used 10 years ago
- We want a "push" solution, not a "pull" solution
  - We need a "telemetry" solution

Original solution

- Use show commands in the CLI
  - Automate through screen-scraping
  - Still a “pull” solution
- Or do BGP-peering to a spot where you want to monitor your router(s)
  - Like a looking-glass server
- Has downsides:
  - Looking-glass server might send you routes by mistake
  - Prone to configuration errors
  - BGP will only send the best-path
    - (can be overcome with add-paths, but that adds complexity)

Solution: a new protocol

- BMP stands for BGP Monitoring Protocol
- RFC7854 (Fernando, Scudder, Stuart)
- Idea is from 2007 or so
- RFC was published June 2016
- Product of the Global Routing Operations Workgroup (grow)
- Simple, efficient
- “Push” solution, not a “pull” solution.
  - No periodic polling.
- Main goal is to just report the routes a router has received from its peers

What is BMP ?

- BGP Monitoring Protocol, to monitor BGP (duh)
- Point-to-point protocol
- Between a router and a BMP-station
  - A “Station” is sometimes called a “Collector”
- Collector is software that runs on a Linux box
- A collector collects events, statistics and routes from BGP
  - Data can be stored in a real data-base
  - Analysis can be done later, at any time
  - Analysis doesn’t consume router resources

The BMP session

- Runs over TCP
  - No well-known port-number. Pick one
  - Can use TCP-keepalives if you want
- Uni-directional
  - Router sends messages to the station
  - Station never ever sends messages to a router
- Simple
  - No hand-shakes, no errors, no state-machine

BMP message types

- Initiation
- Termination
- Peer-up
- Peer-down
- Periodic Statistics Reports
- Route-monitoring
- Route-mirroring

Format of a BMP message

- 6 bytes of BMP header
  - 1 byte protocol version (always 3)
  - 4 bytes of message length
  - 1 byte of message type
- 42 bytes of BMP per-peer header
  - Not for initiation and termination messages
  - Peer-address (ipv4 or ipv6), peer-type, ASN, RD
  - Router-id, timestamp, 8 bits of flags
- Message content
  - BGP Update Message in a Route-monitoring message
  - Counters in a Periodic Statistics Report message
  - OPEN messages in a Peer-up message, etc

Typical life of a BMP session

- Router sets up a TCP connection to the station
- Router sends an Initiation Message
- Send Peer-Up messages for each Established peer
- Send Route-monitoring messages for all received routes
- Send End-Of-RIB messages for all peers, all address-families
- Keep sending Route-monitoring messages when new routes arrive
  - Or withdrawals
- Report peers going down or up via peer-up/down messages
- Maybe send periodic Statistics Reports with counters
- Session ends with a termination message

NOKIA

Examples of BMP Collectors

- pmacct
  - Set of monitoring tools
- OpenBMP
  - Part of toolset called snas.io
- OpenDayLight
- Ryu BMP
- Simple python-scripts (search github)

- Proprietary collectors implemented by hyper-scalers
- Proprietary collectors implemented by router vendors
  - Maybe to feed SDN-controllers

Example configuration for SR-OS

- Configure a bmp-station in the global config

```
configure bmp
  station lys create
    family ipv4 ipv6 vpn-ipv4 label-ipv6
    stats-report-interval 900
    connection
      station-address 192.31.231.16 port 1790
    no shutdown
  no shutdown
```

Example configuration for SR-OS (cont'd)

- Configure which peers you want to monitor

```
configure router bgp
  group internal-peers
    monitor
      station lys braavos myr
      route-monitoring pre-policy post-policy
      no shutdown
```

We want moar !!

- RFC7854 was published in June 2016
- Operators want more
- BMP, like any protocol, can always be improved

- Wish to monitor outgoing routes (Adj-RIB-Out)
- Wish to see best BGP routes (Loc-RIB)

- Want to know why routes were rejected
- Want to know why routes didn’t win best-path selection

NOKIA

Two new BMP drafts

- Draft-ietf-grow-bmp-adj-rib-out-01
  - Allow reporting of outgoing routes, from Adj-RIB-Out
  - Similar to reporting incoming routes
  - Set a bit in the per-peer-header flags-field to distinguish from Adj-RIB-In
  - Two new Periodic Stats Reports counters
- Draft-ietf-grow-bmp-local-rib-00
  - Allow reporting of routes in the BGP Loc-RIB
  - Set peer-type to new value: Loc-RIB Instance Peer
  - Set peer-address to all-zeros

NOKIA

Elegance is not a dispensable luxury but a quality that decides between success and failure

- Edsger W. Dijkstra, 1999
  - Computing Science: Achievements and Challenges
  - https://www.cs.utexas.edu/users/EWD/transcriptions/EWD12xx/EWD1284.html

Limitations of the 2 current proposals

- We only have 8 bits in the peer-flags in the per-peer header
- 4 Bits used now, only 4 bits free for future extensions
  - We still have 249 unused message-types out of potential 256 message-types
- We now know which routes are in the Loc-RIB, but we lost peer information

- We want a solution where we can report all extra state we can think of
- Some state requires a single bit
  - We have only 4 bits left in the per-peer flags field
- Some state requires more information
  - Route-monitoring messages are fixed-format
  - We can't add anything

NOKIA

A new proposal: a new extensible route-monitoring message-format

- Most BMP messages use TLV-based encoding
- Only Route-monitoring messages have a fixed format
  - 6 bytes BMP header
  - 42 bytes per-peer header
  - A full BGP Update Message, including marker, header, attributes and NLRI

- Proposal: use TLV-encoding for the body of a BMP route-monitoring message !

- Requires a new BMP message-type
  - While we're at it, define 3 new message-types:
  - One for Adj-RIB-In, one for Adj-RIB-Out and one for Loc-RIB

NOKIA

Where to find more information

- Draft was published in July 2018
- https://datatracker.ietf.org/doc/draft-hsmit-bmp-extensible-routemon-msgs-00
- New version of the draft will be published soon
  - September or October 2018
  - Backed by Juniper, NTT and hopefully many others

Example of a new BMP route-monitoring message

- Bmp generic header (6 bytes)
- Bmp per-peer header (42 bytes)
- Tlv-header (4 bytes)
- Flags-field content (2 bytes, can be longer)
- Tlv-header (4 bytes)
- BGP update message (marker, header, attributes, NLRI)
- Potentially more TLVs

Flags-field TLV

- Attributes are pre-policy, post-policy, or both
- Route was accepted or rejected by policy
- Route is valid/invalid (e.g. next-hop is unreachable)
- Route is or is not best BGP route after best-path selection
- Route is installed in the general routing table
- Route is best route in the general routing table
- Route is installed in the FIB
- As-path is in 4-byte ASN notation
- NLRI has path-id (add-paths)

Future TLVs

- Tie-break reason why a route did not win best-path selection
- Policy-name or route-map name why a route was rejected
  - Maybe with line-number or entry-number of the exact line in a filter caused rejection

- Got ideas ? What state of a route would you like to see ?

NOKIA

Implementation

- Extensible encoding exists in Nokia's SR-OS today
- But not available to customers (yet)
  - Config command removed from the CLI (and Yang/SNMP)
- Earliest available in 17.0R1 (spring 2019)
  - Ask your friendly Nokia product-manager

- Proposed changes are not very complex
  - So hopefully both BMP-collector implementors and router-vendors can adapt easily
- No need for a configuration-option on the BMP-collector
- Routers need an option to send old-style fixed-format messages (type 0), or send the new tlv-encoded route-monitoring messages (type 7, 8 and 9)

Thank you for your attention

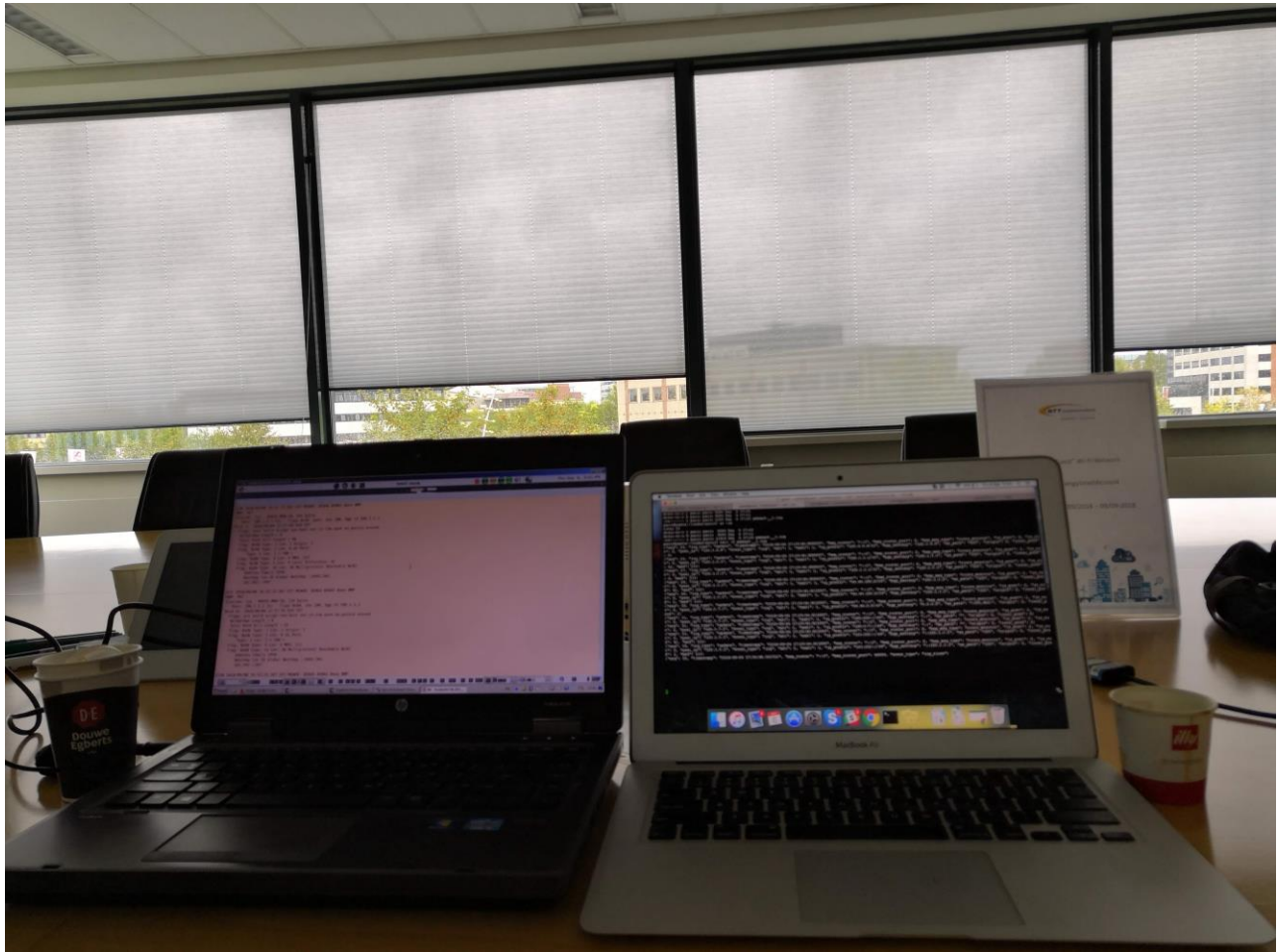
- We hope BMP will be useful for you !

- Contact info:
  - paolo@ntt.net

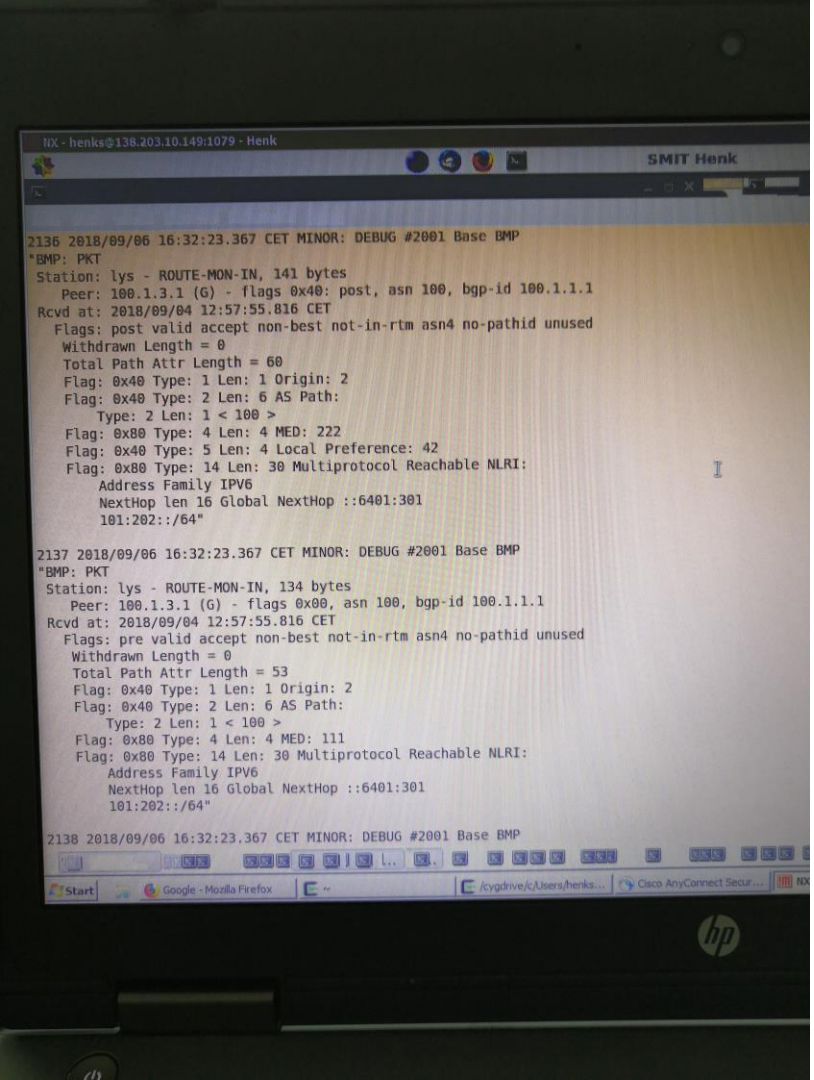
  - henk_hw.smit@nokia.com , or:
  - hhw.smit@xs4all.nl

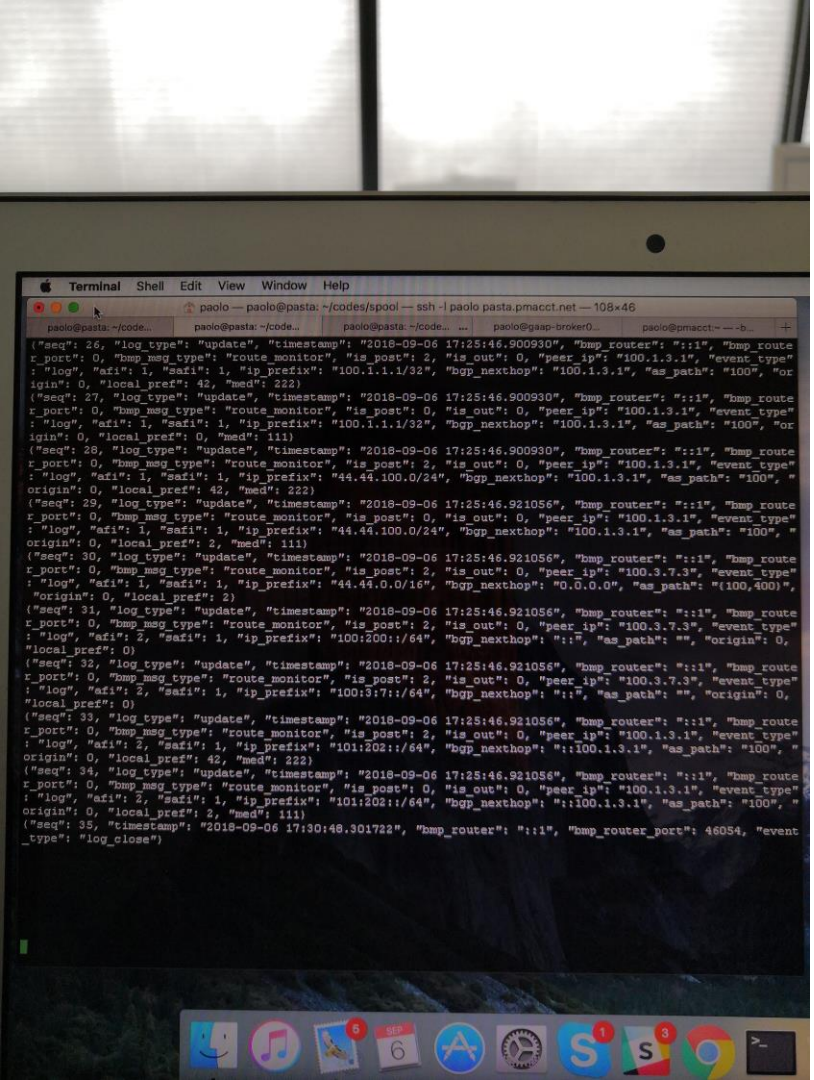
NOKIA

Interoperability



NOKIA


NX - henks@138.203.10.149:1079 - Henk
SMIT Henk
2136 2018/09/06 16:32:23.367 CET MINOR: DEBUG #2001 Base BMP
"BMP: PKT
Station: lys - ROUTE-MON-IN, 141 bytes
Peer: 100.1.3.1 (G) - flags 0x40: post, asn 100, bgp-id 100.1.1.1
Rcvd at: 2018/09/04 12:57:55.816 CET
Flags: post valid accept non-best not-in-rtm asn4 no-pathid unused
Withdrawn Length = 0
Total Path Attr Length = 60
Flag: 0x40 Type: 1 Len: 1 Origin: 2
Flag: 0x40 Type: 2 Len: 6 AS Path:
Type: 2 Len: 1 < 100 >
Flag: 0x80 Type: 4 Len: 4 MED: 222
Flag: 0x40 Type: 5 Len: 4 Local Preference: 42
Flag: 0x80 Type: 14 Len: 30 Multiprotocol Reachable NLRI:
Address Family IPV6
NextHop len 16 Global NextHop ::6401:301
101:202::/64"
2137 2018/09/06 16:32:23.367 CET MINOR: DEBUG #2001 Base BMP
"BMP: PKT
Station: lys - ROUTE-MON-IN, 134 bytes
Peer: 100.1.3.1 (G) - flags 0x00, asn 100, bgp-id 100.1.1.1
Rcvd at: 2018/09/04 12:57:55.816 CET
Flags: pre valid accept non-best not-in-rtm asn4 no-pathid unused
Withdrawn Length = 0
Total Path Attr Length = 53
Flag: 0x40 Type: 1 Len: 1 Origin: 2
Flag: 0x40 Type: 2 Len: 6 AS Path:
Type: 2 Len: 1 < 100 >
Flag: 0x80 Type: 4 Len: 4 MED: 111
Flag: 0x80 Type: 14 Len: 30 Multiprotocol Reachable NLRI:
Address Family IPV6
NextHop len 16 Global NextHop ::6401:301
101:202::/64"
2138 2018/09/06 16:32:23.367 CET MINOR: DEBUG #2001 Base BMP
Start
Google - Mozilla Firefox
/cygdrive/c/Users/henks...
Cisco AnyConnect Secur...


Terminal Shell Edit View Window Help
paolo — paolo@pasta: ~/codes/spool — ssh -l paolo pasta.pmacct.net — 108×46
paolo@pasta: ~/code...
paolo@pasta: ~/code...
paolo@pasta: ~/code...
paolo@gaap-broker0...
paolo@pmacct:~ — -b...
{"seq": 26, "log_type": "update", "timestamp": "2018-09-06 17:25:46.900930", "bmp_router": "::1", "bmp_router_port": 0, "bmp_msg_type": "route_monitor", "is_post": 2, "is_out": 0, "peer_ip": "100.1.3.1", "event_type": "log", "afi": 1, "safi": 1, "ip_prefix": "100.1.1.1/32", "bgp_nexthop": "100.1.3.1", "as_path": "100", "origin": 0, "local_pref": 42, "med": 222}
{"seq": 27, "log_type": "update", "timestamp": "2018-09-06 17:25:46.900930", "bmp_router": "::1", "bmp_router_port": 0, "bmp_msg_type": "route_monitor", "is_post": 0, "is_out": 0, "peer_ip": "100.1.3.1", "event_type": "log", "afi": 1, "safi": 1, "ip_prefix": "100.1.1.1/32", "bgp_nexthop": "100.1.3.1", "as_path": "100", "origin": 0, "local_pref": 0, "med": 111}
{"seq": 28, "log_type": "update", "timestamp": "2018-09-06 17:25:46.900930", "bmp_router": "::1", "bmp_router_port": 0, "bmp_msg_type": "route_monitor", "is_post": 2, "is_out": 0, "peer_ip": "100.1.3.1", "event_type": "log", "afi": 1, "safi": 1, "ip_prefix": "44.44.100.0/24", "bgp_nexthop": "100.1.3.1", "as_path": "100", "origin": 0, "local_pref": 42, "med": 222}
{"seq": 29, "log_type": "update", "timestamp": "2018-09-06 17:25:46.921056", "bmp_router": "::1", "bmp_router_port": 0, "bmp_msg_type": "route_monitor", "is_post": 0, "is_out": 0, "peer_ip": "100.1.3.1", "event_type": "log", "afi": 1, "safi": 1, "ip_prefix": "44.44.100.0/24", "bgp_nexthop": "100.1.3.1", "as_path": "100", "origin": 0, "local_pref": 2, "med": 111}
{"seq": 30, "log_type": "update", "timestamp": "2018-09-06 17:25:46.921056", "bmp_router": "::1", "bmp_router_port": 0, "bmp_msg_type": "route_monitor", "is_post": 2, "is_out": 0, "peer_ip": "100.3.7.3", "event_type": "log", "afi": 1, "safi": 1, "ip_prefix": "44.44.0.0/16", "bgp_nexthop": "0.0.0.0", "as_path": "(100,400)", "origin": 0, "local_pref": 2}
{"seq": 31, "log_type": "update", "timestamp": "2018-09-06 17:25:46.921056", "bmp_router": "::1", "bmp_router_port": 0, "bmp_msg_type": "route_monitor", "is_post": 2, "is_out": 0, "peer_ip": "100.3.7.3", "event_type": "log", "afi": 2, "safi": 1, "ip_prefix": "100:200::/64", "bgp_nexthop": "::", "as_path": "", "origin": 0, "local_pref": 0}
{"seq": 32, "log_type": "update", "timestamp": "2018-09-06 17:25:46.921056", "bmp_router": "::1", "bmp_router_port": 0, "bmp_msg_type": "route_monitor", "is_post": 2, "is_out": 0, "peer_ip": "100.3.7.3", "event_type": "log", "afi": 2, "safi": 1, "ip_prefix": "100:3:7::/64", "bgp_nexthop": "::", "as_path": "", "origin": 0, "local_pref": 0}
{"seq": 33, "log_type": "update", "timestamp": "2018-09-06 17:25:46.921056", "bmp_router": "::1", "bmp_router_port": 0, "bmp_msg_type": "route_monitor", "is_post": 2, "is_out": 0, "peer_ip": "100.1.3.1", "event_type": "log", "afi": 2, "safi": 1, "ip_prefix": "101:202::/64", "bgp_nexthop": "::100.1.3.1", "as_path": "100", "origin": 0, "local_pref": 42, "med": 222}
{"seq": 34, "log_type": "update", "timestamp": "2018-09-06 17:25:46.921056", "bmp_router": "::1", "bmp_router_port": 0, "bmp_msg_type": "route_monitor", "is_post": 0, "is_out": 0, "peer_ip": "100.1.3.1", "event_type": "log", "afi": 2, "safi": 1, "ip_prefix": "101:202::/64", "bgp_nexthop": "::100.1.3.1", "as_path": "100", "origin": 0, "local_pref": 2, "med": 111}
{"seq": 35, "timestamp": "2018-09-06 17:30:48.301722", "bmp_router": "::1", "bmp_router_port": 46054, "event_type": "log_close"}

NOKIA

Paolo’s lunch