

---

# **Knot Resolver**

*Release 3.1.0*

**CZ.NIC Labs**

**Nov 02, 2018**



<b>1</b>	<b>Knot Resolver daemon</b>	<b>3</b>
1.1	Configuration . . . . .	3
1.2	Enabling DNSSEC . . . . .	22
1.3	CLI interface . . . . .	24
1.4	Control sockets . . . . .	24
1.5	Utilizing multiple CPUs . . . . .	24
1.6	Using CLI tools . . . . .	25
<b>2</b>	<b>Knot Resolver modules</b>	<b>27</b>
2.1	Static hints . . . . .	28
2.2	Statistics collector . . . . .	30
2.3	Query policies . . . . .	32
2.4	Views and ACLs . . . . .	37
2.5	Prefetching records . . . . .	38
2.6	HTTP/2 services . . . . .	39
2.7	DNS Application Firewall . . . . .	44
2.8	Rebinding protection . . . . .	46
2.9	Graphite module . . . . .	47
2.10	Etc module . . . . .	47
2.11	DNS64 . . . . .	48
2.12	Renumber . . . . .	49
2.13	DNSSEC validation failure logging . . . . .	49
2.14	Workarounds . . . . .	49
2.15	Dnstap . . . . .	50
2.16	Signaling Trust Anchor Knowledge in DNSSEC . . . . .	50
2.17	Sentinel for Detecting Trusted Root Keys . . . . .	50
2.18	Priming module . . . . .	51
2.19	System time skew detector . . . . .	51
2.20	Detect discontinuous jumps in the system time . . . . .	51
2.21	Root on loopback (RFC 7706) . . . . .	51
2.22	Cache prefilling . . . . .	52
2.23	Serve stale . . . . .	52
<b>3</b>	<b>Building project</b>	<b>53</b>
3.1	Installing from packages . . . . .	53
3.2	Platform considerations . . . . .	53
3.3	Requirements . . . . .	53

3.4	Building from sources . . . . .	55
3.5	Getting Docker image . . . . .	58
<b>4</b>	<b>Knot Resolver library</b>	<b>59</b>
4.1	Requirements . . . . .	59
4.2	For users . . . . .	59
4.3	For developers . . . . .	59
4.4	Writing layers . . . . .	61
4.5	APIs in Lua . . . . .	62
4.6	API reference . . . . .	65
<b>5</b>	<b>Modules API reference</b>	<b>107</b>
5.1	Supported languages . . . . .	107
5.2	The anatomy of an extension . . . . .	107
5.3	Writing a module in Lua . . . . .	108
5.4	Writing a module in C . . . . .	110
5.5	Writing a module in Go . . . . .	110
5.6	Configuring modules . . . . .	112
5.7	Exposing C/Go module properties . . . . .	112
<b>6</b>	<b>Release notes</b>	<b>115</b>
6.1	Knot Resolver 3.1.0 (2018-11-02) . . . . .	115
6.2	Knot Resolver 3.0.0 (2018-08-20) . . . . .	115
6.3	Knot Resolver 2.4.1 (2018-08-02) . . . . .	116
6.4	Knot Resolver 2.4.0 (2018-07-03) . . . . .	117
6.5	Knot Resolver 2.3.0 (2018-04-23) . . . . .	117
6.6	Knot Resolver 2.2.0 (2018-03-28) . . . . .	118
6.7	Knot Resolver 2.1.1 (2018-02-23) . . . . .	118
6.8	Knot Resolver 2.1.0 (2018-02-16) . . . . .	119
6.9	Knot Resolver 2.0.0 (2018-01-31) . . . . .	119
6.10	Knot Resolver 1.5.3 (2018-01-23) . . . . .	120
6.11	Knot Resolver 1.5.2 (2018-01-22) . . . . .	120
6.12	Knot Resolver 1.5.1 (2017-12-12) . . . . .	120
6.13	Knot Resolver 1.5.0 (2017-11-02) . . . . .	121
6.14	Knot Resolver 1.99.1-alpha (2017-10-26) . . . . .	121
6.15	Knot Resolver 1.4.0 (2017-09-22) . . . . .	122
6.16	Knot Resolver 1.3.3 (2017-08-09) . . . . .	122
6.17	Knot Resolver 1.3.2 (2017-07-28) . . . . .	123
6.18	Knot Resolver 1.3.1 (2017-06-23) . . . . .	123
6.19	Knot Resolver 1.3.0 (2017-06-13) . . . . .	123
6.20	Knot Resolver 1.2.6 (2017-04-24) . . . . .	124
6.21	Knot Resolver 1.2.5 (2017-04-05) . . . . .	124
6.22	Knot Resolver 1.2.4 (2017-03-09) . . . . .	125
6.23	Knot Resolver 1.2.3 (2017-02-23) . . . . .	126
6.24	Knot Resolver 1.2.2 (2017-02-10) . . . . .	126
6.25	Knot Resolver 1.2.1 (2017-02-01) . . . . .	126
6.26	Knot Resolver 1.2.0 (2017-01-24) . . . . .	127
6.27	Knot Resolver 1.1.1 (2016-08-24) . . . . .	128
6.28	Knot Resolver 1.1.0 (2016-08-12) . . . . .	128
6.29	Knot Resolver 1.0.0 (2016-05-30) . . . . .	128
<b>7</b>	<b>Indices and tables</b>	<b>129</b>

Knot Resolver is a minimalistic implementation of a caching validating DNS resolver. Modular architecture keeps the core tiny and efficient, and it provides a state-machine like API for extensions.



---

## Knot Resolver daemon

---

The server is in the *daemon* directory, it works out of the box without any configuration.

```
$ kresd -h # Get help
$ kresd -a ::1
```

If you're using our packages, they also provide systemd integration. To start the resolver under systemd, you can use the `kresd@1.service` service. By default, the resolver only binds to local interfaces.

```
$ man kresd.systemd # Help for systemd integration configuration
$ systemctl start kresd@1.service
```

### 1.1 Configuration

- *Configuration example*
- *Configuration syntax*
  - *Dynamic configuration*
  - *Asynchronous events*
- *Configuration reference*
  - *Environment*
  - *Network configuration*
  - *TLS server configuration*
  - *Trust anchors and DNSSEC*
  - *Modules configuration*
  - *Cache configuration*

- *Timers and events*
- *Asynchronous function execution*
- *Scripting worker*

In its simplest form the server requires just a working directory in which it can set up persistent files like cache and the process state. If you don't provide the working directory by parameter, it is going to make itself comfortable in the current working directory.

```
$ kresd /var/cache/knot-resolver
```

And you're good to go for most use cases! If you want to use modules or configure daemon behavior, read on.

There are several choices on how you can configure the daemon, a RPC interface, a CLI, and a configuration file. Fortunately all share common syntax and are transparent to each other.

### 1.1.1 Configuration example

```
-- interfaces
net = { '127.0.0.1', '::1' }
-- load some modules
modules = { 'policy' }
-- 10MB cache
cache.size = 10*MB
```

**Tip:** There are more configuration examples in *etc/* directory for personal, ISP, company internal and resolver cluster use cases.

### 1.1.2 Configuration syntax

The configuration is kept in the `config` file in the daemon working directory, and it's going to get loaded automatically. If there isn't one, the daemon is going to start with sane defaults, listening on *localhost*. The syntax for options is like follows: `group.option = value` or `group.action(parameters)`. You can also comment using a `--` prefix.

A simple example would be to load static hints.

```
modules = {
    'hints' -- no configuration
}
```

If the module accepts configuration, you can call the `module.config({...})` or provide options table. The syntax for table is `{ key1 = value, key2 = value }`, and it represents the unpacked **JSON-encoded** string, that the modules use as the *input configuration*.

```
modules = {
    hints = '/etc/hosts'
}
```

**Warning:** Modules specified including their configuration may not load exactly in the same order as specified.



Modules are inherently ordered by their declaration. Some modules are built-in, so it would be normally impossible to place for example *hints* before *cache*. You can enforce specific order by precedence operators *>* and *<*.

```
modules = {
  'hints > iterate', -- Hints AFTER iterate
  'policy > hints',  -- Policy AFTER hints
  'view < cache'    -- View BEFORE cache
}
modules.list() -- Check module call order
```

This is useful if you're writing a module with a layer, that evaluates an answer before writing it into cache for example.

**Tip:** The configuration and CLI syntax is Lua language, with which you may already be familiar with. If not, you can read the [Learn Lua in 15 minutes](#) for a syntax overview. Spending just a few minutes will allow you to break from static configuration, write more efficient configuration with iteration, and leverage events and hooks. Lua is heavily used for scripting in applications ranging from embedded to game engines, but in DNS world notably in [PowerDNS Recursor](#). Knot Resolver does not simply use Lua modules, but it is the heart of the daemon for everything from configuration, internal events and user interaction.

## Dynamic configuration

Knowing that the the configuration is a Lua in disguise enables you to write dynamic rules. It also helps you to avoid repetitive templating that is unavoidable with static configuration.

```
if hostname() == 'hidden' then
    net.listen(net.eth0, 5353)
else
    net = { '127.0.0.1', net.eth1.addr[1] }
end
```

Another example would show how it is possible to bind to all interfaces, using iteration.

```
for name, addr_list in pairs(net.interfaces()) do
    net.listen(addr_list)
end
```

**Tip:** Some users observed a considerable, close to 100%, performance gain in Docker containers when they bound the daemon to a single interface:ip address pair. One may expand the aforementioned example with browsing available addresses as:

```
addrpref = env.EXPECTED_ADDR_PREFIX
for k, v in pairs(addr_list["addr"]) do
    if string.sub(v,1,string.len(addrpref)) == addrpref then
        net.listen(v)
    end
end
...
```

You can also use third-party packages (available for example through [LuaRocks](#)) as on this example to download cache from parent, to avoid cold-cache start.

```
local http = require('socket.http')
local ltn12 = require('ltn12')
```

(continues on next page)

(continued from previous page)

```

local cache_size = 100*MB
local cache_path = '/var/cache/knot-resolver'
cache.open(cache_size, 'lmbd://' .. cache_path)
if cache.count() == 0 then
    cache.close()
    -- download cache from parent
    http.request {
        url = 'http://parent/data.mdb',
        sink = ltn12.sink.file(io.open(cache_path .. '/data.mdb', 'w'))
    }
    -- reopen cache with 100M limit
    cache.open(cache_size, 'lmbd://' .. cache_path)
end

```

## Asynchronous events

Lua supports a concept called **closures**, this is extremely useful for scripting actions upon various events, say for example - publish statistics each minute and so on. Here's an example of an anonymous function with `event.recurrent()`.

Note that each scheduled event is identified by a number valid for the duration of the event, you may use it to cancel the event at any time.

```

modules.load('stats')

-- log statistics every second
local stat_id = event.recurrent(1 * second, function(evid)
    log(table_print(stats.list()))
end)

-- stop printing statistics after first minute
event.after(1 * minute, function(evid)
    event.cancel(stat_id)
end)

```

If you need to persist state between events, encapsulate even handle in closure function which will provide persistent variable (called `previous`):

```

modules.load('stats')

-- make a closure, encapsulating counter
function speed_monitor()
    local previous = stats.list()
    -- monitoring function
    return function(evid)
        local now = stats.list()
        local total_increment = now['answer.total'] - previous['answer.total']
        local slow_increment = now['answer.slow'] - previous['answer.slow']
        if slow_increment / total_increment > 0.05 then
            log('WARNING! More than 5 %% of queries was slow!')
        end
        previous = now -- store current value in closure
    end
end

```

(continues on next page)

(continued from previous page)

```
-- monitor every minute
local monitor_id = event.recurrent(1 * minute, speed_monitor())
```

Another type of actionable event is activity on a file descriptor. This allows you to embed other event loops or monitor open files and then fire a callback when an activity is detected. This allows you to build persistent services like HTTP servers or monitoring probes that cooperate well with the daemon internal operations. See `event.socket()`

File watchers are possible with `worker.coroutine()` and `cqueues`, see the `cqueues` documentation for more information.

```
local notify = require('cqueues.notify')
local watcher = notify opendir('/etc')
watcher:add('hosts')

-- Watch changes to /etc/hosts
worker.coroutine(function ()
  for flags, name in watcher:changes() do
    for flag in notify.flags(flags) do
      print(name, notify[flag])
    end
  end
end)
end)
```

### 1.1.3 Configuration reference

This is a reference for variables and functions available to both configuration file and CLI.

- *Environment*
- *Network configuration*
- *TLS server configuration*
- *Trust anchors and DNSSEC*
- *Modules configuration*
- *Cache configuration*
- *Timers and events*
- *Asynchronous function execution*
- *Scripting worker*

#### Environment

##### **env** (table)

Return environment variable.

```
env.USER -- equivalent to $USER in shell
```

##### **hostname** ([fqdn])

**Returns** Machine hostname.

If called with a parameter, it will set kresd's internal hostname. If called without a parameter, it will return kresd's internal hostname, or the system's POSIX hostname (see `gethostname(2)`) if kresd's internal hostname is unset.

This affects ephemeral certificates for kresd serving DNS over TLS.

**moduledir** ([dir])

**Returns** Modules directory.

If called with a parameter, it will change kresd's directory for looking up the dynamic modules. If called without a parameter, it will return kresd's modules directory.

**verbose** (true | false)

**Returns** Toggle verbose logging.

**mode** ('strict' | 'normal' | 'permissive')

**Returns** Change resolver strictness checking level.

By default, resolver runs in *normal* mode. There are possibly many small adjustments hidden behind the mode settings, but the main idea is that in *permissive* mode, the resolver tries to resolve a name with as few lookups as possible, while in *strict* mode it spends much more effort resolving and checking referral path. However, if majority of the traffic is covered by DNSSEC, some of the strict checking actions are counter-productive.

Glue type	Modes when it is accepted	Example glue <sup>1</sup>
mandatory glue	strict, normal, permissive	ns1.example.org
in-bailiwick glue	normal, permissive	ns1.example2.org
any glue records	permissive	ns1.example3.net

**reorder\_RR** ([true | false])

**Parameters**

- **value** (*boolean*) – New value for the option (*optional*)

**Returns** The (new) value of the option

If set, resolver will vary the order of resource records within RR-sets. It is disabled by default.

**user** (name, [group])

**Parameters**

- **name** (*string*) – user name
- **group** (*string*) – group name (*optional*)

**Returns** boolean

Drop privileges and run as given user (and group, if provided).

---

**Tip:** Note that you should bind to required network addresses before changing user. At the same time, you should open the cache **AFTER** you change the user (so it remains accessible). A good practice is to divide configuration in two parts:

---

<sup>1</sup> The examples show glue records acceptable from servers authoritative for *org* zone when delegating to *example.org* zone. Unacceptable or missing glue records trigger resolution of names listed in NS records before following respective delegation.

```
-- privileged
net = { '127.0.0.1', '::1' }
-- unprivileged
cache.size = 100*MB
trust_anchors.file = 'root.key'
```

Example output:

```
> user('baduser')
invalid user name
> user('knot-resolver', 'netgrp')
true
> user('root')
Operation not permitted
```

**resolve** (name, type[, class = *kres.class.IN*, options = {}, finish = nil, init = nil])

#### Parameters

- **name** (*string*) – Query name (e.g. 'com.')
- **type** (*number*) – Query type (e.g. *kres.type.NS*)
- **class** (*number*) – Query class (*optional*) (e.g. *kres.class.IN*)
- **options** (*strings*) – Resolution options (see *kr\_qflags*)
- **finish** (*function*) – Callback to be executed when resolution completes (e.g. *function cb (pkt, req) end*). The callback gets a packet containing the final answer and doesn't have to return anything.
- **init** (*function*) – Callback to be executed with the *kr\_request* before resolution starts.

**Returns** boolean

The function can also be executed with a table of arguments instead. This is useful if you'd like to skip some arguments, for example:

```
resolve {
  name = 'example.com',
  type = kres.type.AAAA,
  init = function (req)
  end,
}
```

Example:

```
-- Send query for root DNSKEY, ignore cache
resolve('.', kres.type.DNSKEY, kres.class.IN, 'NO_CACHE')

-- Query for AAAA record
resolve('example.com', kres.type.AAAA, kres.class.IN, 0,
function (answer, req)
  -- Check answer RCODE
  local pkt = kres.pkt_t(answer)
  if pkt:rcode() == kres.rcode.NOERROR then
    -- Print matching records
    local records = pkt:section(kres.section.ANSWER)
```

(continues on next page)

(continued from previous page)

```

    for i = 1, #records do
        local rr = records[i]
        if rr.type == kres.type.AAAA then
            print ('record:', kres.rr2str(rr))
        end
    end
else
    print ('rcode: ', pkt:rcode())
end
end)

```

**package\_version()****Returns** Current package version.

This returns current package version (the version of the binary) as a string.

```

> package_version()
2.1.1

```

**Network configuration**

For when listening on localhost just doesn't cut it.

**Systemd socket configuration**If you're using our packages with systemd with sockets support (not supported on CentOS 7), network interfaces are configured using systemd drop-in files for `kresd.socket` and `kresd-tls.socket`.To configure `kresd` to listen on public interface, create a drop-in file:

```
$ systemctl edit kresd.socket
```

```

# /etc/systemd/system/kresd.socket.d/override.conf
[Socket]
ListenDatagram=192.0.2.115:53
ListenStream=192.0.2.115:53

```

The default port can also be overridden by using an empty `ListenDatagram=` or `ListenStream=` directive. This can be useful if you want to use the Knot DNS with the [dnssproxy module](#) to have both resolver and authoritative server running on the same machine.

```

# /etc/systemd/system/kresd.socket.d/override.conf
[Socket]
ListenDatagram=
ListenStream=
ListenDatagram=127.0.0.1:53000
ListenStream=127.0.0.1:53000
ListenDatagram=[::1]:53000
ListenStream=[::1]:53000

```

The `kresd-tls.socket` can also be configured to listen for TLS connections.

```
$ systemctl edit kresd-tls.socket
```

```
# /etc/systemd/system/kresd-tls.socket.d/override.conf
[Socket]
ListenStream=192.0.2.115:853
```

### Daemon network configuration

If you don't use systemd with sockets to run kresd, network interfaces are configured in the config file.

**Tip:** Use declarative interface for network.

```
net = { '127.0.0.1', net.eth0, net.eth1.addr[1] }
net.ipv4 = false
```

**Warning:** On machines with multiple IP addresses avoid binding to wildcard 0.0.0.0 or :: (see example below). Knot Resolver could answer from different IP in case the ranges overlap and client will probably refuse such a response.

```
net = { '0.0.0.0' }
```

#### **net.ipv6 = true|false**

**Return** boolean (default: true)

Enable/disable using IPv6 for recursion.

#### **net.ipv4 = true|false**

**Return** boolean (default: true)

Enable/disable using IPv4 for recursion.

#### **net.listen** (addresses, [port = 53, flags = {tls = (port == 853)}])

**Returns** boolean

Listen on addresses; port and flags are optional. The addresses can be specified as a string or device, or a list of addresses (recursively). The command can be given multiple times, but note that it silently skips any addresses that have already been bound.

Examples:

```
net.listen(':::1')
net.listen(net.lo, 5353)
net.listen({net.eth0, '127.0.0.1'}, 53853, {tls = true})
```

#### **net.close** (address, [port = 53])

**Returns** boolean

Close opened address/port pair, noop if not listening.

#### **net.list** ()

**Returns** Table of bound interfaces.

Example output:

```
[127.0.0.1] => {
  [port] => 53
  [tcp] => true
  [udp] => true
}
```

**net.interfaces()**

**Returns** Table of available interfaces and their addresses.

Example output:

```
[lo0] => {
  [addr] => {
    [1] => ::1
    [2] => 127.0.0.1
  }
  [mac] => 00:00:00:00:00:00
}
[eth0] => {
  [addr] => {
    [1] => 192.168.0.1
  }
  [mac] => de:ad:be:ef:aa:bb
}
```

---

**Tip:** You can use `net.<iface>` as a shortcut for specific interface, e.g. `net.eth0`

---

**net.bufsize** ([udp\_bufsize])

Get/set maximum EDNS payload available. Default is 4096. You cannot set less than 512 (512 is DNS packet size without EDNS, 1220 is minimum size for DNSSEC) or more than 65535 octets.

Example output:

```
> net.bufsize 4096
> net.bufsize()
4096
```

**net.tcp\_pipeline** ([len])

Get/set per-client TCP pipeline limit, i.e. the number of outstanding queries that a single client connection can make in parallel. Default is 100.

```
> net.tcp_pipeline()
100
> net.tcp_pipeline(50)
50
```

**Warning:** Please note that too large limit may have negative impact on performance and can lead to increased number of SERVFAIL answers.

**net.outgoing\_v4** ([string address])

Get/set the IPv4 address used to perform queries. There is also `net.outgoing_v6` for IPv6. The default is `nil`, which lets the OS choose any address.



## TLS server configuration

**net.tls** ([cert\_path], [key\_path])

Get/set path to a server TLS certificate and private key for DNS/TLS.

Example output:

```
> net.tls("/etc/knot-resolver/server-cert.pem", "/etc/knot-resolver/server-key.pem")
↔
> net.tls()
("/etc/knot-resolver/server-cert.pem", "/etc/knot-resolver/server-key.pem")
> net.listen("::", 853)
> net.listen("::", 443, {tls = true})
```

**net.tls\_padding** ([true | false])

Get/set EDNS(0) padding of answers to queries that arrive over TLS transport. If set to *true* (the default), it will use a sensible default padding scheme, as implemented by libknot if available at compile time. If set to a numeric value  $\geq 2$  it will pad the answers to nearest *padding* boundary, e.g. if set to *64*, the answer will have size of a multiple of 64 (64, 128, 192, ...). If set to *false* (or a number  $< 2$ ), it will disable padding entirely.

**net.tls\_sticket\_secret** ([string with pre-shared *secret*])

Set secret for TLS session resumption via tickets, by [RFC 5077](#).

The server-side key is rotated roughly once per hour. By default or if called without *secret*, the key is random. That is good for long-term forward secrecy, but multiple *kresd* instances won't be able to resume each other's sessions.

If you provide the same *secret* to multiple instances, they will be able to resume each other's sessions *without* any further communication between them. This synchronization works only among instances having the same endianness and *time\_t* structure and size (*sizeof(time\_t)*).

**For good security** the *secret* must have enough entropy to be hard to guess, and it should still be occasionally rotated manually and securely forgotten, to reduce the scope of privacy leak in case the *secret* leaks eventually.

**Warning:** Setting the secret is probably too risky with TLS  $\leq 1.2$ . At this moment no GnuTLS stable release even supports TLS 1.3. Therefore setting the secrets should be considered experimental for now and might not be available on your system.

**net.tls\_sticket\_secret\_file** ([string with path to a file containing pre-shared *secret*])

The same as *net.tls\_sticket\_secret()*, except the *secret* is read from a (binary) file.

## Trust anchors and DNSSEC

**trust\_anchors.config** (keyfile, readonly)

Alias for *add\_file*. It is also equivalent to CLI parameter *-k <keyfile>* and *trust\_anchors.file = keyfile*.

**trust\_anchors.add\_file** (keyfile, readonly)

### Parameters

- **keyfile** (*string*) – path to the file.
- **readonly** – if true, do not attempt to update the file.

The format is standard zone file, though additional information may be persisted in comments. Either DS or DNSKEY records can be used for TAs. If the file does not exist, bootstrapping of *root* TA will be attempted.

Each file can only contain records for a single domain. The TAs will be updated according to **RFC 5011** and persisted in the file (if allowed).

Example output:

```
> trust_anchors.add_file('root.key')
[ ta ] new state of trust anchors for a domain:
.           165488 DS           19036 8 2
↪49AAC11D7B6F6446702E54A1607371607A1A41855200FD2CE1CDDE32F24E8FB5
nil
[ ta ] key: 19036 state: Valid
```

**trust\_anchors.hold\_down\_time = 30 \* day**

**Return** int (default: 30 \* day)

Modify RFC5011 hold-down timer to given value. Example: 30 \* sec

**trust\_anchors.refresh\_time = nil**

**Return** int (default: nil)

Modify RFC5011 refresh timer to given value (not set by default), this will force trust anchors to be updated every N seconds periodically instead of relying on RFC5011 logic and TTLs. Example: 10 \* sec

**trust\_anchors.keep\_removed = 0**

**Return** int (default: 0)

How many Removed keys should be held in history (and key file) before being purged. Note: all Removed keys will be purged from key file after restarting the process.

**trust\_anchors.set\_insecure (nta\_set)**

**Parameters**

- **nta\_list** (*table*) – List of domain names (text format) representing NTAs.

When you use a domain name as an NTA, DNSSEC validation will be turned off at/below these names. Each function call replaces the previous NTA set. You can find the current active set in `trust_anchors.insecure` variable.

---

**Tip:** Use the `trust_anchors.negative = {}` alias for easier configuration.

---

Example output:

```
> trust_anchors.negative = { 'bad.boy', 'example.com' }
> trust_anchors.insecure
[1] => bad.boy
[2] => example.com
```

**Warning:** If you set NTA on a name that is not a zone cut, it may not always affect names not separated from the NTA by a zone cut.

**trust\_anchors.add (rr\_string)**

**Parameters**

- **rr\_string** (*string*) – DS/DNSKEY records in presentation format (e.g. `. 3600 IN DS 19036 8 2 49AAC11...`)

Inserts DS/DNSKEY record(s) into current keyset. These will not be managed or updated, use it only for testing or if you have a specific use case for not using a keyfile.

Example output:

```
> trust_anchors.add('. 3600 IN DS 19036 8 2 49AAC11...')
```

## Modules configuration

The daemon provides an interface for dynamic loading of *daemon modules*.

**Tip:** Use declarative interface for module loading.

```
modules = {
    hints = {file = '/etc/hosts'}
}
```

Equals to:

```
modules.load('hints')
hints.config({file = '/etc/hosts'})
```

**modules.list** ()

**Returns** List of loaded modules.

**modules.load** (name)

**Parameters**

- **name** (*string*) – Module name, e.g. “hints”

**Returns** boolean

Load a module by name.

**modules.unload** (name)

**Parameters**

- **name** (*string*) – Module name

**Returns** boolean

Unload a module by name.

## Cache configuration

The default cache in Knot Resolver is persistent with LMDB backend, this means that the daemon doesn't lose the cached data on restart or crash to avoid cold-starts. The cache may be reused between cache daemons or manipulated from other processes, making for example synchronized load-balanced recursors possible.

**cache.open** (max\_size[, config\_uri])

**Parameters**

- **max\_size** (*number*) – Maximum cache size in bytes.

**Returns** `true` if cache was opened

Open cache with a size limit. The cache will be reopened if already open. Note that the `max_size` cannot be lowered, only increased due to how cache is implemented.

---

**Tip:** Use `kB`, `MB`, `GB` constants as a multiplier, e.g. `100*MB`.

---

As of now, the built-in backend with URI `lmdb://` allows you to change the cache directory.

Example:

```
cache.open(100 * MB, 'lmdb:///var/cache/knot-resolver')
```

### `cache.size`

Set the cache maximum size in bytes. Note that this is only a hint to the backend, which may or may not respect it. See `cache.open()`.

```
cache.size = 100 * MB -- equivalent to `cache.open(100 * MB)`
```

### `cache.current_size`

Get the maximum size in bytes.

```
print(cache.current_size)
```

### `cache.storage`

Set the cache storage backend configuration, see `cache.backends()` for more information. If the new storage configuration is invalid, it is not set.

```
cache.storage = 'lmdb://.'
```

### `cache.current_storage`

Get the storage backend configuration.

```
print(cache.storage)
```

### `cache.backends()`

**Returns** map of backends

The cache supports runtime-changeable backends, using the optional [RFC 3986](#) URI, where the scheme represents backend protocol and the rest of the URI backend-specific configuration. By default, it is a `lmdb` backend in working directory, i.e. `lmdb://`.

Example output:

```
[lmdb://] => true
```

### `cache.count()`

**Returns** Number of entries in the cache. Meaning of the number is an implementation detail and is subject of change.

### `cache.close()`

**Returns** `true` if cache was closed

Close the cache.

---

**Note:** This may or may not clear the cache, depending on the cache backend.

---

`cache.stats()`

**Warning:** Cache statistics are being reworked. Do not rely on current behavior.

Return table of statistics, note that this tracks all operations over cache, not just which queries were answered from cache or not.

Example:

```
print('Insertions:', cache.stats().insert)
```

`cache.max_ttl([ttl])`

**Parameters**

- `ttl` (*number*) – maximum cache TTL in seconds (default: 6 days)

**Returns** current maximum TTL

Get or set maximum cache TTL.

---

**Note:** The *ttl* value must be in range (*min\_ttl*, 4294967295).

---

**Warning:** This settings applies only to currently open cache, it will not persist if the cache is closed or reopened.

```
-- Get maximum TTL
cache.max_ttl()
518400
-- Set maximum TTL
cache.max_ttl(172800)
172800
```

`cache.min_ttl([ttl])`

**Parameters**

- `ttl` (*number*) – minimum cache TTL in seconds (default: 5 seconds)

**Returns** current maximum TTL

Get or set minimum cache TTL. Any entry inserted into cache with TTL lower than minimal will be overridden to minimum TTL. Forcing TTL higher than specified violates DNS standards, use with care.

---

**Note:** The *ttl* value must be in range  $<0, max\_ttl$ .

---

**Warning:** This settings applies only to currently open cache, it will not persist if the cache is closed or reopened.

```
-- Get minimum TTL
cache.min_ttl()
0
-- Set minimum TTL
cache.min_ttl(5)
5
```

`cache.ns_tout` ([timeout])

#### Parameters

- **timeout** (*number*) – NS retry interval in milliseconds (default: `KR_NS_TIMEOUT_RETRY_INTERVAL`)

**Returns** current timeout

Get or set time interval for which a nameserver address will be ignored after determining that it doesn't return (useful) answers. The intention is to avoid waiting if there's little hope; instead, kresd can immediately SERVFAIL or immediately use stale records (with `serve_stale` module).

**Warning:** This settings applies only to the current kresd process.

`cache.get` ([domain])

This function is not implemented at this moment. We plan to re-introduce it soon, probably with a slightly different API.

`cache.clear` ([name], [exact\_name], [rr\_type], [chunk\_size], [callback], [prev\_state])

Purge cache records matching specified criteria. There are two specifics:

- To reliably remove **negative** cache entries you need to clear subtree with the whole zone. E.g. to clear negative cache entries for (formerly non-existing) record `www.example.com`. A you need to flush whole subtree starting at zone apex, e.g. `example.com`.<sup>2</sup>
- This operation is asynchronous and might not be yet finished when call to `cache.clear()` function returns. Return value indicates if clearing continues asynchronously or not.

#### Parameters

- **name** (*string*) – subtree to purge; if the name isn't provided, whole cache is purged (and any other parameters are disregarded).
- **exact\_name** (*bool*) – if set to `true`, only records with *the same* name are removed; default: `false`.
- **rr\_type** (*kres.type*) – you may additionally specify the type to remove, but that is only supported with `exact_name == true`; default: `nil`.
- **chunk\_size** (*integer*) – the number of records to remove in one round; default: 100. The purpose is not to block the resolver for long. The default `callback` repeats the command after one millisecond until all matching data are cleared.

<sup>2</sup> This is a consequence of DNSSEC negative cache which relies on proofs of non-existence on various owner nodes. It is impossible to efficiently flush part of DNS zones signed with NSEC3.

- **callback** (*function*) – a custom code to handle result of the underlying C call. Its parameters are copies of those passed to `cache.clear()` with one additional parameter `rettable` containing table with return value from current call. `count` field contains a return code from `kr_cache_remove_subtree()`.
- **prev\_state** (*table*) – return value from previous run (can be used by callback)

**Return type** table

### Returns

`count` key is always present. Other keys are optional and their presence indicate special conditions.

- **count** (*integer*) - number of items removed from cache by this call (can be 0 if no entry matched criteria)
- **not\_apex** - cleared subtree is not cached as zone apex; proofs of non-existence were probably not removed
- **subtree** (*string*) - hint where zone apex lies (this is estimation from cache content and might not be accurate)
- **chunk\_limit** - more than `chunk_size` items needs to be cleared, clearing will continue asynchronously

Examples:

```
-- Clear whole cache
> cache.clear()
[count] => 76

-- Clear records at and below 'com.'
> cache.clear('com.')
[chunk_limit] => chunk size limit reached; the default callback will continue_
↳asynchronously
[not_apex] => to clear proofs of non-existence call cache.clear('com.')
[count] => 100
[round] => 1
[subtree] => com.
> worker.sleep(0.1)
[cache] asynchronous cache.clear('com', false) finished

-- Clear only 'www.example.com.'
> cache.clear('www.example.com.', true)
[round] => 1
[count] => 1
[not_apex] => to clear proofs of non-existence call cache.clear('example.com.')
[subtree] => example.com.
```

## Timers and events

The timer represents exactly the thing described in the examples - it allows you to execute closures after specified time, or event recurrent events. Time is always described in milliseconds, but there are convenient variables that you can use - `sec`, `minute`, `hour`. For example, `5 * hour` represents five hours, or `5*60*60*100` milliseconds.

**event.after** (time, function)

**Returns** event id

Execute function after the specified time has passed. The first parameter of the callback is the event itself.

Example:

```
event.after(1 * minute, function() print('Hi!') end)
```

**event.recurrent** (interval, function)

**Returns** event id

Similar to `event.after()`, periodically execute function after interval passes.

Example:

```
msg_count = 0
event.recurrent(5 * sec, function(e)
  msg_count = msg_count + 1
  print('Hi #'..msg_count)
end)
```

**event.reschedule** (event\_id, timeout)

Reschedule a running event, it has no effect on canceled events. New events may reuse the event\_id, so the behaviour is undefined if the function is called after another event is started.

Example:

```
local interval = 1 * minute
event.after(1 * minute, function (ev)
  print('Good morning!')
  -- Halven the interval for each iteration
  interval = interval / 2
  event.reschedule(ev, interval)
end)
```

**event.cancel** (event\_id)

Cancel running event, it has no effect on already canceled events. New events may reuse the event\_id, so the behaviour is undefined if the function is called after another event is started.

Example:

```
e = event.after(1 * minute, function() print('Hi!') end)
event.cancel(e)
```

Watch for file descriptor activity. This allows embedding other event loops or simply firing events when a pipe endpoint becomes active. In another words, asynchronous notifications for daemon.

**event.socket** (fd, cb)

**Parameters**

- **fd** (*number*) – file descriptor to watch
- **cb** – closure or callback to execute when fd becomes active

**Returns** event id

Execute function when there is activity on the file descriptor and calls a closure with event id as the first parameter, status as second and number of events as third.

Example:



```
e = event.socket(0, function(e, status, nevents)
    print('activity detected')
end)
e.cancel(e)
```

## Asynchronous function execution

The *event* package provides a very basic mean for non-blocking execution - it allows running code when activity on a file descriptor is detected, and when a certain amount of time passes. It doesn't however provide an easy to use abstraction for non-blocking I/O. This is instead exposed through the *worker* package (if *cqueues* Lua package is installed in the system).

### **worker.coroutine** (function)

Start a new coroutine with given function (closure). The function can do I/O or run timers without blocking the main thread. See *cqueues* for documentation of possible operations and synchronization primitives. The main limitation is that you can't wait for a finish of a coroutine from processing layers, because it's not currently possible to suspend and resume execution of processing layers.

Example:

```
worker.coroutine(function ()
    for i = 0, 10 do
        print('executing', i)
        worker.sleep(1)
    end
end)
```

### **worker.sleep** (seconds)

Pause execution of current function (asynchronously if running inside a worker coroutine).

When daemon is running in forked mode, each process acts independently. This is good because it reduces software complexity and allows for runtime scaling, but not ideal because of additional operational burden. For example, when you want to add a new policy, you'd need to add it to either put it in the configuration, or execute command on each process independently. The daemon simplifies this by promoting process group leader which is able to execute commands synchronously over forks.

Example:

```
worker.sleep(1)
```

### **map** (expr)

Run expression synchronously over all forks, results are returned as a table ordered as forks. Expression can be any valid expression in Lua.

Example:

```
-- Current instance only
hostname()
localhost
-- Mapped to forks
map 'hostname()'
[1] => localhost
[2] => localhost
-- Get worker ID from each fork
map 'worker.id'
[1] => 0
```

(continues on next page)

(continued from previous page)

```
[2] => 1
-- Get cache stats from each fork
map 'cache.stats()'
[1] => {
  [hit] => 0
  [delete] => 0
  [miss] => 0
  [insert] => 0
}
[2] => {
  [hit] => 0
  [delete] => 0
  [miss] => 0
  [insert] => 0
}
```

## Scripting worker

Worker is a service over event loop that tracks and schedules outstanding queries, you can see the statistics or schedule new queries. It also contains information about specified worker count and process rank.

### **worker.count**

Return current total worker count (e.g. *1* for single-process)

### **worker.id**

Return current worker ID (starting from *0* up to *worker.count - 1*)

### **worker.pid**

Current worker process PID (number).

### **worker.stats()**

Return table of statistics.

- `udp` - number of outbound queries over UDP
- `tcp` - number of outbound queries over TCP
- `ipv6` - number of outbound queries over IPv6
- `ipv4` - number of outbound queries over IPv4
- `timeout` - number of timeouted outbound queries
- `concurrent` - number of concurrent queries at the moment
- `queries` - number of inbound queries
- `dropped` - number of dropped inbound queries

Example:

```
print(worker.stats().concurrent)
```

## 1.2 Enabling DNSSEC

The resolver supports DNSSEC including [RFC 5011](#) automated DNSSEC TA updates and [RFC 7646](#) negative trust anchors. To enable it, you need to provide trusted root keys. Bootstrapping of the keys is automated, and kresd fetches

root trust anchors set over a [secure channel](#) from IANA. From there, it can perform [RFC 5011](#) automatic updates for you.

**Note:** Automatic bootstrap requires [luasocket](#) and [luasec](#) installed.

```
$ kresd -k root-new.keys # File for root keys
[ ta ] keyfile 'root-new.keys': doesn't exist, bootstrapping
[ ta ] Root trust anchors bootstrapped over https with pinned certificate.
      You SHOULD verify them manually against original source:
      https://www.iana.org/dnssec/files
[ ta ] Current root trust anchors are:
. 0 IN DS 19036 8 2 49AAC11D7B6F6446702E54A1607371607A1A41855200FD2CE1CDDE32F24E8FB5
. 0 IN DS 20326 8 2 E06D44B80B8F1D39A95C0B0D7C65D08458E880409BBC683457104237C7F8EC8D
[ ta ] next refresh for . in 24 hours
```

Alternatively, you can set it in configuration file with `trust_anchors.file = 'root.keys'`. If the file doesn't exist, it will be automatically populated with root keys validated using root anchors retrieved over HTTPS.

This is equivalent to using `unbound-anchor`:

```
$ unbound-anchor -a "root.keys" || echo "warning: check the key at this point"
$ echo "auto-trust-anchor-file: \"root.keys\"" >> unbound.conf
$ unbound -c unbound.conf
```

**Warning:** Bootstrapping of the root trust anchors is automatic, you are however **encouraged to check** the key over **secure channel**, as specified in [DNSSEC Trust Anchor Publication for the Root Zone](#). This is a critical step where the whole infrastructure may be compromised, you will be warned in the server log.

Configuration is described in *Trust anchors and DNSSEC*.

## 1.2.1 Manually providing root anchors

The root anchors bootstrap may fail for various reasons, in this case you need to provide IANA or alternative root anchors. The format of the keyfile is the same as for Unbound or BIND and contains DS/DNSKEY records.

1. Check the current TA published on [IANA website](#)
2. Fetch current keys (DNSKEY), verify digests
3. Deploy them

```
$ kdig DNSKEY . @k.root-servers.net +noall +answer | grep "DNSKEY[[:space:]]257" >
↪root.keys
$ ldns-key2ds -n root.keys # Only print to stdout
... verify that digest matches TA published by IANA ...
$ kresd -k root.keys
```

You've just enabled DNSSEC!

**Note:** Bootstrapping and automatic update need write access to keyfile directory. If you want to manage root anchors manually you should use `trust_anchors.add_file('root.keys', true)`.

## 1.3 CLI interface

The daemon features a CLI interface, type `help()` to see the list of available commands.

```
$ kresd /var/cache/knot-resolver
[system] started in interactive mode, type 'help()'
> cache.count()
53
```

### 1.3.1 Verbose output

If the verbose logging is compiled in, i.e. not turned off by `-DNOVERBOSELOG`, you can turn on verbose tracing of server operation with the `-v` option. You can also toggle it on runtime with `verbose(true|false)` command.

```
$ kresd -v
```

To run the daemon by hand, such as under `nohup`, use `-f 1` to start a single fork. For example:

```
$ nohup ./daemon/kresd -a 127.0.0.1 -f 1 -v &
```

## 1.4 Control sockets

Unless ran manually, `knot-resolver` is typically started in non-interactive mode. The mode gets triggered by using the `-f` command-line parameter or by passing sockets from `systemd`. You can attach to the the consoles for each process; by default they are in `rundir/tty/$PID`.

---

**Note:** When running `kresd` with `systemd`, you can find the location of the socket(s) using `systemctl status kresd-control@*.socket`. Typically, these are in `/run/knot-resolver/control@*`.

---

```
$ nc -U rundir/tty/3008 # or socat - UNIX-CONNECT:rundir/tty/3008
> cache.count()
53
```

The *direct output* of the CLI command is captured and sent over the socket, while also printed to the daemon standard outputs (for accountability). This gives you an immediate response on the outcome of your command. Error or debug logs aren't captured, but you can find them in the daemon standard outputs.

This is also a way to enumerate and test running instances, the list of files in `tty` corresponds to the list of running processes, and you can test the process for liveness by connecting to the UNIX socket.

## 1.5 Utilizing multiple CPUs

The server can run in multiple independent processes, all sharing the same socket and cache. These processes can be started or stopped during runtime based on the load.

### Using `systemd`

To run multiple daemons using `systemd`, use a different numeric identifier for the instance, for example:



```
$ kresd-query.lua www.sub.nic.cz 'assert(kres.dname2str(req:resolved()).zone_cut.name) == "nic.cz.")' && echo "yes"
yes
$ kresd-query.lua -C 'trust_anchors.config("root.keys")' nic.cz
→ 'assert(req:resolved().flags.DNSSEC_WANT)'
$ echo $?
0
```

---

### Knot Resolver modules

---

- *Static hints*
- *Statistics collector*
- *Query policies*
- *Views and ACLs*
- *Prefetching records*
- *HTTP/2 services*
- *DNS Application Firewall*
- *Rebinding protection*
- *Graphite module*
- *Etc module*
- *DNS64*
- *Renumber*
- *DNSSEC validation failure logging*
- *Workarounds*
- *Dnstap*
- *Signaling Trust Anchor Knowledge in DNSSEC*
- *Sentinel for Detecting Trusted Root Keys*
- *Priming module*
- *System time skew detector*
- *Detect discontinuous jumps in the system time*

- *Root on loopback (RFC 7706)*
- *Cache prefilling*
- *Serve stale*

## 2.1 Static hints

This is a module providing static hints for forward records (A/AAAA) and reverse records (PTR). The records can be loaded from `/etc/hosts`-like files and/or added directly.

You can also use the module to change the root hints; they are used as a safety belt or if the root NS drops out of cache.

### 2.1.1 Examples

```
-- Load hints after iterator (so hints take precedence before caches)
modules = { 'hints > iterate' }
-- Add a custom hosts file
hints.add_hosts('hosts.custom')
-- Override the root hints
hints.root({
  ['j.root-servers.net.'] = { '2001:503:c27::2:30', '192.58.128.30' }
})
-- Add a custom hint
hints['foo.bar'] = '127.0.0.1'
```

---

**Note:** The *policy* module applies before hints, meaning e.g. that hints for special names ([RFC 6761#section-6](#)) like `localhost` or `test` will get shadowed by policy rules by default. That can be worked around e.g. by explicit `policy.PASS` action.

---

### 2.1.2 Properties

`hints.config` ([path])

**Parameters**

- **path** (*string*) – path to hosts-like file, default: no file

**Returns** { result: bool }

Clear any configured hints, and optionally load a hosts-like file as in `hints.add_hosts(path)`. (Root hints are not touched.)

`hints.add_hosts` ([path])

**Parameters**

- **path** (*string*) – path to hosts-like file, default: `/etc/hosts`

Add hints from a host-like file.

`hints.get` (*hostname*)

**Parameters**



- **hostname** (*string*) – i.e. "localhost"

**Returns** { result: [address1, address2, ...] }

Return list of address record matching given name. If no hostname is specified, all hints are returned in the table format used by `hints.root()`.

**hints.set** (*pair*)

**Parameters**

- **pair** (*string*) – hostname address i.e. "localhost 127.0.0.1"

**Returns** { result: bool }

Add a hostname–address pair hint.

---

**Note:** If multiple addresses have been added for a name (in separate `hints.set()` commands), all are returned in a forward query. If multiple names have been added to an address, the last one defined is returned in a corresponding PTR query.

---

**hints.del** (*pair*)

**Parameters**

- **pair** (*string*) – hostname address i.e. "localhost 127.0.0.1", or just hostname

**Returns** { result: bool }

Remove a hostname - address pair hint. If address is omitted, all addresses for the given name are deleted.

**hints.root\_file** (*path*)

Replace current root hints from a zonefile. If the path is omitted, the compiled-in path is used, i.e. the root hints are reset to the default.

**hints.root** (*root\_hints*)

**Parameters**

- **root\_hints** (*table*) – new set of root hints i.e. { ['name'] = 'addr', ... }

**Returns** { ['a.root-servers.net.'] = { '1.2.3.4', '5.6.7.8', ... }, ... }

Replace current root hints and return the current table of root hints.

---

**Tip:** If no parameters are passed, it only returns current root hints set without changing anything.

---

Example:

```
> hints.root({
  ['l.root-servers.net.'] = '199.7.83.42',
  ['m.root-servers.net.'] = '202.12.27.33'
})
[l.root-servers.net.] => {
  [1] => 199.7.83.42
}
[m.root-servers.net.] => {
  [1] => 202.12.27.33
}
```

---

**Tip:** A good rule of thumb is to select only a few fastest root hints. The server learns RTT and NS quality over time, and thus tries all servers available. You can help it by preselecting the candidates.

---

**hints.use\_nodata** (toggle)

**Parameters**

- **toggle** (*bool*) – true if enabling NODATA synthesis, false if disabling

**Returns** { result: bool }

If set to true (the default), NODATA will be synthesised for matching hint name, but mismatching type (e.g. AAAA query when only A hint exists).

## 2.2 Statistics collector

This module gathers various counters from the query resolution and server internals, and offers them as a key-value storage. Any module may update the metrics or simply hook in new ones.

```
-- Enumerate metrics
> stats.list()
[answer.cached] => 486178
[iterator.tcp] => 490
[answer.noerror] => 507367
[answer.total] => 618631
[iterator.udp] => 102408
[query.concurrent] => 149

-- Query metrics by prefix
> stats.list('iter')
[iterator.udp] => 105104
[iterator.tcp] => 490

-- Set custom metrics from modules
> stats['filter.match'] = 5
> stats['filter.match']
5

-- Fetch most common queries
> stats.frequent()
[1] => {
  [type] => 2
  [count] => 4
  [name] => cz.
}

-- Fetch most common queries (sorted by frequency)
> table.sort(stats.frequent(), function (a, b) return a.count > b.count end)

-- Show recently contacted authoritative servers
> stats.upstreams()
[2a01:618:404::1] => {
  [1] => 26 -- RTT
}
[128.241.220.33] => {
```

(continues on next page)

(continued from previous page)

```
[1] => 31 - RTT
}
```

## 2.2.1 Properties

**stats.get** (key)

**Parameters**

- **key** (*string*) – i.e. "answer.total"

**Returns** number

Return nominal value of given metric.

**stats.set** (key, val)

**Parameters**

- **key** (*string*) – i.e. "answer.total"
- **val** (*number*) – i.e. 5

Set nominal value of given metric.

**stats.list** ([prefix])

**Parameters**

- **prefix** (*string*) – optional metric prefix, i.e. "answer" shows only metrics beginning with "answer"

Outputs collected metrics as a JSON dictionary.

**stats.upstreams** ()

Outputs a list of recent upstreams and their RTT. It is sorted by time and stored in a ring buffer of a fixed size. This means it's not aggregated and readable by multiple consumers, but also that you may lose entries if you don't read quickly enough. The default ring size is 512 entries, and may be overridden on compile time by `-DUPSTREAMS_COUNT=X`.

**stats.frequent** ()

Outputs list of most frequent iterative queries as a JSON array. The queries are sampled probabilistically, and include subrequests. The list maximum size is 5000 entries, make diffs if you want to track it over time.

**stats.clear\_frequent** ()

Clear the list of most frequent iterative queries.

## 2.2.2 Built-in statistics

- `answer.total` - total number of answered queries
- `answer.cached` - number of queries answered from cache
- `answer.noerror` - number of **NOERROR** answers
- `answer.nodata` - number of **NOERROR**, but empty answers
- `answer.nxdomain` - number of **NXDOMAIN** answers
- `answer.servfail` - number of **SERVFAIL** answers

- `answer.1ms` - number of answers completed in 1ms
- `answer.10ms` - number of answers completed in 10ms
- `answer.50ms` - number of answers completed in 50ms
- `answer.100ms` - number of answers completed in 100ms
- `answer.250ms` - number of answers completed in 250ms
- `answer.500ms` - number of answers completed in 500ms
- `answer.1000ms` - number of answers completed in 1000ms
- `answer.1500ms` - number of answers completed in 1500ms
- `answer.slow` - number of answers that took more than 1500ms
- `query.edns` - number of queries with EDNS
- `query.dnssec` - number of queries with DNSSEC DO=1

## 2.3 Query policies

This module can block, rewrite, or alter inbound queries based on user-defined policies.

Each policy *rule* has two parts: a *filter* and an *action*. A *filter* selects which queries will be affected by the policy, and *action* which modifies queries matching the associated filter. Typically a rule is defined as follows: `filter(action(action parameters), filter parameters)`. For example, a filter can be `suffix` which matches queries whose suffix part is in specified set, and one of possible actions is `DENY`, which denies resolution. These are combined together into `policy.suffix(policy.DENY, {todname('badguy.example.')})`. The rule is effective when it is added into rule table using `policy.add()`, please see *Policy examples*.

By default, if no rule applies to a query, built-in rules for *special-use* and *locally-served* domain names are applied. These built-in rules can be overridden using action `PASS`, see *Policy examples* below.

### 2.3.1 Filters

A *filter* selects which queries will be affected by specified *action*. There are several policy filters available in the `policy` table:

- `all(action)` - always applies the action
- `pattern(action, pattern)` - applies the action if QNAME matches a [regular expression](#)
- `suffix(action, table)` - applies the action if QNAME suffix matches one of suffixes in the table (useful for “is domain in zone” rules), uses [Aho-Corasick](#) string matching algorithm from [CloudFlare](#) (BSD 3-clause)
- `policy.suffix_common`
- `rpz(default_action, path)` - implements a subset of [RPZ](#) in zonefile format. See below for details: [policy.rpz](#).
- custom filter function

### 2.3.2 Actions

An *action* is function which modifies DNS query. There are several actions available in the `policy` table:

- `PASS` - let the query pass through; it’s useful to make exceptions before wider rules

- DENY - reply NXDOMAIN authoritatively
- DENY\_MSG(msg) - reply NXDOMAIN authoritatively and add explanatory message to additional section
- DROP - terminate query resolution and return SERVFAIL to the requestor
- REFUSE - terminate query resolution and return REFUSED to the requestor
- TC - set TC=1 if the request came through UDP, forcing client to retry with TCP
- FORWARD(ip) - resolve a query via forwarding to an IP while validating and caching locally;
- TLS\_FORWARD({{ip, authentication}}) - resolve a query via TLS connection forwarding to an IP while validating and caching locally; the parameter can be a single IP (string) or a lua list of up to four IPs.
- STUB(ip) - similar to FORWARD(ip) but *without* attempting DNSSEC validation. Each request may be either answered from cache or simply sent to one of the IPs with proxying back the answer.
- MIRROR(ip) - mirror query to given IP and continue solving it (useful for partial snooping); it's a chain action
- REROUTE({{subnet, target}, ...}) - reroute addresses in response matching given subnet to given target, e.g. {'192.0.2.0/24', '127.0.0.0'} will rewrite '192.0.2.55' to '127.0.0.55', see [renumber module](#) for more information.
- QTRACE - pretty-print DNS response packets into the log for the query and its sub-queries. It's useful for debugging weird DNS servers. It's a chain action.
- FLAGS(set, clear) - set and/or clear some flags for the query. There can be multiple flags to set/clear. You can just pass a single flag name (string) or a set of names. It's a chain action.

Most actions stop the policy matching on the query, but “chain actions” allow to keep trying to match other rules, until a non-chain action is triggered.

Also, it is possible to write your own action (i.e. Lua function). It is possible to implement complex heuristics, e.g. to deflect [Slow drip DNS attacks](#) or gray-list resolution of misbehaving zones.

**Warning:** The policy module currently only looks at whole DNS requests. The rules won't be re-applied e.g. when following CNAMEs.

**Note:** The module (and `kres`) expects domain names in wire format, not textual representation. So each label in name is prefixed with its length, e.g. “example.com” equals to “\7example\3com”. You can use convenience function `todname('example.com')` for automatic conversion.

### 2.3.3 Forwarding over TLS protocol (DNS-over-TLS)

Policy `TLS_FORWARD` allows you to forward queries using [Transport Layer Security](#) protocol, which hides the content of your queries from an attacker observing the network traffic. Further details about this protocol can be found in [RFC 7858](#) and [IETF draft dprive-dtls-and-tls-profiles](#).

Queries affected by `TLS_FORWARD` policy will always be resolved over TLS connection. Knot Resolver does not implement fallback to non-TLS connection, so if TLS connection cannot be established or authenticated according to the configuration, the resolution will fail.

To test this feature you need to either [configure Knot Resolver as DNS-over-TLS server](#), or pick some public DNS-over-TLS server. Please see [DNS Privacy Project](#) homepage for list of public servers.

When multiple servers are specified, the one with the lowest round-trip time is used.

## CA+hostname authentication

Traditional PKI authentication requires server to present certificate with specified hostname, which is issued by one of trusted CAs. Example policy is:

```
policy.TLS_FORWARD({
    {'2001:DB8::d0c', hostname='res.example.com'}})
```

- *hostname* must exactly match hostname in server's certificate, i.e. in most cases it must not contain trailing dot (*res.example.com*).
- System CA certificate store will be used if no *ca\_file* option is specified.
- Optional *ca\_file* option can specify path to CA certificate (or certificate bundle) in PEM format.

## TLS Examples

```
modules = { 'policy' }
-- forward all queries over TLS to the specified server
policy.add(policy.all(policy.TLS_FORWARD({'192.0.2.1', pin_sha256='YQ=='})))
-- for brevity, other TLS examples omit policy.add(policy.all())
-- single server authenticated using its certificate pin_sha256
policy.TLS_FORWARD({'192.0.2.1', pin_sha256='YQ=='}) -- pin_sha256 is base64-
↳encoded
-- single server authenticated using hostname and system-wide CA certificates
policy.TLS_FORWARD({'192.0.2.1', hostname='res.example.com'})
-- single server using non-standard port
policy.TLS_FORWARD({'192.0.2.1@443', pin_sha256='YQ=='}) -- use @ or # to
↳specify port
-- single server with multiple valid pins (e.g. anycast)
policy.TLS_FORWARD({'192.0.2.1', pin_sha256={'YQ==', 'Wg=='}})
-- multiple servers, each with own authenticator
policy.TLS_FORWARD({ -- please note that { here starts list of servers
    {'192.0.2.1', pin_sha256='Wg=='},
    -- server must present certificate issued by specified CA and hostname must
↳match
    {'2001:DB8::d0c', hostname='res.example.com', ca_file='/etc/knot-resolver/
↳tlsca.crt'}
})
```

## 2.3.4 Policy examples

```
-- Whitelist 'www[0-9].badboy.cz'
policy.add(policy.pattern(policy.PASS, '\4www[0-9]\6badboy\2cz'))
-- Block all names below badboy.cz
policy.add(policy.suffix(policy.DENY, {todname('badboy.cz.')}))

-- Custom rule
local ffi = require('ffi')
local function genRR (state, req)
    local answer = req.answer
    local qry = req:current()
    if qry.stype ~= kres.type.A then
        return state
    end
end
```

(continues on next page)

(continued from previous page)

```

ffi.C.kr_pkt_make_auth_header(answer)
answer:rcode(kres.rcode.NOERROR)
answer:begin(kres.section.ANSWER)
answer:put(qry.sname, 900, answer:qclass(), kres.type.A, '\192\168\1\3')
return kres.DONE
end
policy.add(policy.suffix(genRR, { todname('my.example.cz.' )}))

-- Disallow ANY queries
policy.add(function (req, query)
    if query.stype == kres.type.ANY then
        return policy.DROP
    end
end)

-- Enforce local RPZ
policy.add(policy.rpz(policy.DENY, 'blacklist.rpz'))
-- Forward all queries below 'company.se' to given resolver
policy.add(policy.suffix(policy.FORWARD('192.168.1.1'), {todname('company.se')}))
-- Forward all queries matching pattern
policy.add(policy.pattern(policy.FORWARD('2001:DB8::1'), '\4bad[0-9]\2cz'))
-- Forward all queries (to public resolvers https://www.nic.cz/odvr)
policy.add(policy.all(policy.FORWARD({'2001:678:1::206', '193.29.206.206'})))
-- Print all responses with matching suffix
policy.add(policy.suffix(policy.QTRACE, {todname('rhybar.cz.')}))
-- Print all responses
policy.add(policy.all(policy.QTRACE))
-- Mirror all queries and retrieve information
local rule = policy.add(policy.all(policy.MIRROR('127.0.0.2')))
-- Print information about the rule
print(string.format('id: %d, matched queries: %d', rule.id, rule.count))
-- Reroute all addresses found in answer from 192.0.2.0/24 to 127.0.0.x
-- this policy is enforced on answers, therefore 'postrule'
local rule = policy.add(policy.REROUTE({'192.0.2.0/24', '127.0.0.0'}, true))
-- Delete rule that we just created
policy.del(rule.id)

```

### 2.3.5 Additional properties

Most properties (actions, filters) are described above.

**policy.add** (rule, postrule)

#### Parameters

- **rule** – added rule, i.e. `policy.pattern(policy.DENY, '[0-9]+\2cz')`
- **postrule** – boolean, if true the rule will be evaluated on answer instead of query

#### Returns rule description

Add a new policy rule that is executed either on queries or answers, depending on the `postrule` parameter. You can then use the returned rule description to get information and unique identifier for the rule, as well as match count.

**policy.del** (id)

#### Parameters

- **id** – identifier of a given rule

**Returns** boolean

Remove a rule from policy list.

`policy.suffix_common` (action, suffix\_table[, common\_suffix])

**Parameters**

- **action** – action if the pattern matches QNAME
- **suffix\_table** – table of valid suffixes
- **common\_suffix** – common suffix of entries in suffix\_table

Like suffix match, but you can also provide a common suffix of all matches for faster processing (nil otherwise). This function is faster for small suffix tables (in the order of “hundreds”).

`policy.rpz` (action, path)

**Parameters**

- **action** – the default action for match in the zone; typically you want `policy.DENY`
- **path** – path to zone file | database

Enforce **RPZ** rules. This can be used in conjunction with published blacklist feeds. The **RPZ** operation is well described in this [Jan-Piet Mens’s post](#), or the [Pro DNS and BIND](#) book. Here’s compatibility table:

Policy Action	RH Value	Support
action is used	.	yes, if action is DENY
action is used	*.	<i>partial</i> <sup>1</sup>
policy.PASS	rpz-passthru.	yes
policy.DROP	rpz-drop.	yes
policy.TC	rpz-tcp-only.	yes
Modified	anything	no

Policy Trigger	Support
QNAME	yes
CLIENT-IP	<i>partial</i> , may be done with <i>views</i>
IP	no
NSDNAME	no
NS-IP	no

`policy.todnames` ({name, ...})

**Param** names table of domain names in textual format

Returns table of domain names in wire format converted from strings.

```

-- Convert single name
assert(todname('example.com') == '\7example\3com\0')
-- Convert table of names
policy.todnames({'example.com', 'me.cz'})
{ '\7example\3com\0', '\2me\2cz\0' }

```

This module is enabled by default because it implements mandatory **RFC 6761** logic. For debugging purposes you can add `modules.unload('policy')` to your config to unload the module.

<sup>1</sup> The specification for \*. wants a NODATA answer. For now, `policy.DENY` action doing `NXDOMAIN` is typically used instead.



## 2.4 Views and ACLs

The *policy* module implements policies for global query matching, e.g. solves “how to react to certain query”. This module combines it with query source matching, e.g. “who asked the query”. This allows you to create personalized blacklists, filters and ACLs, sort of like ISC BIND views.

There are two identification mechanisms:

- `addr` - identifies the client based on his subnet
- `tsig` - identifies the client based on a TSIG key

You can combine this information with *policy* rules.

```
view:addr('10.0.0.1', policy.suffix(policy.TC, {'\7example\3com'}))
```

This will force given client subnet to TCP for names in `example.com`. You can combine view selectors with `RPZ` to create personalized filters for example.

### 2.4.1 Example configuration

```
-- Load modules
modules = { 'policy', 'view' }
-- Whitelist queries identified by TSIG key
view:tsig('\5mykey', function (req, qry) return policy.PASS end)
-- Block local clients (ACL like)
view:addr('127.0.0.1', function (req, qry) return policy.DENY end)
-- Drop queries with suffix match for remote client
view:addr('10.0.0.0/8', policy.suffix(policy.DROP, {'\3xxx'}))
-- RPZ for subset of clients
view:addr('192.168.1.0/24', policy.rpz(policy.PASS, 'whitelist.rpz'))
-- Forward all queries from given subnet to proxy
view:addr('10.0.0.0/8', policy.all(policy.FORWARD('2001:DB8::1')))
-- Drop everything that hasn't matched
view:addr('0.0.0.0/0', function (req, qry) return policy.DROP end)
```

### 2.4.2 Properties

**view:addr** (subnet, rule)

#### Parameters

- **subnet** – client subnet, i.e. `10.0.0.1`
- **rule** – added rule, i.e. `policy.pattern(policy.DENY, '[0-9]+\2cz')`

Apply rule to clients in given subnet.

**view:tsig** (key, rule)

#### Parameters

- **key** – client TSIG key domain name, i.e. `\5mykey`
- **rule** – added rule, i.e. `policy.pattern(policy.DENY, '[0-9]+\2cz')`

Apply rule to clients with given TSIG key.

**Warning:** This just selects rule based on the key name, it doesn't verify the key or signature yet.

## 2.5 Prefetching records

The module refreshes records that are about to expire when they're used (having less than 1% of original TTL). This improves latency for frequently used records, as they are fetched in advance.

It is also able to learn usage patterns and repetitive queries that the server makes. For example, if it makes a query every day at 18:00, the resolver expects that it is needed by that time and prefetches it ahead of time. This is helpful to minimize the perceived latency and keeps the cache hot.

---

**Tip:** The tracking window and period length determine memory requirements. If you have a server with relatively fast query turnover, keep the period low (hour for start) and shorter tracking window (5 minutes). For personal slower resolver, keep the tracking window longer (i.e. 30 minutes) and period longer (a day), as the habitual queries occur daily. Experiment to get the best results.

---

### 2.5.1 Example configuration

```
modules = {
  predict = {
    window = 15, -- 15 minutes sampling window
    period = 6*(60/15) -- track last 6 hours
  }
}
```

Defaults are 15 minutes window, 6 hours period.

---

**Tip:** Use period 0 to turn off prediction and just do prefetching of expiring records. That works even without the *stats* module.

---

---

**Note:** Otherwise this module requires *stats* module and loads it if not present.

---

### 2.5.2 Exported metrics

To visualize the efficiency of the predictions, the module exports following statistics.

- `predict.epoch` - current prediction epoch (based on time of day and sampling window)
- `predict.queue` - number of queued queries in current window
- `predict.learned` - number of learned queries in current window

### 2.5.3 Properties

`predict.config` ({ window = 15, period = 24})

Reconfigure the predictor to given tracking window and period length. Both parameters are optional. Window

length is in minutes, period is a number of windows that can be kept in memory. e.g. if a window is 15 minutes, a period of “24” means 6 hours.

## 2.6 HTTP/2 services

This is a module that does the heavy lifting to provide an HTTP/2 enabled server that provides endpoint for other modules in order to enable them to export restful APIs and websocket streams. One example is statistics module that can stream live metrics on the website, or publish metrics on request for Prometheus scraper.

The server allows other modules to either use default endpoint that provides built-in webpage, restful APIs and websocket streams, or create new endpoints.

By default the server provides plain HTTP and TLS on the same port. See below if you want to use only one of these.

**Warning:** This module provides access to various API endpoints and must not be directly exposed to untrusted parties. Use [reverse-proxy](#) like [Apache](#) or [Nginx](#) if you need to authenticate API clients.

### 2.6.1 Example configuration

By default, the web interface starts HTTPS/2 on port 8053 using an ephemeral certificate that is valid for 90 days and is automatically renewed. It is of course self-signed. Why not use something like [Let's Encrypt](#)?

```
-- Load HTTP module with defaults
modules = {
  http = {
    host = 'localhost', -- Default: 'localhost'
    port = 8053,        -- Default: 8053
    geoip = 'GeoLite2-City.mmdb' -- Optional, see
    -- e.g. https://dev.maxmind.com/geoip/geoip2/geolite2/
    -- and install mmdblua library
    endpoints = {},
  }
}
```

Now you can reach the web services and APIs, done!

```
$ curl -k https://localhost:8053
$ curl -k https://localhost:8053/stats
```

### 2.6.2 Configuring TLS

You can disable unencrypted HTTP and enforce HTTPS by passing `tls = true` option.

```
http = {
  tls = true,
}
```

If you want to provide your own certificate and key, you're welcome to do so:

```
http = {
  cert = 'mycert.crt',
  key  = 'mykey.key',
}
```

The format of both certificate and key is expected to be PEM, e.g. equivalent to the outputs of following:

```
openssl ecparam -genkey -name prime256v1 -out mykey.key
openssl req -new -key mykey.key -out csr.pem
openssl req -x509 -days 90 -key mykey.key -in csr.pem -out mycert.crt
```

It is also possible to disable HTTPS altogether by passing `tls = false` option. Plain HTTP gets handy if you want to use [reverse-proxy](#) like [Apache](#) or [Nginx](#) for authentication to API etc. (Unencrypted HTTP could be fine for localhost tests as, for example, Safari doesn't allow WebSockets over HTTPS with a self-signed certificate. Major drawback is that current browsers won't do HTTP/2 over insecure connection.)

## 2.6.3 Built-in services

The HTTP module has several built-in services to use.

Endpoint	Service	Description
<code>/stats</code>	Statistics/metrics	Exported metrics in JSON.
<code>/metrics</code>	Prometheus metrics	Exported metrics for <a href="#">Prometheus</a>
<code>/trace/:name/:type</code>	Tracking	Trace resolution of the query and return the verbose logs.

## 2.6.4 Enabling Prometheus metrics endpoint

The module exposes `/metrics` endpoint that serves internal metrics in [Prometheus](#) text format. You can use it out of the box:

```
$ curl -k https://localhost:8053/metrics | tail
# TYPE latency histogram
latency_bucket{le=10} 2.000000
latency_bucket{le=50} 2.000000
latency_bucket{le=100} 2.000000
latency_bucket{le=250} 2.000000
latency_bucket{le=500} 2.000000
latency_bucket{le=1000} 2.000000
latency_bucket{le=1500} 2.000000
latency_bucket{le=+Inf} 2.000000
latency_count 2.000000
latency_sum 11.000000
```

You can namespace the metrics in configuration, using `http.prometheus.namespace` attribute:

```
http = {
  host = 'localhost',
}

-- Set Prometheus namespace
http.prometheus.namespace = 'resolver_'
```

You can also add custom metrics or rewrite existing metrics before they are returned to Prometheus client.

```

http = {
    host = 'localhost',
}

-- Add an arbitrary metric to Prometheus
http.prometheus.finalize = function (metrics)
    table.insert(metrics, 'build_info{version="1.2.3"} 1')
end

```

## 2.6.5 Tracing requests

With the `/trace` endpoint you can trace various aspects of the request execution. The basic mode allows you to resolve a query and trace verbose logs (and messages received):

```

$ curl https://localhost:8053/trace/e.root-servers.net
[ 8138] [iter] 'e.root-servers.net.' type 'A' created outbound query, parent id 0
[ 8138] [ rc ] => rank: 020, lowest 020, e.root-servers.net. A
[ 8138] [ rc ] => satisfied from cache
[ 8138] [iter] <= answer received:
;; ->>HEADER<<- opcode: QUERY; status: NOERROR; id: 8138
;; Flags: qr aa QUERY: 1; ANSWER: 0; AUTHORITY: 0; ADDITIONAL: 0

;; QUESTION SECTION
e.root-servers.net.      A

;; ANSWER SECTION
e.root-servers.net.    3556353 A      192.203.230.10

[ 8138] [iter] <= rcode: NOERROR
[ 8138] [resl] finished: 4, queries: 1, mempool: 81952 B

```

## 2.6.6 How to expose services over HTTP

The module provides a table `endpoints` of already existing endpoints, it is free for reading and writing. It contains tables describing a triplet - `{mime, on_serve, on_websocket}`. In order to register a new service, simply add it to the table:

```

local on_health = {'application/json'},
function (h, stream)
    -- API call, return a JSON table
    return {state = 'up', uptime = 0}
end,
function (h, ws)
    -- Stream current status every second
    local ok = true
    while ok do
        local push = tojson('up')
        ok = ws:send(tojson({'up'}))
        require('cqueues').sleep(1)
    end
    -- Finalize the WebSocket
    ws:close()
end}
-- Load module

```

(continues on next page)

(continued from previous page)

```
modules = {
  http = {
    endpoints = { ['/health'] = on_health }
  }
}
```

Then you can query the API endpoint, or tail the WebSocket using curl.

```
$ curl -k https://localhost:8053/health
{"state":"up","uptime":0}
$ curl -k -i -N -H "Connection: Upgrade" -H "Upgrade: websocket" -H "Host:
↪localhost:8053/health" -H "Sec-WebSocket-Key: nope" -H "Sec-WebSocket-Version: 13"
↪https://localhost:8053/health
HTTP/1.1 101 Switching Protocols
upgrade: websocket
sec-websocket-accept: eg18mwU7CDRGUF1Q+EJwPM335eM=
connection: upgrade

?["up"]?["up"]?["up"]
```

Since the stream handlers are effectively coroutines, you are free to keep state and yield using cqueues. This is especially useful for WebSockets, as you can stream content in a simple loop instead of chains of callbacks.

Last thing you can publish from modules are “*snippets*”. Snippets are plain pieces of HTML code that are rendered at the end of the built-in webpage. The snippets can be extended with JS code to talk to already exported restful APIs and subscribe to WebSockets.

```
http.snippets['/health'] = {'Health service', '<p>UP!</p>'}
```

## 2.6.7 How to expose RESTful services

A RESTful service is likely to respond differently to different type of methods and requests, there are three things that you can do in a service handler to send back results. First is to just send whatever you want to send back, it has to respect MIME type that the service declared in the endpoint definition. The response code would then be 200 OK, any non-string responses will be packed to JSON. Alternatively, you can respond with a number corresponding to the HTTP response code or send headers and body yourself.

```
-- Our upvalue
local value = 42

-- Expose the service
local service = {'application/json',
function (h, stream)
  -- Get request method and deal with it properly
  local m = h:get(':method')
  local path = h:get(':path')
  log('[service] method %s path %s', m, path)
  -- Return table, response code will be '200 OK'
  if m == 'GET' then
    return {key = path, value = value}
  -- Save body, perform check and either respond with 505 or 200 OK
  elseif m == 'POST' then
    local data = stream:get_body_as_string()
    if not tonumber(data) then
      return 500, 'Not a good request'
```

(continues on next page)

(continued from previous page)

```

        end
        value = tonumber(data)
        -- Unsupported method, return 405 Method not allowed
    else
        return 405, 'Cannot do that'
    end
end}
-- Load the module
modules = {
    http = {
        endpoints = { ['/service'] = service }
    }
}

```

In some cases you might need to send back your own headers instead of default provided by HTTP handler, you can do this, but then you have to return `false` to notify handler that it shouldn't try to generate a response.

```

local headers = require('http.headers')
function (h, stream)
    -- Send back headers
    local hsend = headers.new()
    hsend:append(':status', '200')
    hsend:append('content-type', 'binary/octet-stream')
    assert(stream:write_headers(hsend, false))
    -- Send back data
    local data = 'binary-data'
    assert(stream:write_chunk(data, true))
    -- Disable default handler action
    return false
end

```

## 2.6.8 How to expose more interfaces

Services exposed in the previous part share the same external interface. This means that it's either accessible to the outside world or internally, but not one or another. This is not always desired, i.e. you might want to offer DNS/HTTPS to everyone, but allow application firewall configuration only on localhost. `http` module allows you to create additional interfaces with custom endpoints for this purpose.

```

http.add_interface {
    endpoints = {
        ['/conf'] = {
            'application/json', function (h, stream)
                return 'configuration API\n'
            end
        },
    },
    -- Same options as the config() method
    host = 'localhost',
    port = '8054',
}

```

This way you can have different internal-facing and external-facing services at the same time.

## 2.6.9 Dependencies

- lua-http (>= 0.1) available in LuaRocks

If you're installing via Homebrew on OS X, you need OpenSSL too.

```
$ brew update
$ brew install openssl
$ brew link openssl --force # Override system OpenSSL
```

Any other system can install from LuaRocks directly:

```
$ luarocks install http
```

- mmdblua available in LuaRocks

```
$ luarocks install --server=https://luarocks.org/dev mmdblua
$ curl -O https://geolite.maxmind.com/download/geoip/database/GeoLite2-
  ↪City.mmdb.gz
$ gzip -d GeoLite2-City.mmdb.gz
```

## 2.7 DNS Application Firewall

This module is a high-level interface for other powerful filtering modules and DNS views. It provides an easy interface to apply and monitor DNS filtering rules and a persistent memory for them. It also provides a restful service interface and an HTTP interface.

### 2.7.1 Example configuration

Firewall rules are declarative and consist of filters and actions. Filters have `field operator operand` notation (e.g. `qname = example.com`), and may be chained using AND/OR keywords. Actions may or may not have parameters after the action name.

```
-- Let's write some daft rules!
modules = { 'daf' }

-- Block all queries with QNAME = example.com
daf.add 'qname = example.com deny'

-- Filters can be combined using AND/OR...
-- Block all queries with QNAME match regex and coming from given subnet
daf.add 'qname ~ %w+.example.com AND src = 192.0.2.0/24 deny'

-- We also can reroute addresses in response to alternate target
-- This reroutes 1.2.3.4 to localhost
daf.add 'src = 127.0.0.0/8 reroute 192.0.2.1-127.0.0.1'

-- Subnets work too, this reroutes a whole subnet
-- e.g. 192.0.2.55 to 127.0.0.55
daf.add 'src = 127.0.0.0/8 reroute 192.0.2.0/24-127.0.0.0'

-- This rewrites all A answers for 'example.com' from
-- whatever the original address was to 127.0.0.2
daf.add 'src = 127.0.0.0/8 rewrite example.com A 127.0.0.2'
```

(continues on next page)



(continued from previous page)

```

-- Mirror queries matching given name to DNS logger
daf.add 'qname ~ %w+.example.com mirror 127.0.0.2'
daf.add 'qname ~ example-%d.com mirror 127.0.0.3@5353'

-- Forward queries from subnet
daf.add 'src = 127.0.0.1/8 forward 127.0.0.1@5353'
-- Forward to multiple targets
daf.add 'src = 127.0.0.1/8 forward 127.0.0.1@5353,127.0.0.2@5353'

-- Truncate queries based on destination IPs
daf.add 'dst = 192.0.2.51 truncate'

-- Disable a rule
daf.disable 2
-- Enable a rule
daf.enable 2
-- Delete a rule
daf.del 2

```

If you're not sure what firewall rules are in effect, see `daf.rules`:

```

-- Show active rules
> daf.rules
[1] => {
  [rule] => {
    [count] => 42
    [id] => 1
    [cb] => function: 0x1a3eda38
  }
  [info] => qname = example.com AND src = 127.0.0.1/8 deny
  [policy] => function: 0x1a3eda38
}
[2] => {
  [rule] => {
    [suspended] => true
    [count] => 123522
    [id] => 2
    [cb] => function: 0x1a3ede88
  }
  [info] => qname ~ %w+.facebook.com AND src = 127.0.0.1/8 deny...
  [policy] => function: 0x1a3ede88
}

```

## 2.7.2 Web interface

If you have *HTTP/2* loaded, the firewall automatically loads as a snippet. You can create, track, suspend and remove firewall rules from the web interface. If you load both modules, you have to load *daf* after *http*.

## 2.7.3 RESTful interface

The module also exports a RESTful API for operations over rule chains.

URL	HTTP Verb	Action
/daf	GET	Return JSON list of active rules.
/daf	POST	Insert new rule, rule string is expected in body. Returns rule information in JSON.
/daf/<id>	GET	Retrieve a rule matching given ID.
/daf/<id>	DELETE	Delete a rule matching given ID.
/daf/<id>/<prop>/<val>	PATCH	Modify given rule, for example /daf/3/active/false suspends rule 3.

This interface is used by the web interface for all operations, but you can also use it directly for testing.

```
# Get current rule set
$ curl -s -X GET http://localhost:8053/daf | jq .
{}

# Create new rule
$ curl -s -X POST -d "src = 127.0.0.1 pass" http://localhost:8053/daf | jq .
{
  "count": 0,
  "active": true,
  "info": "src = 127.0.0.1 pass",
  "id": 1
}

# Disable rule
$ curl -s -X PATCH http://localhost:8053/daf/1/active/false | jq .
true

# Retrieve a rule information
$ curl -s -X GET http://localhost:8053/daf/1 | jq .
{
  "count": 4,
  "active": true,
  "info": "src = 127.0.0.1 pass",
  "id": 1
}

# Delete a rule
$ curl -s -X DELETE http://localhost:8053/daf/1 | jq .
true
```

## 2.8 Rebinding protection

This module provides protection from [DNS Rebinding attack](#) by blocking answers which contain IPv4 or IPv6 addresses for private use (or some other special-use addresses).

To enable this module insert following line into your configuration file:

```
modules.load('rebinding < iterate')
```

Please note that this module does not offer stable configuration interface yet. For this reason it is suitable mainly for public resolver operators who do not need to whitelist certain subnets.

**Warning:** Some like to “misuse” such addresses, e.g. `127.*.*` in blacklists served over DNS, and this module will block such uses.

## 2.9 Graphite module

The module sends statistics over the [Graphite](#) protocol to either [Graphite](#), [Metronome](#), [InfluxDB](#) or any compatible storage. This allows powerful visualization over metrics collected by Knot Resolver.

**Tip:** The Graphite server is challenging to get up and running, [InfluxDB](#) combined with [Grafana](#) are much easier, and provide richer set of options and available front-ends. [Metronome](#) by PowerDNS alternatively provides a mini-graphite server for much simpler setups.

### 2.9.1 Example configuration

Only the `host` parameter is mandatory.

By default the module uses UDP so it doesn't guarantee the delivery, set `tcp = true` to enable Graphite over TCP. If the TCP consumer goes down or the connection with Graphite is lost, resolver will periodically attempt to reconnect with it.

```
modules = {
  graphite = {
    prefix = hostname(), -- optional metric prefix
    host = '127.0.0.1', -- graphite server address
    port = 2003, -- graphite server port
    interval = 5 * sec, -- publish interval
    tcp = false -- set to true if want TCP mode
  }
}
```

The module supports sending data to multiple servers at once.

```
modules = {
  graphite = {
    host = { '127.0.0.1', '1.2.3.4', '::1' },
  }
}
```

### 2.9.2 Dependencies

- [luasocket](#) available in LuaRocks

```
$ luarocks install luasocket
```

## 2.10 Etcd module

The module connects to Etcd peers and watches for configuration change. By default, the module looks for the subtree under `/knot-resolver` directory, but you can change this in the configuration.

The subtree structure corresponds to the configuration variables in the declarative style.

```
$ etcdctl set /knot-resolvevr/net/127.0.0.1 53
$ etcdctl set /knot-resolver/cache/size 10000000
```

Configures all listening nodes to following configuration:

```
net = { '127.0.0.1' }
cache.size = 10000000
```

### 2.10.1 Example configuration

```
modules = {
  etcd = {
    prefix = '/knot-resolver',
    peer = 'http://127.0.0.1:7001'
  }
}
```

**Warning:** Work in progress!

### 2.10.2 Dependencies

- lua-etcd available in LuaRocks

```
$ luarocks install etcd --from=https://mah0x211.github.io/rocks/
```

## 2.11 DNS64

The module for [RFC 6147](#) DNS64 AAAA-from-A record synthesis, it is used to enable client-server communication between an IPv6-only client and an IPv4-only server. See the well written [introduction](#) in the PowerDNS documentation. If no address is passed (i.e. nil), the well-known prefix `64:ff9b::` is used.

**Warning:** The module currently won't work well with *policy.STUB*. Also, the IPv6 passed in configuration is assumed to be /96, and PTR synthesis and “exclusion prefixes” aren't implemented.

---

**Tip:** The A record sub-requests will be DNSSEC secured, but the synthetic AAAA records can't be. Make sure the last mile between stub and resolver is secure to avoid spoofing.

---

### 2.11.1 Example configuration

```
-- Load the module with a NAT64 address
modules = { dns64 = 'fe80::21b:77ff:0:0' }
-- Reconfigure later
dns64.config('fe80::21b:aabb:0:0')
```

## 2.12 Renumber

The module renumbers addresses in answers to different address space. e.g. you can redirect malicious addresses to a blackhole, or use private address ranges in local zones, that will be remapped to real addresses by the resolver.

**Warning:** While requests are still validated using DNSSEC, the signatures are stripped from final answer. The reason is that the address synthesis breaks signatures. You can see whether an answer was valid or not based on the AD flag.

### 2.12.1 Example configuration

```
modules = {
  renumber = {
    -- Source subnet, destination subnet
    {'10.10.10.0/24', '192.168.1.0'},
    -- Remap /16 block to localhost address range
    {'166.66.0.0/16', '127.0.0.0'}
  }
}
```

## 2.13 DNSSEC validation failure logging

This module adds error message for each DNSSEC validation failure. It is meant to provide hint to operators which queries should be investigated using diagnostic tools like [DNSViz](#).

Add following line to your configuration file to enable it:

```
modules.load('bogus_log')
```

Example of error message logged by this module:

```
DNSSEC validation failure dnssec-failed.org. DNSKEY
```

Please note that in future this module might be replaced with some other way to log this information.

## 2.14 Workarounds

A simple module that alters resolver behavior on specific broken sub-domains. Currently it mainly disables case randomization on them.

### 2.14.1 Running

```
modules = { 'workarounds < iterate' }
```

## 2.15 Dnstap

Dnstap module currently supports logging dns responses to a unix socket in dnstap format using fstrm framing library. The unix socket and the socket reader should be present before starting kresd.

### 2.15.1 Configuration

Tunables:

- `socket_path`: the the unix socket file where dnstap messages will be sent
- `log_responses`: if true responses in wire format will be logged

```
modules = {
  dnstap = {
    socket_path = "/tmp/dnstap.sock",
    log_responses = true
  }
}
```

## 2.16 Signaling Trust Anchor Knowledge in DNSSEC

The module for Signaling Trust Anchor Knowledge in DNSSEC Using Key Tag Query, implemented according to [RFC 8145#section-5](#).

This feature allows validating resolvers to signal to authoritative servers which keys are referenced in their chain of trust. The data from such signaling allow zone administrators to monitor the progress of rollovers in a DNSSEC-signed zone.

This mechanism serve to measure the acceptance and use of new DNSSEC trust anchors and key signing keys (KSKs). This signaling data can be used by zone administrators as a gauge to measure the successful deployment of new keys. This is of particular interest for the DNS root zone in the event of key and/or algorithm rollovers that rely on [RFC 5011](#) to automatically update a validating DNS resolver's trust anchor.

This module is enabled by default. You may use `modules.unload('ta_signal_query')` in your configuration.

## 2.17 Sentinel for Detecting Trusted Root Keys

The module implementing A Root Key Trust Anchor Sentinel for DNSSEC according to [draft-ietf-dnsop-kskroll-sentinel-12](#).

This feature allows users of validating resolver to detect which root keys are configured in their chain of trust. The data from such signaling are necessary to monitor the progress of the DNSSEC root key rollover.

This module is enabled by default and we urge users not to disable it. If it is absolutely necessary you may add `modules.unload('ta_sentinel')` to your configuration to disable it.

## 2.18 Priming module

The module for Initializing a DNS Resolver with Priming Queries implemented according to [RFC 8109](#). Purpose of the module is to keep up-to-date list of root DNS servers and associated IP addresses.

Result of successful priming query replaces root hints distributed with the resolver software. Unlike other DNS resolvers, Knot Resolver caches result of priming query on disk and keeps the data between restarts until TTL expires.

This module is enabled by default and it is not recommended to disable it. For debugging purposes you may disable the module by appending `modules.unload('priming')` to your configuration.

## 2.19 System time skew detector

This module compares local system time with inception and expiration time bounds in DNSSEC signatures for . NS records. If the local system time is outside of these bounds, it is likely a misconfiguration which will cause all DNSSEC validation (and resolution) to fail.

In case of mismatch, a warning message will be logged to help with further diagnostics.

**Warning:** Information printed by this module can be forged by a network attacker! System administrator **MUST** verify values printed by this module and fix local system time using a trusted source.

This module is useful for debugging purposes. It runs only once during resolver start does not anything after that. It is enabled by default. You may disable the module by appending `modules.unload('detect_time_skew')` to your configuration.

## 2.20 Detect discontinuous jumps in the system time

This module detect discontinuous jumps in the system time when resolver is running. It clears cache when a significant backward time jumps occurs.

Time jumps are usually created by NTP time change or by admin intervention. These change can affect cache records as they store timestamp and TTL in real time.

If you want to preserve cache during time travel you should disable this module by `modules.unload('detect_time_jump')`.

Due to the way monotonic system time works on typical systems, suspend-resume cycles will be perceived as forward time jumps, but this direction of shift does not have the risk of using records beyond their intended TTL, so forward jumps do not cause erasing the cache.

## 2.21 Root on loopback (RFC 7706)

Knot Resolver developers decided that pure implementation of [RFC 7706](#) is a bad idea so it is not implemented in the form envisioned by the RFC. You can get the very similar effect without its downsides by combining [prefill](#) and [serve\\_stale](#) modules with Aggressive Use of DNSSEC-Validated Cache ([RFC 8198](#)) behavior which is enabled automatically together with DNSSEC validation.

## 2.22 Cache prefilling

This module provides ability to periodically prefill DNS cache by importing root zone data obtained over HTTPS.

Intended users of this module are big resolver operators which will benefit from decreased latencies and smaller amount of traffic towards DNS root servets.

Example configuration is:

```
modules.load('prefill')
prefill.config({
    ['.'] = {
        url = 'https://www.internic.net/domain/root.zone',
        ca_file = '/etc/pki/tls/certs/ca-bundle.crt',
        interval = 86400 -- seconds
    }
})
```

This configuration downloads zone file from URL `https://www.internic.net/domain/root.zone` and imports it into cache every 86400 seconds (1 day). The HTTPS connection is authenticated using CA certificate from file `/etc/pki/tls/certs/ca-bundle.crt` and signed zone content is validated using DNSSEC.

Root zone to import must be signed using DNSSEC and the resolver must have valid DNSSEC configuration. (For further details please see [Enabling DNSSEC](#).)

Parameter	Description
<code>ca_file</code>	path to CA certificate bundle used to authenticate the HTTPS connection
<code>interval</code>	number of seconds between zone data refresh attempts
<code>url</code>	URL of a file in <a href="#">RFC 1035</a> zone file format

Only root zone import is supported at the moment.

### 2.22.1 Dependencies

Depends on the `luasec` library.

## 2.23 Serve stale

Demo module that allows using timed-out records in case kresd is unable to contact upstream servers.

By default it allows stale-ness by up to one day, after roughly four seconds trying to contact the servers. It's quite configurable/flexible; see the beginning of the module source for details. See also the RFC [draft](#) (not fully followed) and `cache.ns_tout`.

### 2.23.1 Running

```
modules = { 'serve_stale < cache' }
```



### 3.1 Installing from packages

The resolver is packaged for Debian, Fedora+EPEL, Ubuntu, Docker, NixOS/NixPkgs, FreeBSD, HomeBrew, and Turris Omnia. Some of these are maintained directly by the knot-resolver team.

Refer to [project page](#) for information about installing from packages. If packages are not available for your OS, see following sections to see how you can build it from sources (or package it), or use official [Docker images](#).

### 3.2 Platform considerations

Knot-resolver is written for UNIX-like systems, mainly in C99. Portable I/O is provided by [libuv](#). Some 64-bit systems with LuaJIT 2.1 may be affected by a [problem](#) – Linux on x86\_64 is unaffected but [Linux on aarch64](#) is.

Windows systems might theoretically work without large changes, but it's most likely broken and currently not planned to be supported.

### 3.3 Requirements

The following is a list of software required to build Knot Resolver from sources.

Requirement	Required by	Notes
<a href="#">GNU Make 3.80+</a>	<i>all</i>	<i>(build only)</i>
C and C++ compiler	<i>all</i>	<i>(build only)</i> <sup>1</sup>
<a href="#">pkg-config</a>	<i>all</i>	<i>(build only)</i> <sup>2</sup>
<a href="#">hexdump</a> or <a href="#">xxd</a>	daemon	<i>(build only)</i>
<a href="#">libknot 2.7.2+</a>	<i>all</i>	Knot DNS libraries - requires autotools, GnuTLS, ...
<a href="#">LuaJIT 2.0+</a>	daemon	Embedded scripting language.
<a href="#">libuv 1.7+</a>	daemon	Multiplatform I/O and services ( <a href="#">libuv 1.0</a> with limitations <sup>3</sup> ).
<a href="#">lmdb</a>	daemon	If missing, a static version is embedded.

There are also *optional* packages that enable specific functionality in Knot Resolver, they are useful mainly for developers to build documentation and tests.

Optional	Needed for	Notes
lua-http	modules/http	HTTP/2 client/server for Lua.
luasocket	trust anchors, modules/stats	Sockets for Lua.
luasec	trust anchors	TLS for Lua.
cmocka	unit tests	Unit testing framework.
Doxygen	documentation	Generating API documentation.
Sphinx and sphinx_rtd_theme	documentation	Building this HTML/PDF documentation.
breathe	documentation	Exposing Doxygen API doc to Sphinx.
libsystemd	daemon	Systemd socket activation support.
libprotobuf 3.0+	modules/dnstap	Protocol Buffers support for dnstap.
libprotobuf-c 1.0+	modules/dnstap	C bindings for Protobuf.
libfstrm 0.2+	modules/dnstap	Frame Streams data transport protocol.
luacheck	lint-lua	Syntax and static analysis checker for Lua.
clang-tidy	lint-c	Syntax and static analysis checker for C.
luacov	check-config	Code coverage analysis for Lua modules.

### 3.3.1 Packaged dependencies

Most of the dependencies can be resolved from packages, here's an overview for several platforms.

- **Debian** (since *sid*) - current stable doesn't have libknot and libuv, which must be installed from sources.

```
sudo apt-get install pkg-config libknot-dev libuv1-dev libcmocka-dev libluajit-5.1-dev
```

- **Ubuntu** - unknown.
- **Fedora**

```
# minimal build
sudo dnf install @buildsys-build knot-devel libuv-devel luajit-devel
# unit tests
sudo dnf install libcmocka-devel
# integration tests
sudo dnf install cmake git python-dns python-jinja2
# optional features
sudo dnf install lua-sec-compat lua-socket-compat systemd-devel
# docs
sudo dnf install doxygen python-breathe python-sphinx
```

- **RHEL/CentOS** - unknown.
- **openSUSE** - there is an [experimental package](#).

<sup>1</sup> Requires C99, `__attribute__((cleanup))` and `-MMD -MP` for dependency file generation. GCC, Clang and ICC are supported.

<sup>2</sup> You can use variables `<dependency>_CFLAGS` and `<dependency>_LIBS` to configure dependencies manually (i.e. `libknot_CFLAGS` and `libknot_LIBS`).

<sup>3</sup> libuv 1.7 brings `SO_REUSEPORT` support that is needed for multiple forks. libuv < 1.7 can be still used, but only in single-process mode. Use *different method* for load balancing.

- **FreeBSD** - when installing from ports, all dependencies will install automatically, corresponding to the selected options.
- **NetBSD** - unknown.
- **OpenBSD** - unknown.
- **Mac OS X** - the dependencies can be found through [Homebrew](#).

```
brew install pkg-config libuv luajit cmocka
```

## 3.4 Building from sources

Initialize git submodules first.

```
$ git submodule update --init --recursive
```

The Knot Resolver depends on the the Knot DNS library, recent version of [libuv](#), and [LuaJIT](#).

```
$ make info # See what's missing
```

When you have all the dependencies ready, you can build and install.

```
$ make PREFIX="/usr/local"
$ make install PREFIX="/usr/local"
```

**Note:** Always build with `PREFIX` if you want to install, as it is hardcoded in the executable for module search path. Production code should be compiled with `-DNDEBUG`. If you build the binary with `-DNOVERBOSELOG`, it won't be possible to turn on verbose logging; we advise packagers against using that flag.

**Note:** If you build with `PREFIX`, you may need to also set the `LDFLAGS` for the libraries:

```
make LDFLAGS="-Wl,-rpath=/usr/local/lib" PREFIX="/usr/local"
```

Alternatively you can build only specific parts of the project, i.e. `library`.

```
$ make lib
$ make lib-install
```

**Note:** Documentation is not built by default, run `make doc` to build it.

### 3.4.1 Building with security compiler flags

Knot Resolver enables certain [security compile-time flags](#) that do not affect performance. You can add more flags to the build by appending them to `CFLAGS` variable, e.g. `make CFLAGS="-fstack-protector"`.

Method	Status	Notes
-fstack-protector	<i>dis-abled</i>	(must be specifically enabled in CFLAGS)
- D_FORTIFY_SOURCE=2	<b>en-abled</b>	
-pie	<b>en-abled</b>	enables ASLR for kresd (disable with make HARDENING=no)
RELRO	<b>en-abled</b>	full <sup>4</sup>

You can also disable linker hardening when it's unsupported with `make HARDENING=no`.

### 3.4.2 Building for packages

The build system supports `DESTDIR`

```
$ make install DESTDIR=/tmp/stage
```

**Tip:** There is a template for service file and AppArmor profile to help you kickstart the package.

### 3.4.3 Default paths

The default installation follows FHS with several custom paths for configuration and modules. All paths are prefixed with `PREFIX` variable by default if not specified otherwise.

Component	Variable	Default	Notes
library	LIBDIR	<code>\$(PREFIX)/lib</code>	pkg-config is auto-generated <sup>5</sup>
daemon	SBINDIR	<code>\$(PREFIX)/sbin</code>	
configuration	ETCDIR	<code>\$(PREFIX)/etc/knot-resolver</code>	Configuration file, templates.
modules	MODULEDIR	<code>\$(LIBDIR)/kdns_modules</code>	Runtime directory for loading dynamic modules <sup>6</sup> .
trust anchor file	KEYFILE_DEFAULT	<code>\$(PREFIX)/etc/knot-resolver/KEYFILE_DEFAULT</code>	Path to read-only trust anchor file, which is used as fallback when no other file is specified. <sup>7</sup>
work directory		the current directory	Run directory for daemon. (Only relevant during run time, not e.g. during installation.)

**Note:** Each module is self-contained and may install additional bundled files within `$(MODULEDIR)/$(modulename)`. These files should be read-only, non-executable.

<sup>4</sup> See `checksec.sh`

<sup>5</sup> The `libkres.pc` is installed in `$(LIBDIR)/pkgconfig`.

<sup>6</sup> The default `moduledir` can be changed with `-m` option to `kresd` daemon or by calling `moduledir()` function from lua.

<sup>7</sup> If no other trust anchor is specified by user, the compiled-in path `KEYFILE_DEFAULT` must contain a valid trust anchor. This is typically used by distributions which provide DNSSEC root trust anchors as part of distribution package. Users can disable the built-in trust anchor by adding `trust_anchors.keyfile_default = nil` to their configuration.

### 3.4.4 Static or dynamic?

By default the resolver library is built as a dynamic library with versioned ABI. You can revert to static build with `BUILDMODE` variable.

```
$ make BUILDMODE=dynamic # Default, create dynamic library
$ make BUILDMODE=static # Create static library
```

When the library is linked statically, it usually produces a smaller binary. However linking it to various C modules might violate ODR and increase the size.

### 3.4.5 Resolving dependencies

The build system relies on `pkg-config` to find dependencies. You can override it to force custom versions of the software by environment variables.

```
$ make libknot_CFLAGS="-I/opt/include" libknot_LIBS="-L/opt/lib -lknot -ldnssec"
```

Optional dependencies may be disabled as well using `HAS_x=yes|no` variable.

```
$ make HAS_go=no HAS_cmocka=no
```

**Warning:** If the dependencies lie outside of library search path, you need to add them somehow. Try `LD_LIBRARY_PATH` on Linux/BSD, and `DYLD_FALLBACK_LIBRARY_PATH` on OS X. Otherwise you need to add the locations to linker search path.

Several dependencies may not be in the packages yet, the script pulls and installs all dependencies in a chroot. You can avoid rebuilding dependencies by specifying `BUILD_IGNORE` variable, see the `Dockerfile` for example. Usually you only really need to rebuild `libknot`.

```
$ export FAKEROOT="${HOME}/.local"
$ export PKG_CONFIG_PATH="${FAKEROOT}/lib/pkgconfig"
$ export BUILD_IGNORE="..." # Ignore installed dependencies
$ ./scripts/bootstrap-dependends.sh ${FAKEROOT}
```

### 3.4.6 Building extras

The project can be built with code coverage tracking using the `COVERAGE=1` variable.

The `make coverage` target gathers both `gcov` code coverage for C files, and `luacov` code coverage for Lua files and merges it for analysis. It requires `lcov` to be installed.

```
$ make coverage
```

### 3.4.7 Running unit and integration tests

The linter requires `luacheck` and `clang-tidy` and is executed by `make lint`. The unit tests require `cmocka` and are executed by `make check`. Tests for the `dnstap` module need `go` and are executed by `make check-dnstap`.

The integration tests use `Deckard`, the `DNS test harness`.

```
$ make check-integration
```

Note that the daemon and modules must be installed first before running integration tests, the reason is that the daemon is otherwise unable to find and load modules.

Read the [documentation](#) for more information about requirements, how to run it and extend it.

### 3.5 Getting Docker image

Docker images require only either Linux or a Linux VM (see [boot2docker](#) on OS X).

```
$ docker run cznic/knot-resolver
```

See the [Docker images](#) page for more information and options. You can hack on the container by changing the container entrypoint to shell like:

```
$ docker run -it --entrypoint=/bin/bash cznic/knot-resolver
```

---

**Tip:** You can build the Docker image yourself with `docker build -t knot-resolver scripts`.

---

## 4.1 Requirements

- `libknot 2.0` (Knot DNS high-performance DNS library.)

## 4.2 For users

The library as described provides basic services for name resolution, which should cover the usage, examples are in the *resolve API* documentation.

---

**Tip:** If you're migrating from `getaddrinfo()`, see “*synchronous*” API, but the library offers iterative API as well to plug it into your event loop for example.

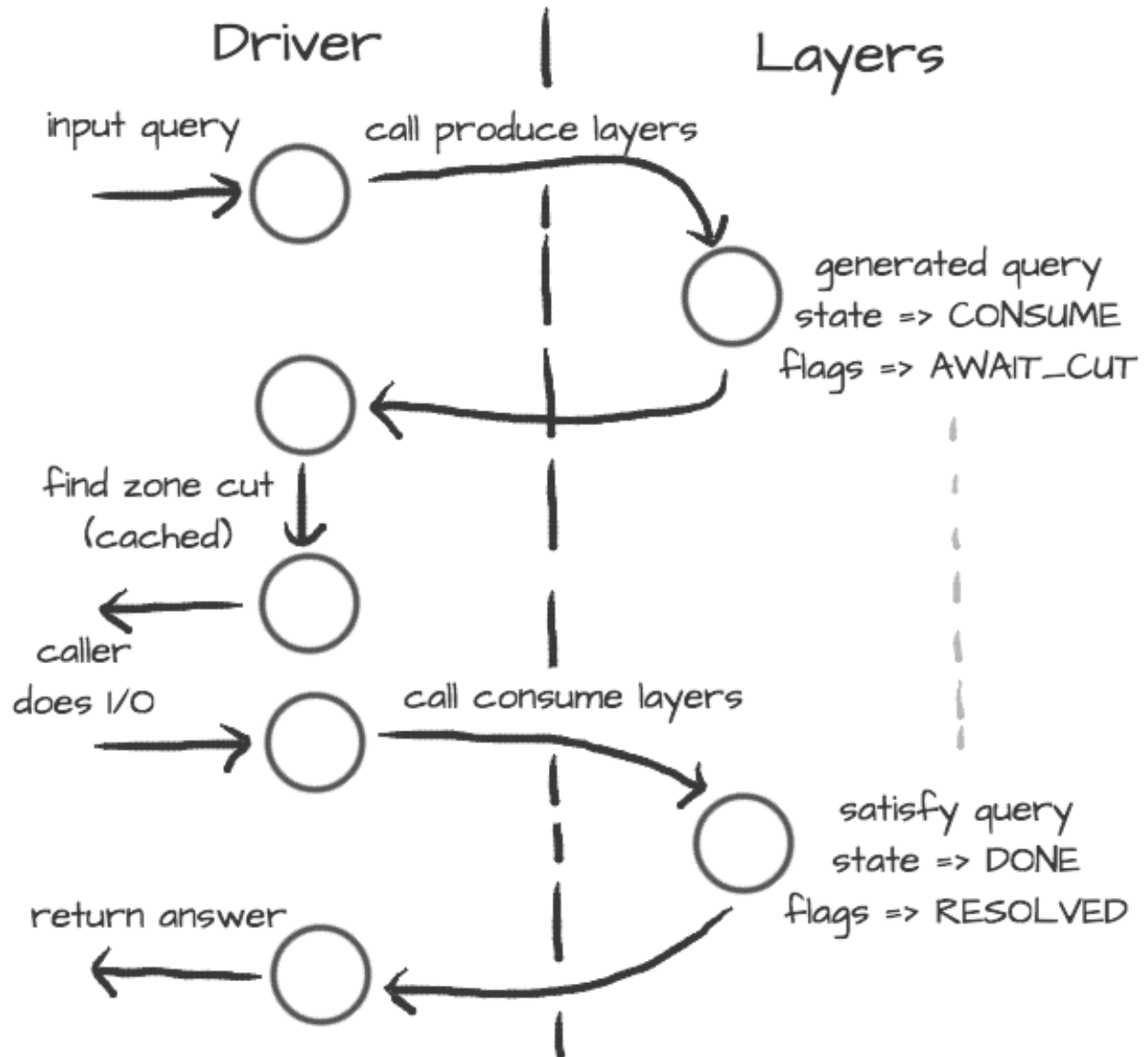
---

## 4.3 For developers

The resolution process starts with the functions in *resolve.c*, they are responsible for:

- reacting to state machine state (i.e. calling consume layers if we have an answer ready)
- interacting with the library user (i.e. asking caller for I/O, accepting queries)
- fetching assets needed by layers (i.e. zone cut)

This is the *driver*. The driver is not meant to know “*how*” the query resolves, but rather “*when*” to execute “*what*”.



On the other side are *layers*. They are responsible for dissecting the packets and informing the driver about the results. For example, a *produce* layer generates query, a *consume* layer validates answer.

**Tip:** Layers are executed asynchronously by the driver. If you need some asset beforehand, you can signalize the driver using returning state or current query flags. For example, setting a flag `AWAIT_CUT` forces driver to fetch zone cut information before the packet is consumed; setting a `RESOLVED` flag makes it pop a query after the current set of layers is finished; returning `FAIL` state makes it fail current query.

Layers can also change course of resolution, for example by appending additional queries.

```
consume = function (state, req, answer)
    answer = kres.pkt_t(answer)
    if answer:qtype() == kres.type.NS then
```

(continues on next page)



(continued from previous page)

```

        req = kres.request_t(req)
        local qry = req:push(answer:qname(), kres.type.SOA, kres.class.IN)
        qry.flags.AWAIT_CUT = true
    end
    return state
end

```

This **doesn't** block currently processed query, and the newly created sub-request will start as soon as driver finishes processing current. In some cases you might need to issue sub-request and process it **before** continuing with the current, i.e. validator may need a DNSKEY before it can validate signatures. In this case, layers can yield and resume afterwards.

```

consume = function (state, req, answer)
    answer = kres.pkt_t(answer)
    if state == kres.YIELD then
        print('continuing yielded layer')
        return kres.DONE
    else
        if answer:qtype() == kres.type.NS then
            req = kres.request_t(req)
            local qry = req:push(answer:qname(), kres.type.SOA, kres.
↪class.IN)

            qry.flags.AWAIT_CUT = true
            print('planned SOA query, yielding')
            return kres.YIELD
        end
        return state
    end
end
end

```

The YIELD state is a bit special. When a layer returns it, it interrupts current walk through the layers. When the layer receives it, it means that it yielded before and now it is resumed. This is useful in a situation where you need a sub-request to determine whether current answer is valid or not.

## 4.4 Writing layers

**Warning:** FIXME: this dev-docs section is outdated! Better see comments in files instead, for now.

The resolver *library* leverages the processing API from the libknot to separate packet processing code into layers.

**Note:** This is only crash-course in the library internals, see the resolver *library* documentation for the complete overview of the services.

The library offers following services:

- *Cache* - MVCC cache interface for retrieving/storing resource records.
- *Resolution plan* - Query resolution plan, a list of partial queries (with hierarchy) sent in order to satisfy original query. This contains information about the queries, nameserver choice, timing information, answer and its class.
- *Nameservers* - Reputation database of nameservers, this serves as an aid for nameserver choice.

A processing layer is going to be called by the query resolution driver for each query, so you're going to work with `struct kr_request` as your per-query context. This structure contains pointers to resolution context, resolution plan and also the final answer.

```
int consume(kr_layer_t *ctx, knot_pkt_t *pkt)
{
    struct kr_request *req = ctx->req;
    struct kr_query *qry = req->current_query;
}
```

This is only passive processing of the incoming answer. If you want to change the course of resolution, say satisfy a query from a local cache before the library issues a query to the nameserver, you can use states (see the *Static hints* for example).

```
int produce(kr_layer_t *ctx, knot_pkt_t *pkt)
{
    struct kr_request *req = ctx->req;
    struct kr_query *qry = req->current_query;

    /* Query can be satisfied locally. */
    if (can_satisfy(qry)) {
        /* This flag makes the resolver move the query
         * to the "resolved" list. */
        qry->flags.RESOLVED = true;
        return KR_STATE_DONE;
    }

    /* Pass-through. */
    return ctx->state;
}
```

It is possible to not only act during the query resolution, but also to view the complete resolution plan afterwards. This is useful for analysis-type tasks, or “*per answer*” hooks.

```
int finish(kr_layer_t *ctx)
{
    struct kr_request *req = ctx->req;
    struct kr_rplan *rplan = req->rplan;

    /* Print the query sequence with start time. */
    char qname_str[KNOT_DNAME_MAXLEN];
    struct kr_query *qry = NULL
    WALK_LIST(qry, rplan->resolved) {
        knot_dname_to_str(qname_str, qry->sname, sizeof(qname_str));
        printf("%s at %u\n", qname_str, qry->timestamp);
    }

    return ctx->state;
}
```

## 4.5 APIs in Lua

The APIs in Lua world try to mirror the C APIs using LuaJIT FFI, with several differences and enhancements. There is not comprehensive guide on the API yet, but you can have a look at the [bindings](#) file.

### 4.5.1 Elementary types and constants

- States are directly in `kres` table, e.g. `kres.YIELD`, `kres.CONSUME`, `kres.PRODUCE`, `kres.DONE`, `kres.FAIL`.
- DNS classes are in `kres.class` table, e.g. `kres.class.IN` for Internet class.
- DNS types are in `kres.type` table, e.g. `kres.type.AAAA` for AAAA type.
- DNS rcodes types are in `kres.rcode` table, e.g. `kres.rcode.NOERROR`.
- Packet sections (QUESTION, ANSWER, AUTHORITY, ADDITIONAL) are in the `kres.section` table.

### 4.5.2 Working with domain names

The internal API usually works with domain names in label format, you can convert between text and wire freely.

```
local dname = kres.str2dname('business.se')
local strname = kres.dname2str(dname)
```

### 4.5.3 Working with resource records

Resource records are stored as tables.

```
local rr = { owner = kres.str2dname('owner'),
             ttl = 0,
             class = kres.class.IN,
             type = kres.type.CNAME,
             rdata = kres.str2dname('someplace') }
print(kres.rr2str(rr))
```

RRSets in packet can be accessed using FFI, you can easily fetch single records.

```
local rrset = { ... }
local rr = rrset:get(0) -- Return first RR
print(kres.dname2str(rr:owner()))
print(rr:ttl())
print(kres.rr2str(rr))
```

### 4.5.4 Working with packets

Packet is the data structure that you're going to see in layers very often. They consists of a header, and four sections: QUESTION, ANSWER, AUTHORITY, ADDITIONAL. The first section is special, as it contains the query name, type, and class; the rest of the sections contain RRsets.

First you need to convert it to a type known to FFI and check basic properties. Let's start with a snippet of a *consume* layer.

```
consume = function (state, req, pkt)
    pkt = kres.pkt_t(answer)
    print('rcode:', pkt:rcode())
    print('query:', kres.dname2str(pkt:qname()), pkt:qclass(), pkt:qtype())
    if pkt:rcode() ~= kres.rcode.NOERROR then
        print('error response')
```

(continues on next page)

(continued from previous page)

```

    end
end

```

You can enumerate records in the sections.

```

local records = pkt:section(kres.section.ANSWER)
for i = 1, #records do
    local rr = records[i]
    if rr.type == kres.type.AAAA then
        print(kres.rr2str(rr))
    end
end
end

```

During *produce* or *begin*, you might want to write to packet. Keep in mind that you have to write packet sections in sequence, e.g. you can't write to ANSWER after writing AUTHORITY, it's like stages where you can't go back.

```

pkt:rcode(kres.rcode.NXDOMAIN)
-- Clear answer and write QUESTION
pkt:recycle()
pkt:question('\7blocked', kres.class.IN, kres.type.SOA)
-- Start writing data
pkt:begin(kres.section.ANSWER)
-- Nothing in answer
pkt:begin(kres.section.AUTHORITY)
local soa = { owner = '\7blocked', ttl = 900, class = kres.class.IN, type = kres.type.
↳SOA, rdata = '...' }
pkt:put(soa.owner, soa.ttl, soa.class, soa.type, soa.rdata)

```

#### 4.5.5 Working with requests

The request holds information about currently processed query, enabled options, cache, and other extra data. You primarily need to retrieve currently processed query.

```

consume = function (state, req, pkt)
    req = kres.request_t(req)
    print(req.options)
    print(req.state)

    -- Print information about current query
    local current = req:current()
    print(kres.dname2str(current.owner))
    print(current.stype, current.sclass, current.id, current.flags)
end

```

In layers that either begin or finalize, you can walk the list of resolved queries.

```

local last = req:resolved()
print(last.stype)

```

As described in the layers, you can not only retrieve information about current query, but also push new ones or pop old ones.

```

-- Push new query
local qry = req:push(pkt:qname(), kres.type.SOA, kres.class.IN)
qry.flags.AWAIT_CUT = true

-- Pop the query, this will erase it from resolution plan
req:pop(qry)

```

## 4.5.6 Significant Lua API changes

### Incompatible changes since 3.0.0

In the main `kres.*` lua binding, there was only change in struct `knot_rrset_t`:

- constructor now accepts TTL as additional parameter (defaulting to zero)
- `add_rdata()` doesn't accept TTL anymore (and will throw an error if passed)

In case you used `knot_*` functions and structures bound to lua:

- `knot_dname_is_sub(a, b)`: `knot_dname_in_bailiwick(a, b) > 0`
- `knot_rdata_rrlen()`: `knot_rdataset_at().len`
- `knot_rdata_data()`: `knot_rdataset_at().data`
- `knot_rdata_array_size()`: `offsetof(struct knot_data_t, data) + knot_rdataset_at().len`
- struct `knot_rdataset`: field names were renamed to `.count` and `.rdata`
- some functions got inlined from headers, but you can use their `kr_*` clones: `kr_rrsig_sig_inception()`, `kr_rrsig_sig_expiration()`, `kr_rrsig_type_covered()`. Note that these functions now accept `knot_rdata_t*` instead of a pair `knot_rdataset_t*` and `size_t` - you can use `knot_rdataset_at()` for that.
- `knot_rrset_add_rdata()` doesn't take TTL parameter anymore
- `knot_rrset_init_empty()` was inlined, but in lua you can use the constructor
- `knot_rrset_ttl()` was inlined, but in lua you can use `:ttl()` method instead
- `knot_pkt_qname()`, `_qtype()`, `_qclass()`, `_rr()`, `_section()` were inlined, but in lua you can use methods instead, e.g. `myPacket:qname()`
- `knot_pkt_free()` takes `knot_pkt_t*` instead of `knot_pkt_t**`, but from lua you probably didn't want to use that; constructor ensures garbage collection.

## 4.6 API reference

- *Name resolution*
- *Cache*
- *Nameservers*
- *Modules*
- *Utilities*
- *Generics library*

## 4.6.1 Name resolution

The API provides an API providing a “consumer-producer”-like interface to enable user to plug it into existing event loop or I/O code.

*Example usage of the iterative API:*

```
// Create request and its memory pool
struct kr_request req = {
    .pool = {
        .ctx = mp_new (4096),
        .alloc = (mm_alloc_t) mp_alloc
    }
};

// Setup and provide input query
int state = kr_resolve_begin(&req, ctx, final_answer);
state = kr_resolve_consume(&req, query);

// Generate answer
while (state == KR_STATE_PRODUCE) {

    // Additional query generate, do the I/O and pass back answer
    state = kr_resolve_produce(&req, &addr, &type, query);
    while (state == KR_STATE_CONSUME) {
        int ret = sendrecv(addr, proto, query, resp);

        // If I/O fails, make "resp" empty
        state = kr_resolve_consume(&request, addr, resp);
        knot_pkt_clear(resp);
    }
    knot_pkt_clear(query);
}

// "state" is either DONE or FAIL
kr_resolve_finish(&request, state);
```

### Defines

**kr\_request\_selected**(req)

Initializer for an array of \*\_selected.

### Enums

**enum kr\_rank**

RRset rank - for cache and ranked\_rr\_\*.

The rank meaning consists of one independent flag - KR\_RANK\_AUTH, and the rest have meaning of values where only one can hold at any time. You can use one of the enums as a safe initial value, optionally | KR\_RANK\_AUTH; otherwise it's best to manipulate ranks via the kr\_rank\_\* functions.

See also: <https://tools.ietf.org/html/rfc2181#section-5.4.1> <https://tools.ietf.org/html/rfc4035#section-4.3>

**Note** The representation is complicated by restrictions on integer comparison:

- AUTH must be > than !AUTH
- AUTH INSECURE must be > than AUTH (because it attempted validation)

- !AUTH SECURE must be > than AUTH (because it's valid)

Values:

**KR\_RANK\_INITIAL** = 0

Did not attempt to validate.

It's assumed compulsory to validate (or prove insecure).

**KR\_RANK\_OMIT**

Do not attempt to validate.

(And don't consider it a validation failure.)

**KR\_RANK\_TRY**

Attempt to validate, but failures are non-fatal.

**KR\_RANK\_INDET** = 4

Unable to determine whether it should be secure.

**KR\_RANK\_BOGUS**

Ought to be secure but isn't.

**KR\_RANK\_MISMATCH**

**KR\_RANK\_MISSING**

Unable to obtain a good signature.

**KR\_RANK\_INSECURE** = 8

Proven to be insecure, i.e.

we have a chain of trust from TAs that cryptographically denies the possibility of existence of a positive chain of trust from the TAs to the record.

**KR\_RANK\_AUTH** = 16

Authoritative data flag; the chain of authority was "verified".

Even if not set, only in-bailiwick stuff is acceptable, i.e. almost authoritative (example: mandatory glue and its NS RR).

**KR\_RANK\_SECURE** = 32

Verified whole chain of trust from the closest TA.

## Functions

bool **kr\_rank\_check** (uint8\_t rank)

Check that a rank value is valid.

Meant for assertions.

static bool **kr\_rank\_test** (uint8\_t rank, uint8\_t kr\_flag)

Test the presence of any flag/state in a rank, i.e.

including KR\_RANK\_AUTH.

static void **kr\_rank\_set** (uint8\_t \* rank, uint8\_t kr\_flag)

Set the rank state.

The \_AUTH flag is kept as it was.

*KR\_EXPORT* int **kr\_resolve\_begin** (struct *kr\_request* \* request, struct *kr\_context* \* ctx, knot\_pkt\_t \* answer)

Begin name resolution.

**Note** Expects a request to have an initialized mempool, the “answer” packet will be kept during the resolution and will contain the final answer at the end.

**Return** CONSUME (expecting query)

**Parameters**

- request: request state with initialized mempool
- ctx: resolution context
- answer: allocated packet for final answer

*KR\_EXPORT* int **kr\_resolve\_consume** (struct *kr\_request* \* request, const struct sockaddr \* src, knot\_pkt\_t \* packet)

Consume input packet (may be either first query or answer to query originated from *kr\_resolve\_produce()*)

**Note** If the I/O fails, provide an empty or NULL packet, this will make iterator recognize nameserver failure.

**Return** any state

**Parameters**

- request: request state (awaiting input)
- src: [in] packet source address
- packet: [in] input packet

*KR\_EXPORT* int **kr\_resolve\_produce** (struct *kr\_request* \* request, struct sockaddr \*\* dst, int \* type, knot\_pkt\_t \* packet)

Produce either next additional query or finish.

If the CONSUME is returned then dst, type and packet will be filled with appropriate values and caller is responsible to send them and receive answer. If it returns any other state, then content of the variables is undefined.

**Return** any state

**Parameters**

- request: request state (in PRODUCE state)
- dst: [out] possible address of the next nameserver
- type: [out] possible used socket type (SOCK\_STREAM, SOCK\_DGRAM)
- packet: [out] packet to be filled with additional query

*KR\_EXPORT* int **kr\_resolve\_checkout** (struct *kr\_request* \* request, struct sockaddr \* src, struct sock-addr \* dst, int type, knot\_pkt\_t \* packet)

Finalises the outbound query packet with the knowledge of the IP addresses.

**Note** The function must be called before actual sending of the request packet.

**Return** kr\_ok() or error code

**Parameters**

- request: request state (in PRODUCE state)
- src: address from which the query is going to be sent
- dst: address of the name server



- `type`: used socket type (SOCK\_STREAM, SOCK\_DGRAM)
- `packet`: [in,out] query packet to be finalised

*KR\_EXPORT* int **kr\_resolve\_finish** (struct *kr\_request* \* *request*, int *state*)

Finish resolution and commit results if the state is DONE.

**Note** The structures will be deinitialized, but the assigned memory pool is not going to be destroyed, as it's owned by caller.

**Return** DONE

#### Parameters

- `request`: request state
- `state`: either DONE or FAIL state

*KR\_EXPORT KR\_PURE* struct *kr\_rplan*\* **kr\_resolve\_plan** (struct *kr\_request* \* *request*)

Return resolution plan.

**Return** pointer to rplan

#### Parameters

- `request`: request state

*KR\_EXPORT KR\_PURE* knot\_mm\_t\* **kr\_resolve\_pool** (struct *kr\_request* \* *request*)

Return memory pool associated with request.

**Return** mempool

#### Parameters

- `request`: request state

**struct kr\_context**

*#include <resolve.h>* Name resolution context.

Resolution context provides basic services like cache, configuration and options.

**Note** This structure is persistent between name resolutions and may be shared between threads.

#### Public Members

struct *kr\_qflags* **options**

knot\_rrset\_t\* **opt\_rr**

*map\_t* **trust\_anchors**

*map\_t* **negative\_anchors**

struct *kr\_zonecut* **root\_hints**

struct *kr\_cache* **cache**

kr\_nsrep\_rtt\_lru\_t\* **cache\_rtt**

unsigned **cache\_rtt\_tout\_retry\_interval**

kr\_nsrep\_lru\_t\* **cache\_rep**

```
module_array_t* modules
struct kr_cookie_ctx cookie_ctx
kr_cookie_lru_t* cache_cookie
int32_t tls_padding
    See net.tls_padding in ../daemon/README.rst -1 is "true" (default policy), 0 is "false" (no padding)
knot_mm_t* pool
```

**struct kr\_request**

*#include <resolve.h>* Name resolution request.

Keeps information about current query processing between calls to processing APIs, i.e. current resolved query, resolution plan, ... Use this instead of the simple interface if you want to implement multiplexing or custom I/O.

**Note** All data for this request must be allocated from the given pool.

**Public Members**

```
struct kr_context* ctx
knot_pkt_t* answer
struct kr_query* current_query
    Current evaluated query.
const knot_rrset_t* key
const struct sockaddr* addr
    Address that originated the request.
    Current upstream address.
    NULL for internal origin.
const struct sockaddr* dst_addr
    Address that accepted the request.
    NULL for internal origin.
const knot_pkt_t* packet
const knot_rrset_t* opt
bool tcp
    true if the request is on tcp; only meaningful if (dst_addr)
size_t size
    query packet size
struct kr_request::@6 qsource
unsigned rtt
    Current upstream RTT.
struct kr_request::@7 upstream
    Upstream information, valid only in consume() phase.
struct kr_qflags options
int state
```

ranked\_rr\_array\_t **answ\_selected**

ranked\_rr\_array\_t **auth\_selected**

ranked\_rr\_array\_t **add\_selected**

rr\_array\_t **additional**

bool **answ\_validated**

internal to validator; beware of caching, etc.

bool **auth\_validated**

see `answ_validated` ^^ ; TODO

uint8\_t **rank**

Overall rank for the request.

Values from `kr_rank`, currently just `KR_RANK_SECURE` and `_INITIAL`. Only read this in finish phase and after validator, please. Meaning of `_SECURE`: all RRs in answer+authority are `_SECURE`, including any negative results implied (NXDOMAIN, NODATA).

struct *kr\_rplan* **rplan**

int **has\_tls**

*trace\_log\_f* **trace\_log**

Logging tracepoint.

*trace\_callback\_f* **trace\_finish**

Request finish tracepoint.

int **vars\_ref**

Reference to per-request variable table.

LUA\_NOREF if not set.

knot\_mm\_t **pool**

unsigned int **uid**

## Typedefs

**typedef** int32\_t (\* **kr\_stale\_cb**) (int32\_t *t*, const knot\_dname\_t \**owner*, uint16\_t *type*, const struct *kr\_query* \**q*)

Callback for serve-stale decisions.

**Return** the adjusted TTL (typically 1) or < 0.

### Parameters

- *t*: the expired TTL (i.e. it's < 0)

## Functions

*KR\_EXPORT* void **kr\_qflags\_set** (struct *kr\_qflags* \**f1*, struct *kr\_qflags* *f2*)

Combine flags together.

This means set union for simple flags.

*KR\_EXPORT* void **kr\_qflags\_clear** (struct *kr\_qflags* \**f1*, struct *kr\_qflags* *f2*)

Remove flags.

This means set-theoretic difference.

*KR\_EXPORT* int **kr\_rplan\_init** (struct *kr\_rplan* \* *rplan*, struct *kr\_request* \* *request*, knot\_mm\_t \* *pool*)  
Initialize resolution plan (empty).

**Parameters**

- *rplan*: plan instance
- *request*: resolution request
- *pool*: ephemeral memory pool for whole resolution

*KR\_EXPORT* void **kr\_rplan\_deinit** (struct *kr\_rplan* \* *rplan*)  
Deinitialize resolution plan, aborting any uncommitted transactions.

**Parameters**

- *rplan*: plan instance

*KR\_EXPORT KR\_PURE* bool **kr\_rplan\_empty** (struct *kr\_rplan* \* *rplan*)  
Return true if the resolution plan is empty (i.e. finished or initialized)

**Return** true or false

**Parameters**

- *rplan*: plan instance

*KR\_EXPORT* struct *kr\_query*\* **kr\_rplan\_push\_empty** (struct *kr\_rplan* \* *rplan*, struct *kr\_query* \* *parent*)  
Push empty query to the top of the resolution plan.

**Note** This query serves as a cookie query only.

**Return** query instance or NULL

**Parameters**

- *rplan*: plan instance
- *parent*: query parent (or NULL)

*KR\_EXPORT* struct *kr\_query*\* **kr\_rplan\_push** (struct *kr\_rplan* \* *rplan*, struct *kr\_query* \* *parent*, const knot\_dname\_t \* *name*, uint16\_t *cls*, uint16\_t *type*)  
Push a query to the top of the resolution plan.

**Note** This means that this query takes precedence before all pending queries.

**Return** query instance or NULL

**Parameters**

- *rplan*: plan instance
- *parent*: query parent (or NULL)
- *name*: resolved name
- *cls*: resolved class
- *type*: resolved type

*KR\_EXPORT* int **kr\_rplan\_pop** (struct *kr\_rplan* \* *rplan*, struct *kr\_query* \* *qry*)  
Pop existing query from the resolution plan.

**Note** Popped queries are not discarded, but moved to the resolved list.

**Return** 0 or an error

**Parameters**

- `rplan`: plan instance
- `qry`: resolved query

*KR\_EXPORT KR\_PURE* bool **kr\_rplan\_satisfies** (struct *kr\_query* \* *closure*, const knot\_dname\_t \* *name*, uint16\_t *cls*, uint16\_t *type*)

Return true if resolution chain satisfies given query.

*KR\_EXPORT KR\_PURE* struct *kr\_query*\* **kr\_rplan\_resolved** (struct *kr\_rplan* \* *rplan*)

Return last resolved query.

*KR\_EXPORT KR\_PURE* struct *kr\_query*\* **kr\_rplan\_last** (struct *kr\_rplan* \* *rplan*)

Return last query (either currently being solved or last resolved).

This is necessary to retrieve the last query in case of resolution failures (e.g. time limit reached).

*KR\_EXPORT KR\_PURE* struct *kr\_query*\* **kr\_rplan\_find\_resolved** (struct *kr\_rplan* \* *rplan*, struct *kr\_query* \* *parent*, const knot\_dname\_t \* *name*, uint16\_t *cls*, uint16\_t *type*)

Check if a given query already resolved.

**Return** query instance or NULL

**Parameters**

- `rplan`: plan instance
- `parent`: query parent (or NULL)
- `name`: resolved name
- `cls`: resolved class
- `type`: resolved type

**struct kr\_qflags**

*#include <rplan.h>* Query flags.

**Public Members**

bool **NO\_MINIMIZE**  
Don't minimize QNAME.

bool **NO\_THROTTLE**  
No query/slow NS throttling.

bool **NO\_IPV6**  
Disable IPv6.

bool **NO\_IPV4**  
Disable IPv4.

bool **TCP**  
Use TCP for this query.

- bool **RESOLVED**  
Query is resolved.  
Note that *kr\_query* gets RESOLVED before following a CNAME chain; see .CNAME.
- bool **AWAIT\_IPV4**  
Query is waiting for A address.
- bool **AWAIT\_IPV6**  
Query is waiting for AAAA address.
- bool **AWAIT\_CUT**  
Query is waiting for zone cut lookup.
- bool **SAFEMODE**  
Don't use fancy stuff (EDNS, 0x20, ...)
- bool **CACHED**  
Query response is cached.
- bool **NO\_CACHE**  
No cache for lookup; exception: finding NSs and subqueries.
- bool **EXPIRING**  
Query response is cached, but expiring.
- bool **ALLOW\_LOCAL**  
Allow queries to local or private address ranges.
- bool **DNSSEC\_WANT**  
Want DNSSEC secured answer; exception: +cd, i.e.  
`knot_wire_set_cd(request->answer->wire)`.
- bool **DNSSEC\_BOGUS**  
Query response is DNSSEC bogus.
- bool **DNSSEC\_INSECURE**  
Query response is DNSSEC insecure.
- bool **DNSSEC\_CD**  
Instruction to set CD bit in request.
- bool **STUB**  
Stub resolution, accept received answer as solved.
- bool **ALWAYS\_CUT**  
Always recover zone cut (even if cached).
- bool **DNSSEC\_WEXPAND**  
Query response has wildcard expansion.
- bool **PERMISSIVE**  
Permissive resolver mode.
- bool **STRICT**  
Strict resolver mode.
- bool **BADCOOKIE\_AGAIN**  
Query again because bad cookie returned.
- bool **CNAME**  
Query response contains CNAME in answer section.

bool **REORDER\_RR**  
Reorder cached RRs.

bool **TRACE**  
Also log answers if verbose.

bool **NO\_0X20**  
Disable query case randomization .

bool **DNSSEC\_NODS**  
DS non-existence is proven.

bool **DNSSEC\_OPTOUT**  
Closest encloser proof has optout.

bool **NONAUTH**  
Non-authoritative in-bailiwick records are enough.  
TODO: utilize this also outside cache.

bool **FORWARD**  
Forward all queries to upstream; validate answers.

bool **DNS64\_MARK**  
Internal mark for dns64 module.

bool **CACHE\_TRIED**  
Internal to cache module.

bool **NO\_NS\_FOUND**  
No valid NS found during last PRODUCE stage.

**struct kr\_query**  
*#include <rplan.h>* Single query representation.

### Public Members

struct *kr\_query*\* **parent**

knot\_dname\_t\* **sname**  
The name to resolve - lower-cased, uncompressed.

uint16\_t **stype**

uint16\_t **sclass**

uint16\_t **id**

struct *kr\_qflags* flags **forward\_flags**

uint32\_t **secret**

uint16\_t **fails**

uint16\_t **reorder**  
Seed to reorder (cached) RRs in answer or zero.

uint64\_t **creation\_time\_mono**

uint64\_t **timestamp\_mono**  
Time of query created or time of query to upstream resolver (milliseconds).

struct timeval **timestamp**  
Real time for TTL+DNSSEC checks (.tv\_sec only).

struct *kr\_zonecut* **zone\_cut**

struct *kr\_layer\_pickle*\* **deferred**

uint32\_t **uid**

Query iteration number, unique within the *kr\_rplan*.

struct *kr\_query*\* **cname\_parent**

Pointer to the query that originated this one because of following a CNAME (or NULL).

struct *kr\_request*\* **request**

Parent resolution request.

*kr\_stale\_cb* **stale\_cb**

See the type.

struct *kr\_nsrep* **ns**

## **struct kr\_rplan**

*#include <rplan.h>* Query resolution plan structure.

The structure most importantly holds the original query, answer and the list of pending queries required to resolve the original query. It also keeps a notion of current zone cut.

## **Public Members**

*kr\_qarray\_t* **pending**

List of pending queries.

*kr\_qarray\_t* **resolved**

List of resolved queries.

struct *kr\_request*\* **request**

Parent resolution request.

*knot\_mm\_t*\* **pool**

Temporary memory pool.

uint32\_t **next\_uid**

Next value for *kr\_query::uid* (incremental).

## **4.6.2 Cache**

### **Functions**

int **cache\_peek** (*kr\_layer\_t* \* *ctx*, *knot\_pkt\_t* \* *pkt*)

int **cache\_stash** (*kr\_layer\_t* \* *ctx*, *knot\_pkt\_t* \* *pkt*)

*KR\_EXPORT* int **kr\_cache\_open** (*struct kr\_cache* \* *cache*, *const struct kr\_cdb\_api* \* *api*, *struct kr\_cdb\_opts* \* *opts*, *knot\_mm\_t* \* *mm*)

Open/create cache with provided storage options.

**Return** 0 or an error code

### **Parameters**

- *cache*: cache structure to be initialized
- *api*: storage engine API



- `opts`: storage-specific options (may be NULL for default)
- `mm`: memory context.

*KR\_EXPORT* void **kr\_cache\_close** (struct *kr\_cache* \* *cache*)  
Close persistent cache.

**Note** This doesn't clear the data, just closes the connection to the database.

#### Parameters

- `cache`: structure

*KR\_EXPORT* int **kr\_cache\_sync** (struct *kr\_cache* \* *cache*)  
Run after a row of operations to release transaction/lock if needed.

static bool **kr\_cache\_is\_open** (struct *kr\_cache* \* *cache*)  
Return true if cache is open and enabled.

static void **kr\_cache\_make\_checkpoint** (struct *kr\_cache* \* *cache*)  
(Re)set the time pair to the current values.

*KR\_EXPORT* int **kr\_cache\_insert\_rr** (struct *kr\_cache* \* *cache*, const knot\_rrset\_t \* *rr*, const knot\_rrset\_t \* *rrsig*, uint8\_t *rank*, uint32\_t *timestamp*)  
Insert RRSet into cache, replacing any existing data.

**Return** 0 or an errcode

#### Parameters

- `cache`: cache structure
- `rr`: inserted RRSet
- `rrsig`: RRSIG for inserted RRSet (optional)
- `rank`: rank of the data
- `timestamp`: current time

*KR\_EXPORT* int **kr\_cache\_clear** (struct *kr\_cache* \* *cache*)  
Clear all items from the cache.

**Return** 0 or an errcode

#### Parameters

- `cache`: cache structure

*KR\_EXPORT* int **kr\_cache\_peek\_exact** (struct *kr\_cache* \* *cache*, const knot\_dname\_t \* *name*, uint16\_t *type*, struct *kr\_cache\_p* \* *peek*)

*KR\_EXPORT* int32\_t **kr\_cache\_ttl** (const struct *kr\_cache\_p* \* *peek*, const struct *kr\_query* \* *qry*, const knot\_dname\_t \* *name*, uint16\_t *type*)

*KR\_EXPORT* int **kr\_cache\_materialize** (knot\_rdataset\_t \* *dst*, const struct *kr\_cache\_p* \* *ref*, knot\_mm\_t \* *pool*)

*KR\_EXPORT* int **kr\_cache\_remove** (struct *kr\_cache* \* *cache*, const knot\_dname\_t \* *name*, uint16\_t *type*)  
Remove an entry from cache.

**Return** number of deleted records, or negative error code

**Note** only "exact hits" are considered ATM, and some other information may be removed alongside.

**Parameters**

- `cache`: cache structure
- `name`: `dname`
- `type`: rr type

*KR\_EXPORT* int **kr\_cache\_match** (struct *kr\_cache* \* *cache*, const knot\_dname\_t \* *name*, bool *exact\_name*,  
knot\_db\_val\_t *keyval*[][2], int *maxcount*)

Get keys matching a `dname` if prefix.

**Return** result count or an errcode

**Note** the cache keys are matched by prefix, i.e. it very much depends on their structure; `CACHE_KEY_DEF`.

**Parameters**

- `cache`: cache structure
- `name`: `dname`
- `exact_name`: whether to only consider exact name matches
- `keyval`: matched key-value pairs
- `maxcount`: limit on the number of returned key-value pairs

*KR\_EXPORT* int **kr\_cache\_remove\_subtree** (struct *kr\_cache* \* *cache*, const knot\_dname\_t \* *name*,  
bool *exact\_name*, int *maxcount*)

Remove a subtree in cache.

It's like `_match` but removing them instead of returning.

**Return** number of deleted entries or an errcode

*KR\_EXPORT* int **kr\_cache\_closest\_apex** (struct *kr\_cache* \* *cache*, const knot\_dname\_t \* *name*,  
bool *is\_DS*, knot\_dname\_t \*\* *apex*)

Find the closest cached zone apex for a name (in cache).

**Return** the number of labels to remove from the name, or negative error code

**Note** timestamp is found by a syscall, and stale-serving is not considered

**Parameters**

- `is_DS`: start searching one name higher

*KR\_EXPORT* int **kr\_unpack\_cache\_key** (knot\_db\_val\_t *key*, knot\_dname\_t \* *buf*, uint16\_t \* *type*)

Unpack `dname` and `type` from db key.

**Return** length of `dname` or an errcode

**Note** only “exact hits” are considered ATM, moreover `xNAME` records are “hidden” as NS. (see comments in `struct entry_h`)

**Parameters**

- `key`: db key representation
- `buf`: output buffer of domain name in `dname` format
- `type`: output for `type`

## Variables

const size\_t **PKT\_SIZE\_NOWIRE** = -1

When knot\_pkt is passed from cache without ->wire, this is the ->size.

*KR\_EXPORT* const char\* **kr\_cache\_emergency\_file\_to\_remove**

Path to cache file to remove on critical out-of-space error.

(do NOT modify it)

**struct kr\_cache**

*#include <api.h>* Cache structure, keeps API, instance and metadata.

### Public Members

knot\_db\_t\* **db**

Storage instance.

const struct kr\_cdb\_api\* **api**

Storage engine.

uint32\_t **hit**

Number of cache hits.

uint32\_t **miss**

Number of cache misses.

uint32\_t **insert**

Number of insertions.

uint32\_t **delete**

Number of deletions.

struct *kr\_cache::*@0 **stats**

uint32\_t **t1\_min**

uint32\_t **t1\_max**

TTL limits.

struct timeval **checkpoint\_walltime**

Wall time on the last check-point.

uint64\_t **checkpoint\_monotime**

Monotonic milliseconds on the last check-point.

**struct kr\_cache\_p**

### Public Members

uint32\_t **time**

The time of inception.

uint32\_t **t1**

TTL at inception moment.

Assuming it fits into int32\_t ATM.

uint8\_t **rank**

See enum kr\_rank.

```
void* raw_data
void * raw_bound
struct kr_cache_p::@1 kr_cache_p::@2
```

## 4.6.3 Nameservers

### Defines

#### **KR\_NS\_DEAD**

See `kr_nsrep_update_rtt()`

#### **KR\_NS\_TIMEOUT\_RETRY\_INTERVAL**

If once NS was marked as “timeouted”, it won’t participate in NS elections at least `KR_NS_TIMEOUT_RETRY_INTERVAL` milliseconds (now: one minute).

#### **KR\_NSREP\_MAXADDR**

### Typedefs

```
typedef struct kr_nsrep_rtt_lru_entry kr_nsrep_rtt_lru_entry_t
```

### Enums

#### **enum kr\_ns\_score**

NS RTT score (special values).

**Note** RTT is measured in milliseconds.

*Values:*

**KR\_NS\_MAX\_SCORE** = `KR_CONN_RTT_MAX`

**KR\_NS\_TIMEOUT** = `(95 * KR_NS_MAX_SCORE) / 100`

**KR\_NS\_LONG** = `(3 * KR_NS_TIMEOUT) / 4`

**KR\_NS\_UNKNOWN** = `KR_NS_TIMEOUT / 2`

**KR\_NS\_PENALTY** = `100`

**KR\_NS\_GLUED** = `10`

#### **enum kr\_ns\_rep**

NS QoS flags.

*Values:*

**KR\_NS\_NOIP4** = `1 << 0`

NS has no IPv4.

**KR\_NS\_NOIP6** = `1 << 1`

NS has no IPv6.

**KR\_NS\_NOEDNS** = `1 << 2`

NS has no EDNS support.

**enum kr\_ns\_update\_mode**

NS RTT update modes.

First update is always KR\_NS\_RESET unless KR\_NS\_UPDATE\_NORESET mode had choosen.

*Values:*

**KR\_NS\_UPDATE = 0**

Update as smooth over last two measurements.

**KR\_NS\_UPDATE\_NORESET**

Same as KR\_NS\_UPDATE, but disable fallback to KR\_NS\_RESET on newly added entries.

Zero is used as initial value.

**KR\_NS\_RESET**

Set to given value.

**KR\_NS\_ADD**

Increment current value.

**KR\_NS\_MAX**

Set to maximum of current/proposed value.

**Functions**

typedef **lru\_t** (*kr\_nsrep\_rtt\_lru\_entry\_t*)

NS QoS tracking.

typedef **lru\_t** (unsigned)

NS reputation tracking.

**KR\_EXPORT** int **kr\_nsrep\_set** (struct *kr\_query* \* *qry*, size\_t *index*, const struct sockaddr \* *sock*)

Set given NS address.

**Return** 0 or an error code

**Parameters**

- *qry*: updated query
- *index*: index of the updated target
- *sock*: socket address to use (sockaddr\_in or sockaddr\_in6 or NULL)

**KR\_EXPORT** int **kr\_nsrep\_select** (struct *kr\_query* \* *qry*, struct *kr\_context* \* *ctx*)

Elect best nameserver/address pair from the nsset.

**Return** 0 or an error code

**Parameters**

- *qry*: updated query
- *ctx*: resolution context

**KR\_EXPORT** int **kr\_nsrep\_select\_addr** (struct *kr\_query* \* *qry*, struct *kr\_context* \* *ctx*)

Elect best nameserver/address pair from the nsset.

**Return** 0 or an error code

**Parameters**

- `qry`: updated query
- `ctx`: resolution context

*KR\_EXPORT* int **kr\_nsrep\_update\_rtt** (struct *kr\_nsrep* \* *ns*, const struct sockaddr \* *addr*, unsigned *score*, kr\_nsrep\_rtt\_lru\_t \* *cache*, int *umode*)

Update NS address RTT information.

In KR\_NS\_UPDATE mode reputation is smoothed over last N measurements.

**Return** 0 on success, error code on failure

**Parameters**

- `ns`: updated NS representation
- `addr`: chosen address (NULL for first)
- `score`: new score (i.e. RTT), see enum `kr_ns_score` after two calls with `score = KR_NS_DEAD` and `umode = KR_NS_UPDATE` server will be guaranteed to have `score >= KR_NS_TIMEOUT`
- `cache`: RTT LRU cache
- `umode`: update mode (KR\_NS\_UPDATE or KR\_NS\_RESET or KR\_NS\_ADD)

*KR\_EXPORT* int **kr\_nsrep\_update\_rep** (struct *kr\_nsrep* \* *ns*, unsigned *reputation*, kr\_nsrep\_lru\_t \* *cache*)

Update NSSET reputation information.

**Return** 0 on success, error code on failure

**Parameters**

- `ns`: updated NS representation
- `reputation`: combined reputation flags, see enum `kr_ns_rep`
- `cache`: LRU cache

int **kr\_nsrep\_copy\_set** (struct *kr\_nsrep* \* *dst*, const struct *kr\_nsrep* \* *src*)  
Copy NSSET reputation information and resets score.

**Return** 0 on success, error code on failure

**Parameters**

- `dst`: updated NS representation
- `src`: source NS representation

*KR\_EXPORT* int **kr\_nsrep\_sort** (struct *kr\_nsrep* \* *ns*, kr\_nsrep\_rtt\_lru\_t \* *rtt\_cache*)  
Sort addresses in the query nsrep list.

**Return** 0 or an error code

**Note** ns reputation is zeroed, as KR\_NS\_NOIP{4,6} flags are useless in STUB/FORWARD mode.

**Parameters**

- `ns`: updated *kr\_nsrep*
- `rtt_cache`: RTT LRU cache

**struct kr\_nsrep\_rtt\_lru\_entry**

## Public Members

unsigned **score**

uint64\_t **tout\_timestamp**

**struct kr\_nsrep**

*#include <nsrep.h>* Name server representation.

Contains extra information about the name server, e.g. score or other metadata.

## Public Members

unsigned **score**

NS score.

unsigned **reputation**

NS reputation.

const knot\_dname\_t\* **name**

NS name.

struct *kr\_context*\* **ctx**

Resolution context.

**union inaddr kr\_nsrep::addr[KR\_NSREP\_MAXADDR]**

NS address(es)

## Functions

*KR\_EXPORT* int **kr\_zonecut\_init** (struct *kr\_zonecut* \* *cut*, const knot\_dname\_t \* *name*, knot\_mm\_t \* *pool*)

Populate root zone cut with SBELT.

**Return** 0 or error code

### Parameters

- *cut*: zone cut
- *name*:
- *pool*:

*KR\_EXPORT* void **kr\_zonecut\_deinit** (struct *kr\_zonecut* \* *cut*)

Clear the structure and free the address set.

### Parameters

- *cut*: zone cut

*KR\_EXPORT* void **kr\_zonecut\_set** (struct *kr\_zonecut* \* *cut*, const knot\_dname\_t \* *name*)

Reset zone cut to given name and clear address list.

**Note** This clears the address list even if the name doesn't change. TA and DNSKEY don't change.

### Parameters

- *cut*: zone cut to be set

- name: new zone cut name

*KR\_EXPORT* int **kr\_zonecut\_copy** (struct *kr\_zonecut* \* *dst*, const struct *kr\_zonecut* \* *src*)

Copy zone cut, including all data.

Does not copy keys and trust anchor.

**Return** 0 or an error code; If it fails with `kr_error(ENOMEM)`, it may be in a half-filled state, but it's safe to `deinit...`

**Note** addresses for names in `src` get replaced and others are left as they were.

**Parameters**

- *dst*: destination zone cut
- *src*: source zone cut

*KR\_EXPORT* int **kr\_zonecut\_copy\_trust** (struct *kr\_zonecut* \* *dst*, const struct *kr\_zonecut* \* *src*)

Copy zone trust anchor and keys.

**Return** 0 or an error code

**Parameters**

- *dst*: destination zone cut
- *src*: source zone cut

*KR\_EXPORT* int **kr\_zonecut\_add** (struct *kr\_zonecut* \* *cut*, const knot\_dname\_t \* *ns*, const knot\_rdata\_t \* *rdata*)

Add address record to the zone cut.

The record will be merged with existing data, it may be either A/AAAA type.

**Return** 0 or error code

**Parameters**

- *cut*: zone cut to be populated
- *ns*: nameserver name
- *rdata*: nameserver address (as rdata)

*KR\_EXPORT* int **kr\_zonecut\_del** (struct *kr\_zonecut* \* *cut*, const knot\_dname\_t \* *ns*, const knot\_rdata\_t \* *rdata*)

Delete nameserver/address pair from the zone cut.

**Return** 0 or error code

**Parameters**

- *cut*:
- *ns*: name server name
- *rdata*: name server address

*KR\_EXPORT* int **kr\_zonecut\_del\_all** (struct *kr\_zonecut* \* *cut*, const knot\_dname\_t \* *ns*)

Delete all addresses associated with the given name.

**Return** 0 or error code



**Parameters**

- `cut`:
- `ns`: name server name

*KR\_EXPORT KR\_PURE* `pack_t* kr_zonecut_find` (struct *kr\_zonecut* \* *cut*, const knot\_dname\_t \* *ns*)  
Find nameserver address list in the zone `cut`.

**Note** This can be used for membership test, a non-null pack is returned if the nameserver name exists.

**Return** pack of addresses or NULL

**Parameters**

- `cut`:
- `ns`: name server name

*KR\_EXPORT* `int kr_zonecut_set_sbelt` (struct *kr\_context* \* *ctx*, struct *kr\_zonecut* \* *cut*)  
Populate zone cut with a root zone using SBELT :rfc:1034

**Return** 0 or error code

**Parameters**

- `ctx`: resolution context (to fetch root hints)
- `cut`: zone cut to be populated

*KR\_EXPORT* `int kr_zonecut_find_cached` (struct *kr\_context* \* *ctx*, struct *kr\_zonecut* \* *cut*, const knot\_dname\_t \* *name*, const struct *kr\_query* \* *qry*, bool \*restrict *secured*)

Populate zone cut address set from cache.

**Return** 0 or error code (ENOENT if it doesn't find anything)

**Parameters**

- `ctx`: resolution context (to fetch data from LRU caches)
- `cut`: zone cut to be populated
- `name`: QNAME to start finding zone cut for
- `qry`: query for timestamp and stale-serving decisions
- `secured`: set to true if want secured zone cut, will return false if it is provably insecure

*KR\_EXPORT* `bool kr_zonecut_is_empty` (struct *kr\_zonecut* \* *cut*)  
Check if any address is present in the zone cut.

**Return** true/false

**Parameters**

- `cut`: zone cut to check

**struct `kr_zonecut`**  
*#include* <zonecut.h> Current zone cut representation.

## Public Members

`knot_dname_t*` **name**  
Zone cut name.

`knot_rrset_t*` **key**  
Zone cut DNSKEY.

`knot_rrset_t*` **trust\_anchor**  
Current trust anchor.

struct `kr_zonecut*` **parent**  
Parent zone cut.

`trie_t*` **nsset**  
Map of nameserver => address\_set (pack\_t).

`knot_mm_t*` **pool**  
Memory pool.

## 4.6.4 Modules

Module API definition and functions for (un)loading modules.

### Defines

**KR\_MODULE\_EXPORT** (module)  
Export module API version (place this at the end of your module).

#### Parameters

- `module`: module name (f.e. hints)

### **KR\_MODULE\_API**

### Typedefs

**typedef** `uint32_t` ( `module_api_cb` ) (void)

**typedef** `char*` ( `kr_prop_cb` ) (void *\*env*, struct `kr_module` *\*self*, const char *\*input*)  
Module property callback.

Input and output is passed via a JSON encoded in a string.

**Return** a free-form JSON output (malloc-ated)

#### Parameters

- `env`: pointer to the lua engine, i.e. struct engine *\*env* (TODO: explicit type)
- `input`: parameter (NULL if missing/nil on lua level)

## Functions

*KR\_EXPORT* int **kr\_module\_load** (struct *kr\_module* \* *module*, const char \* *name*, const char \* *path*)  
Load a C module instance into memory.

**Return** 0 or an error

### Parameters

- *module*: module structure
- *name*: module name
- *path*: module search path

*KR\_EXPORT* void **kr\_module\_unload** (struct *kr\_module* \* *module*)  
Unload module instance.

### Parameters

- *module*: module structure

*KR\_EXPORT* const struct *kr\_module*\* **kr\_module\_embedded** (const char \* *name*)  
Get embedded module prototype by name (or NULL).

### struct **kr\_module**

*#include <module.h>* Module representation.

The five symbols (*init*, ...) may be defined by the module as *name\_init*(), etc; all are optional and missing symbols are represented as NULLs;

## Public Members

char\* **name**

int **init**() (struct *kr\_module* \**self*)

Constructor.

Called after loading the module.

**Return** error code.

int **deinit**() (struct *kr\_module* \**self*)

Destructor.

Called before unloading the module.

**Return** error code.

int **config**() (struct *kr\_module* \**self*, const char \**input*)

Configure with encoded JSON (NULL if missing).

**Return** error code.

const *kr\_layer\_api\_t*\* **layer**() (struct *kr\_module* \**self*)

Get a pointer to packet processing API specs.

See docs on that type.

const struct *kr\_prop*\* **props**() (void)

Get a pointer to list of properties, terminated by { NULL, NULL, NULL }.

**void\* lib**  
Shared library handle or RTLD\_DEFAULT.

**void\* data**  
Custom data context.

**struct kr\_prop**  
*#include <module.h>* Module property (named callable).

### Public Members

**kr\_prop\_cb\* cb**  
**const char\* name**  
**const char\* info**

## 4.6.5 Utilities

### Defines

**kr\_log\_info** (*fmt*, ...)

**kr\_log\_error** (*fmt*, ...)

**kr\_log\_trace\_enabled** (query)  
Return true if the query has request log handler installed.

**VERBOSE\_STATUS**  
Block run in verbose mode; optimized when not run.

**WITH\_VERBOSE** (query)

**kr\_log\_verbose**

**KR\_DNAME\_GET\_STR** (dname\_str, dname)

**KR\_RRTYPE\_GET\_STR** (rrtype\_str, rrtype)

**static\_assert** (cond, msg)

**RDATA\_ARR\_MAX**

**KR\_RRKEY\_LEN**

**SWAP** (x, y)  
Swap two places.

Note: the parameters need to be without side effects.

### Typedefs

**typedef void (\* trace\_callback\_f)** (struct *kr\_request* \*request)  
Callback for request events.

**typedef void (\* trace\_log\_f)** (const struct *kr\_query* \*query, const char \*source, const char \*msg)  
Callback for request logging handler.

## Functions

*KR\_EXPORT* bool **kr\_verbose\_set** (bool *status*)

Set verbose mode.

Not available if compiled with -DNOVERBOSELOG.

*KR\_EXPORT* **KR\_PRINTF** (1)

Log a message if in verbose mode.

*KR\_EXPORT* **KR\_PRINTF** (3)

Log a message through the request log handler.

**Return** true if the message was logged

### Parameters

- *query*: current query
- *source*: message source
- *fmt*: message format

static long **time\_diff** (struct timeval \* *begin*, struct timeval \* *end*)

Return time difference in milliseconds.

**Note** based on the `_BSD_SOURCE` `timersub()` macro

*KR\_EXPORT* char\* **kr\_strcatdup** (unsigned *n*, ...)

Concatenate N strings.

int **kr\_rand\_reseed** (void)

Reseed CSPRNG context.

*KR\_EXPORT* uint32\_t **kr\_rand\_uint** (uint32\_t *max*)

Get pseudo-random value between zero and *max*-1 (inclusive).

Passing zero means that any `uint32_t` should be returned (it's also faster).

*KR\_EXPORT* int **kr\_memreserve** (void \* *baton*, char \*\* *mem*, size\_t *elm\_size*, size\_t *want*, size\_t \* *have*)

Memory reservation routine for `knot_mm_t`.

*KR\_EXPORT* int **kr\_pkt\_recycle** (knot\_pkt\_t \* *pkt*)

*KR\_EXPORT* int **kr\_pkt\_clear\_payload** (knot\_pkt\_t \* *pkt*)

*KR\_EXPORT* int **kr\_pkt\_put** (knot\_pkt\_t \* *pkt*, const knot\_dname\_t \* *name*, uint32\_t *ttl*, uint16\_t *rclass*,  
uint16\_t *rtype*, const uint8\_t \* *rdata*, uint16\_t *rdlen*)

Construct and put record to packet.

*KR\_EXPORT* void **kr\_pkt\_make\_auth\_header** (knot\_pkt\_t \* *pkt*)

Set packet header suitable for authoritative answer.

(for policy module)

*KR\_EXPORT* *KR\_PURE* const char\* **kr\_inaddr** (const struct sockaddr \* *addr*)

Address bytes for given family.

*KR\_EXPORT* *KR\_PURE* int **kr\_inaddr\_family** (const struct sockaddr \* *addr*)

Address family.

*KR\_EXPORT KR\_PURE* int **kr\_inaddr\_len** (const struct sockaddr \* *addr*)

Address length for given family, i.e.

sizeof(struct in\*\_addr).

*KR\_EXPORT KR\_PURE* int **kr\_sockaddr\_len** (const struct sockaddr \* *addr*)

Sockaddr length for given family, i.e.

sizeof(struct sockaddr\_in\*).

*KR\_EXPORT KR\_PURE* int **kr\_sockaddr\_cmp** (const struct sockaddr \* *left*, const struct sockaddr \* *right*)

Compare two given sockaddr.

return 0 - addresses are equal, error code otherwise.

*KR\_EXPORT KR\_PURE* uint16\_t **kr\_inaddr\_port** (const struct sockaddr \* *addr*)

Port.

*KR\_EXPORT* void **kr\_inaddr\_set\_port** (struct sockaddr \* *addr*, uint16\_t *port*)

Set port.

*KR\_EXPORT* int **kr\_inaddr\_str** (const struct sockaddr \* *addr*, char \* *buf*, size\_t \* *buflen*)

String representation for given address as “<addr>#<port>”.

*KR\_EXPORT KR\_PURE* int **kr\_straddr\_family** (const char \* *addr*)

Return address type for string.

*KR\_EXPORT KR\_CONST* int **kr\_family\_len** (int *family*)

Return address length in given family (struct in\*\_addr).

*KR\_EXPORT* struct sockaddr\* **kr\_straddr\_socket** (const char \* *addr*, int *port*)

Create a sockaddr\* from string+port representation (also accepts IPv6 link-local).

*KR\_EXPORT* int **kr\_straddr\_subnet** (void \* *dst*, const char \* *addr*)

Parse address and return subnet length (bits).

**Warning** ‘dst’ must be at least sizeof(struct in6\_addr) long.

*KR\_EXPORT* int **kr\_straddr\_split** (const char \* *addr*, char \* *buf*, size\_t *buflen*, uint16\_t \* *port*)

Splits ip address specified as “addr@port” or “addr#port” into addr and port and performs validation.

**Note** if #port part isn’t present, then port will be set to 0. buf and port can be set to NULL.

**Return** kr\_error(EINVAL) - addr part doesn’t contains valid ip address or #port part is out-of-range (either < 0 either > UINT16\_MAX) kr\_error(ENOSP) - buflen is too small

*KR\_EXPORT* int **kr\_straddr\_join** (const char \* *addr*, uint16\_t *port*, char \* *buf*, size\_t \* *buflen*)

Formats ip address and port in “addr#port” format.

and performs validation.

**Note** Port always formatted as five-character string with leading zeros.

**Return** kr\_error(EINVAL) - addr or buf is NULL or buflen is 0 or addr doesn’t contain a valid ip address  
kr\_error(ENOSP) - buflen is too small

*KR\_EXPORT KR\_PURE* int **kr\_bitcmp** (const char \* *a*, const char \* *b*, int *bits*)

Compare memory bitwise.

The semantics is “the same” as for memcmp(). The partial byte is considered with more-significant bits first, so this is e.g. suitable for comparing IP prefixes.

static uint8\_t **KEY\_FLAG\_RANK** (const char \* *key*)

static bool **KEY\_COVERING\_RRSIG** (const char \* *key*)

*KR\_EXPORT* int **kr\_rrkey** (char \* *key*, uint16\_t *class*, const knot\_dname\_t \* *owner*, uint16\_t *type*,  
uint16\_t *additional*)

Create unique null-terminated string key for RR.

**Return** key length if successful or an error

#### Parameters

- *key*: Destination buffer for key size, MUST be KR\_RRKEY\_LEN or larger.
- *class*: RR class.
- *owner*: RR owner name.
- *type*: RR type.
- *additional*: flags (for instance can be used for storing covered type when RR type is RRSIG).

*KR\_EXPORT* int **kr\_ranked\_rrarray\_add** (ranked\_rr\_array\_t \* *array*, const knot\_rrset\_t \* *rr*,  
uint8\_t *rank*, bool *to\_wire*, uint32\_t *qry\_uid*, knot\_mm\_t  
\* *pool*)

int **kr\_ranked\_rrarray\_set\_wire** (ranked\_rr\_array\_t \* *array*, bool *to\_wire*, uint32\_t *qry\_uid*,  
bool *check\_dups*, bool(\**extraCheck*)(const ranked\_rr\_array\_entry\_t  
\*))

*KR\_PURE* char\* **kr\_pkt\_text** (const knot\_pkt\_t \* *pkt*)

*KR\_PURE* char\* **kr\_rrset\_text** (const knot\_rrset\_t \* *rr*)

static *KR\_PURE* char\* **kr\_dname\_text** (const knot\_dname\_t \* *name*)

static *KR\_CONST* char\* **kr\_rrtype\_text** (const uint16\_t *rrtype*)

*KR\_EXPORT* char\* **kr\_module\_call** (struct *kr\_context* \* *ctx*, const char \* *module*, const char \* *prop*, const  
char \* *input*)

Call module property.

static uint16\_t **kr\_rrset\_type\_maysig** (const knot\_rrset\_t \* *rr*)

Return the (covered) type of a nonempty RRset.

static const char\* **lua\_push\_printf** (lua\_State \* *L*, const char \* *fmt*, ...)

Printf onto the lua stack, avoiding additional copy (thin wrapper).

static char\* **kr\_straddr** (const struct sockaddr \* *addr*)

*KR\_EXPORT* uint64\_t **kr\_now** ()

The current time in monotonic milliseconds.

**Note** it may be outdated in case of long callbacks; see uv\_now().

int **knot\_dname\_lf2wire** (knot\_dname\_t \* *dst*, uint8\_t *len*, const uint8\_t \* *lf*)

Convert name from lookup format to wire.

See knot\_dname\_lf

**Note** *len* bytes are read and *len*+1 are written with *normal* LF, but it's also allowed that the final zero byte is omitted in LF.

**Return** the number of bytes written (>0) or error code (<0)

static int **kr\_dname\_lf** (uint8\_t \* *dst*, const knot\_dname\_t \* *src*, bool *add\_wildcard*)  
Patched knot\_dname\_lf.

LF for “.” has length zero instead of one, for consistency. (TODO: consistency?)

**Note** packet is always NULL

### Parameters

- *add\_wildcard*: append the wildcard label

*KR\_EXPORT* void **kr\_rrset\_init** (knot\_rrset\_t \* *rrset*, knot\_dname\_t \* *owner*, uint16\_t *type*,  
uint16\_t *rclass*, uint32\_t *ttd*)

*KR\_EXPORT* uint16\_t **kr\_pkt\_qclass** (const knot\_pkt\_t \* *pkt*)

*KR\_EXPORT* uint16\_t **kr\_pkt\_qtype** (const knot\_pkt\_t \* *pkt*)

*KR\_EXPORT* uint32\_t **kr\_rrsig\_sig\_inception** (const knot\_rdata\_t \* *rdata*)

*KR\_EXPORT* uint32\_t **kr\_rrsig\_sig\_expiration** (const knot\_rdata\_t \* *rdata*)

*KR\_EXPORT* uint16\_t **kr\_rrsig\_type\_covered** (const knot\_rdata\_t \* *rdata*)

### Variables

*KR\_EXPORT* bool **kr\_verbose\_status**  
Whether in verbose mode.

Only use this for reading.

*KR\_EXPORT* const char\* **source**

*KR\_EXPORT* const char const char\* **fmt**

const uint8\_t **KEY\_FLAG\_RRSIG** = 0x02

**union inaddr**

*#include <utils.h>* Simple storage for IPx address or AF\_UNSPEC.

### Public Members

struct sockaddr **ip**

struct sockaddr\_in **ip4**

struct sockaddr\_in6 **ip6**

### Defines

**KR\_EXPORT**

**KR\_CONST**

**KR\_PURE**

**KR\_NORETURN**

**KR\_COLD**

**KR\_PRINTF** (n)

**uint**



`kr_ok``kr_strerror(x)`

## Typedefs

`typedef unsigned int uint`

## Functions

`static int KR_COLD kr_error(int x)`

### 4.6.6 Generics library

This small collection of “generics” was born out of frustration that I couldn’t find no such thing for C. It’s either bloated, has poor interface, null-checking is absent or doesn’t allow custom allocation scheme. BSD-licensed (or compatible) code is allowed here, as long as it comes with a test case in `tests/test_generics.c`.

- *array* - a set of simple macros to make working with dynamic arrays easier.
- *queue* - a FIFO + LIFO queue.
- *map* - a Crit-bit tree key-value map implementation (public domain) that comes with tests.
- *set* - set abstraction implemented on top of map (unused now).
- *pack* - length-prefixed list of objects (i.e. array-list).
- *lru* - LRU-like hash table
- *trie* - a trie-based key-value map, taken from knot-dns

#### array

A set of simple macros to make working with dynamic arrays easier.

```
MIN(array_push(arr, val), other)
```

**Note** The C has no generics, so it is implemented mostly using macros. Be aware of that, as direct usage of the macros in the evaluating macros may lead to different expectations:

May evaluate the code twice, leading to unexpected behaviour. This is a price to pay for the absence of proper generics.

*Example usage:*

```
array_t(const char*) arr;
array_init(arr);

// Reserve memory in advance
if (array_reserve(arr, 2) < 0) {
    return ENOMEM;
}

// Already reserved, cannot fail
array_push(arr, "princess");
array_push(arr, "leia");
```

(continues on next page)

```

// Not reserved, may fail
if (array_push(arr, "han") < 0) {
    return ENOMEM;
}

// It does not hide what it really is
for (size_t i = 0; i < arr.len; ++i) {
    printf("%s\n", arr.at[i]);
}

// Random delete
array_del(arr, 0);

```

## Defines

### **array\_t** (type)

Declare an array structure.

### **array\_init** (array)

Zero-initialize the array.

### **array\_clear** (array)

Free and zero-initialize the array (plain malloc/free).

### **array\_clear\_mm** (array, free, baton)

Make the array empty and free pointed-to memory.

Mempool usage: pass mm\_free and a knot\_mm\_t\* .

### **array\_reserve** (array, n)

Reserve capacity for at least n elements.

**Return** 0 if success, <0 on failure

### **array\_reserve\_mm** (array, n, reserve, baton)

Reserve capacity for at least n elements.

Mempool usage: pass kr\_memreserve and a knot\_mm\_t\* .

**Return** 0 if success, <0 on failure

### **array\_push\_mm** (array, val, reserve, baton)

Push value at the end of the array, resize it if necessary.

Mempool usage: pass kr\_memreserve and a knot\_mm\_t\* .

**Note** May fail if the capacity is not reserved.

**Return** element index on success, <0 on failure

### **array\_push** (array, val)

Push value at the end of the array, resize it if necessary (plain malloc/free).

**Note** May fail if the capacity is not reserved.

**Return** element index on success, <0 on failure

### **array\_pop** (array)

Pop value from the end of the array.

**array\_del** (array, i)

Remove value at given index.

**Return** 0 on success, <0 on failure

**array\_tail** (array)

Return last element of the array.

**Warning** Undefined if the array is empty.

## Functions

static size\_t **array\_next\_count** (size\_t want)

Simplified Qt containers growth strategy.

static int **array\_std\_reserve** (void \* baton, char \*\* mem, size\_t elm\_size, size\_t want, size\_t \* have)

static void **array\_std\_free** (void \* baton, void \* p)

## queue

A queue, usable for FIFO and LIFO simultaneously.

Both the head and tail of the queue can be accessed and pushed to, but only the head can be popped from.

Example usage:

```
// define new queue type, and init a new queue instance
typedef queue_t(int) queue_int_t;
queue_int_t q;
queue_init(q);
// do some operations
queue_push(q, 1);
queue_push(q, 2);
queue_push(q, 3);
queue_push(q, 4);
queue_pop(q);
assert(queue_head(q) == 2);
assert(queue_tail(q) == 4);

// you may iterate
typedef queue_it_t(int) queue_it_int_t;
for (queue_it_int_t it = queue_it_begin(q); !queue_it_finished(it);
     queue_it_next(it)) {
    ++queue_it_val(it);
}
assert(queue_tail(q) == 5);

queue_push_head(q, 0);
++queue_tail(q);
assert(queue_tail(q) == 6);
// free it up
queue_deinit(q);

// you may use dynamic allocation for the type itself
queue_int_t *qm = malloc(sizeof(queue_int_t));
```

(continues on next page)

```
queue_init (*qm);
queue_deinit (*qm);
free (qm);
```

**Note** The implementation uses a singly linked list of blocks where each block stores an array of values (for better efficiency).

## Defines

### **queue\_t** (type)

The type for queue, parametrized by value type.

### **queue\_init** (q)

Initialize a queue.

You can malloc() it the usual way.

### **queue\_deinit** (q)

De-initialize a queue: make it invalid and free any inner allocations.

### **queue\_push** (q, data)

Push data to queue's tail.

(Type-safe version; use `_impl()` otherwise.)

### **queue\_push\_head** (q, data)

Push data to queue's head.

(Type-safe version; use `_impl()` otherwise.)

### **queue\_pop** (q)

Remove the element at the head.

### **queue\_head** (q)

Return a "reference" to the element at the head (it's an L-value) .

### **queue\_tail** (q)

Return a "reference" to the element at the tail (it's an L-value) .

### **queue\_len** (q)

Return the number of elements in the queue.

### **queue\_it\_t** (type)

Type for queue iterator, parametrized by value type.

It's a simple structure that owns no other resources. You may NOT use it after doing any push or pop (without `_begin` again).

### **queue\_it\_begin** (q)

Initialize a queue iterator at the head of the queue.

If you use this in assignment (instead of initialization), you will unfortunately need to add corresponding type-cast in front. Beware: there's no type-check between queue and iterator!

### **queue\_it\_val** (it)

Return a "reference" to the current element (it's an L-value) .

### **queue\_it\_finished** (it)

Test if the iterator has gone past the last element.

If it has, you may not use `_val` or `_next`.

**queue\_it\_next** (it)

Advance the iterator to the next element.

**map**

A Crit-bit tree key-value map implementation.

*Example usage:*

**Warning** If the user provides a custom allocator, it must return addresses aligned to 2B boundary.

```
map_t map = map_make(NULL);

// Custom allocator (optional)
map.malloc = &mymalloc;
map.baton = &mymalloc_context;

// Insert k-v pairs
int values = { 42, 53, 64 };
if (map_set(&map, "princess", &values[0]) != 0 ||
    map_set(&map, "prince", &values[1]) != 0 ||
    map_set(&map, "leia", &values[2]) != 0) {
    fail();
}

// Test membership
if (map_contains(&map, "leia")) {
    success();
}

// Prefix search
int i = 0;
int count(const char *k, void *v, void *ext) { (*(int *)ext)++; return 0; }
if (map_walk_prefixed(map, "princ", count, &i) == 0) {
    printf("%d matches\n", i);
}

// Delete
if (map_del(&map, "badkey") != 0) {
    fail(); // No such key
}

// Clear the map
map_clear(&map);
```

**Defines**

**map\_walk** (map, callback, baton)

**Functions**

static *map\_t* **map\_make** (struct knot\_mm \* pool)

Creates an new empty critbit map.

Pass NULL for malloc+free.

int **map\_contains** (*map\_t* \* *map*, const char \* *str*)  
Returns non-zero if map contains str.

void\* **map\_get** (*map\_t* \* *map*, const char \* *str*)  
Returns value if map contains str.

Note: NULL may mean two different things.

int **map\_set** (*map\_t* \* *map*, const char \* *str*, void \* *val*)  
Inserts str into map.

Returns 0 if new, 1 if replaced, or ENOMEM.

int **map\_del** (*map\_t* \* *map*, const char \* *str*)  
Deletes str from the map, returns 0 on success.

void **map\_clear** (*map\_t* \* *map*)  
Clears the given map.

int **map\_walk\_prefixed** (*map\_t* \* *map*, const char \* *prefix*, int(\**callback*)(const char \*, void \*, void \*),  
void \* *baton*)

Calls callback for all strings in map with the given prefix.

Returns value immediately if a callback returns nonzero.

### Parameters

- map:
- prefix: required string prefix (empty => all strings)
- callback: callback parameters are (key, value, baton)
- baton: passed user value

**struct map\_t**  
*#include <map.h>* Main data structure.

### Public Members

void\* **root**  
struct knot\_mm\* **pool**

## set

A set abstraction implemented on top of map.

*Example usage:*

**Note** The API is based on map.h, see it for more examples.

```
set_t set = set_make(NULL);  
  
// Insert keys  
if (set_add(&set, "princess") != 0 ||  
    set_add(&set, "prince") != 0 ||  
    set_add(&set, "leia") != 0) {  
    fail();  
}
```

(continues on next page)

(continued from previous page)

```

// Test membership
if (set_contains(&set, "leia")) {
    success();
}

// Prefix search
int i = 0;
int count(const char *s, void *n) { (*(int *)n)++; return 0; }
if (set_walk_prefixed(set, "princ", count, &i) == 0) {
    printf("%d matches\n", i);
}

// Delete
if (set_del(&set, "badkey") != 0) {
    fail(); // No such key
}

// Clear the set
set_clear(&set);

```

## Defines

### **set\_make**

Creates a new, empty critbit set

### **set\_contains** (set, str)

Returns non-zero if set contains str

### **set\_add** (set, str)

Inserts str into set. Returns 0 if new, 1 if already present, or ENOMEM.

### **set\_del** (set, str)

Deletes str from the set, returns 0 on success

### **set\_clear** (set)

Clears the given set

### **set\_walk** (set, callback, baton)

Calls callback for all strings in map

### **set\_walk\_prefixed** (set, prefix, callback, baton)

Calls callback for all strings in set with the given prefix

## Typedefs

```
typedef map_t set_t
```

```
typedef int( set_walk_cb) (const char *, void *)
```

## pack

A length-prefixed list of objects, also an array list.

Each object is prefixed by item length, unlike array this structure permits variable-length data. It is also equivalent to forward-only list backed by an array.

*Example usage:*

**Note** Maximum object size is  $2^{16}$  bytes, see [pack\\_objlen\\_t](#) If some mistake happens somewhere, the access may end up in an infinite loop. (equality comparison on pointers)

```
pack_t pack;
pack_init(pack);

// Reserve 2 objects, 6 bytes total
pack_reserve(pack, 2, 4 + 2);

// Push 2 objects
pack_obj_push(pack, U8("jedi"), 4)
pack_obj_push(pack, U8("\xbe\xef"), 2);

// Iterate length-value pairs
uint8_t *it = pack_head(pack);
while (it != pack_tail(pack)) {
    uint8_t *val = pack_obj_val(it);
    it = pack_obj_next(it);
}

// Remove object
pack_obj_del(pack, U8("jedi"), 4);

pack_clear(pack);
```

### Defines

**pack\_init** (pack)

Zero-initialize the pack.

**pack\_clear** (pack)

Make the pack empty and free pointed-to memory (plain malloc/free).

**pack\_clear\_mm** (pack, free, baton)

Make the pack empty and free pointed-to memory.

Mempool usage: pass mm\_free and a knot\_mm\_t\* .

**pack\_reserve** (pack, objs\_count, objs\_len)

Reserve space for *additional* objects in the pack (plain malloc/free).

**Return** 0 if success, <0 on failure

**pack\_reserve\_mm** (pack, objs\_count, objs\_len, reserve, baton)

Reserve space for *additional* objects in the pack.

Mempool usage: pass kr\_memreserve and a knot\_mm\_t\* .

**Return** 0 if success, <0 on failure

**pack\_head** (pack)

Return pointer to first packed object.

Recommended way to iterate: for (uint8\_t \*it = [pack\\_head\(pack\)](#); it != [pack\\_tail\(pack\)](#); it = [pack\\_obj\\_next\(it\)](#))

**pack\_tail** (pack)

Return pack end pointer.



## Typedefs

**typedef** `uint16_t pack_objlen_t`  
Packed object length type.

## Functions

**typedef** `array_t (uint8_t)`  
Pack is defined as an array of bytes.

**static** `pack_objlen_t pack_obj_len (uint8_t * it)`  
Return packed object length.

**static** `uint8_t* pack_obj_val (uint8_t * it)`  
Return packed object value.

**static** `uint8_t* pack_obj_next (uint8_t * it)`  
Return pointer to next packed object.

**static** `uint8_t* pack_last (pack_t pack)`  
Return pointer to the last packed object.

**static** `int pack_obj_push (pack_t * pack, const uint8_t * obj, pack_objlen_t len)`  
Push object to the end of the pack.

**Return** 0 on success, negative number on failure

**static** `uint8_t* pack_obj_find (pack_t * pack, const uint8_t * obj, pack_objlen_t len)`  
Returns a pointer to packed object.

**Return** pointer to packed object or NULL

**static** `int pack_obj_del (pack_t * pack, const uint8_t * obj, pack_objlen_t len)`  
Delete object from the pack.

**Return** 0 on success, negative number on failure

**static** `int pack_clone (pack_t ** dst, const pack_t * src, knot_mm_t * pool)`  
Clone a pack, replacing destination pack; (\*dst == NULL) is valid input.

**Return** `kr_error(ENOMEM)` on allocation failure.

## lru

A lossy cache.

Example usage:

```
// Define new LRU type
typedef lru_t(int) lru_int_t;

// Create LRU
lru_int_t *lru;
lru_create(&lru, 5, NULL, NULL);
```

(continues on next page)

(continued from previous page)

```

// Insert some values
int *pi = lru_get_new(lru, "luke", strlen("luke"), NULL);
if (pi)
    *pi = 42;
pi = lru_get_new(lru, "leia", strlen("leia"), NULL);
if (pi)
    *pi = 24;

// Retrieve values
int *ret = lru_get_try(lru, "luke", strlen("luke"), NULL);
if (!ret) printf("luke dropped out!\n");
else printf("luke's number is %d\n", *ret);

char *enemies[] = {"goro", "raiden", "subzero", "scorpion"};
for (int i = 0; i < 4; ++i) {
    int *val = lru_get_new(lru, enemies[i], strlen(enemies[i]), NULL);
    if (val)
        *val = i;
}

// We're done
lru_free(lru);

```

**Note** The implementation tries to keep frequent keys and avoid others, even if “used recently”, so it may refuse to store it on `lru_get_new()`. It uses hashing to split the problem pseudo-randomly into smaller groups, and within each it tries to approximate relative usage counts of several most frequent keys/hashes. This tracking is done for *more* keys than those that are actually stored.

## Defines

### `lru_t` (type)

The type for LRU, parametrized by value type.

### `lru_create` (ptable, max\_slots, mm\_ctx\_array, mm\_ctx)

Allocate and initialize an LRU with default associativity.

The real limit on the number of slots can be a bit larger but less than double.

**Note** The pointers to memory contexts need to remain valid during the whole life of the structure (or be NULL).

### Parameters

- `ptable`: pointer to a pointer to the LRU
- `max_slots`: number of slots
- `mm_ctx_array`: memory context to use for the huge array, NULL for default
- `mm_ctx`: memory context to use for individual key-value pairs, NULL for default

### `lru_free` (table)

Free an LRU created by `lru_create` (it can be NULL).

### `lru_reset` (table)

Reset an LRU to the empty state (but preserve any settings).

### `lru_get_try` (table, key\_, len\_)

Find key in the LRU and return pointer to the corresponding value.

**Return** pointer to data or NULL if not found

**Parameters**

- `table`: pointer to LRU
- `key_`: lookup key
- `len_`: key length

**lru\_get\_new** (`table`, `key_`, `len_`, `res`)

Return pointer to value, inserting if needed (zeroed).

**Return** pointer to data or NULL (can be even if memory could be allocated!)

**Parameters**

- `table`: pointer to LRU
- `key_`: lookup key
- `len_`: key lengthkeys
- `res`: pointer to bool to store result of operation (true if entry is newly added, false otherwise; can be NULL).

**lru\_apply** (`table`, `function`, `baton`)

Apply a function to every item in LRU.

**Parameters**

- `table`: pointer to LRU
- `function`: enum `lru_apply_do` (`*function`)(`const char *key`, `uint len`, `val_type *val`, `void *baton`)  
See enum `lru_apply_do` for the return type meanings.
- `baton`: extra pointer passed to each function invocation

**lru\_capacity** (`table`)

Return the real capacity - maximum number of keys holdable within.

**Parameters**

- `table`: pointer to LRU

## Enums

**enum lru\_apply\_do**

Possible actions to do with an element.

*Values:*

**LRU\_APPLY\_DO\_NOTHING**

**LRU\_APPLY\_DO\_EVICT**

## Functions

static *uint* **round\_power** (*uint size*, *uint power*)

Round the value up to a multiple of (1 << power).

## trie

### Typedefs

**typedef** void\* **trie\_val\_t**

Native API of QP-tries:

- keys are char strings, not necessarily zero-terminated, the structure copies the contents of the passed keys
- values are void\* pointers, typically you get an ephemeral pointer to it
- key lengths are limited by  $2^{32}-1$  ATM

XXX EDITORS: `trie.{h,c}` are synced from <https://gitlab.labs.nic.cz/knot/knot-dns/tree/68352fc969/src/contrib/qp-trie> only with tiny adjustments, mostly `#includes` and `KR_EXPORT`.

Element value.

**typedef** struct trie **trie\_t**

Opaque structure holding a QP-trie.

**typedef** struct trie\_it **trie\_it\_t**

Opaque type for holding a QP-trie iterator.

### Functions

*KR\_EXPORT* **trie\_t\*** **trie\_create** (knot\_mm\_t \* mm)

Create a trie instance. Pass NULL to use malloc+free.

*KR\_EXPORT* void **trie\_free** (trie\_t \* tbl)

Free a trie instance.

*KR\_EXPORT* void **trie\_clear** (trie\_t \* tbl)

Clear a trie instance (make it empty).

*KR\_EXPORT* size\_t **trie\_weight** (const trie\_t \* tbl)

Return the number of keys in the trie.

*KR\_EXPORT* trie\_val\_t\* **trie\_get\_try** (trie\_t \* tbl, const char \* key, uint32\_t len)

Search the trie, returning NULL on failure.

*KR\_EXPORT* trie\_val\_t\* **trie\_get\_first** (trie\_t \* tbl, char \*\* key, uint32\_t \* len)

Return pointer to the minimum. Optionally with key and its length.

*KR\_EXPORT* trie\_val\_t\* **trie\_get\_ins** (trie\_t \* tbl, const char \* key, uint32\_t len)

Search the trie, inserting NULL `trie_val_t` on failure.

*KR\_EXPORT* int **trie\_get\_leq** (trie\_t \* tbl, const char \* key, uint32\_t len, trie\_val\_t \*\* val)

Search for less-or-equal element.

**Return** KNOT\_EOK for exact match, 1 for previous, KNOT\_ENOENT for not-found, or KNOT\_E\*.

#### Parameters

- `tbl`: Trie.
- `key`: Searched key.
- `len`: Key length.
- `val`: Must be valid; it will be set to NULL if not found or errored.

int **trie\_apply** (*trie\_t* \* *tbl*, int(\*f)(*trie\_val\_t* \*, void \*), void \* *d*)

Apply a function to every *trie\_val\_t*, in order.

**Return** First nonzero from *f*() or zero (i.e. KNOT\_EOK).

**Parameters**

- *d*: Parameter passed as the second argument to *f*().

*KR\_EXPORT* int **trie\_del** (*trie\_t* \* *tbl*, const char \* *key*, uint32\_t *len*, *trie\_val\_t* \* *val*)

Remove an item, returning KNOT\_EOK if succeeded or KNOT\_ENOENT if not found.

If *val*!=NULL and deletion succeeded, the deleted value is set.

*KR\_EXPORT* int **trie\_del\_first** (*trie\_t* \* *tbl*, char \* *key*, uint32\_t \* *len*, *trie\_val\_t* \* *val*)

Remove the first item, returning KNOT\_EOK on success.

You may optionally get the key and/or value. The key is copied, so you need to pass sufficient *len*, otherwise *kr\_error*(ENOSPC) is returned.

*KR\_EXPORT* *trie\_it\_t*\* **trie\_it\_begin** (*trie\_t* \* *tbl*)

Create a new iterator pointing to the first element (if any).

*KR\_EXPORT* void **trie\_it\_next** (*trie\_it\_t* \* *it*)

Advance the iterator to the next element.

Iteration is in ascending lexicographical order. In particular, the empty string would be considered as the very first.

**Note** You may not use this function if the trie's key-set has been modified during the lifetime of the iterator (modifying values only is OK).

*KR\_EXPORT* bool **trie\_it\_finished** (*trie\_it\_t* \* *it*)

Test if the iterator has gone past the last element.

*KR\_EXPORT* void **trie\_it\_free** (*trie\_it\_t* \* *it*)

Free any resources of the iterator. It's OK to call it on NULL.

*KR\_EXPORT* const char\* **trie\_it\_key** (*trie\_it\_t* \* *it*, size\_t \* *len*)

Return pointer to the key of the current element.

**Note** The optional *len* is *uint32\_t* internally but *size\_t* is better for our usage, as it is without an additional type conversion.

*KR\_EXPORT* *trie\_val\_t*\* **trie\_it\_val** (*trie\_it\_t* \* *it*)

Return pointer to the value of the current element (writable).



- *Supported languages*
- *The anatomy of an extension*
- *Writing a module in Lua*
- *Writing a module in C*
- *Writing a module in Go*
- *Configuring modules*
- *Exposing C/Go module properties*

## 5.1 Supported languages

Currently modules written in C and LuaJIT are supported. There is also a support for writing modules in Go 1.5+ — the library has no native Go bindings, library is accessible using `CGO`.

## 5.2 The anatomy of an extension

A module is a shared object or script defining specific functions, here's an overview.

*Note* — the *Modules* header documents the module loading and API.

C/Go	Lua	Params	Comment
<code>X_api()</code> <sup>1</sup>			API version
<code>X_init()</code>	<code>X.init()</code>	module	Constructor
<code>X_deinit()</code>	<code>X.deinit()</code>	module, key	Destructor
<code>X_config()</code>	<code>X.config()</code>	module	Configuration
<code>X_layer()</code>	<code>X.layer</code>	module	<i>Module layer</i>
<code>X_props()</code>			List of properties

The X corresponds to the module name, if the module name is `hints`, then the prefix for constructor would be `hints_init()`. This doesn't apply for Go, as it for now always implements *main* and requires capitalized first letter in order to export its symbol.

---

**Note:** The resolution context struct `kr_context` holds loaded modules for current context. A module can be registered with `kr_context_register()`, which triggers module constructor *immediately* after the load. Module destructor is automatically called when the resolution context closes.

If the module exports a layer implementation, it is automatically discovered by `kr_resolver()` on resolution init and plugged in. The order in which the modules are registered corresponds to the call order of layers.

---

## 5.3 Writing a module in Lua

The probably most convenient way of writing modules is Lua since you can use already installed modules from system and have first-class access to the scripting engine. You can also tap to all the events, that the C API has access to, but keep in mind that transitioning from the C to Lua function is slower than the other way round.

---

**Note:** The Lua functions retrieve an additional first parameter compared to the C counterparts - a "state". There is no Lua wrapper for C structures used in the resolution context, until they're implemented you can inspect the structures using the `ffi` library.

---

The modules follow the [Lua way](#), where the module interface is returned in a named table.

```

--- @module Count incoming queries
local counter = {}

function counter.init(module)
    counter.total = 0
    counter.last = 0
    counter.failed = 0
end

function counter.deinit(module)
    print('counted', counter.total, 'queries')
end

-- @function Run the q/s counter with given interval.
function counter.config(conf)
    -- We can use the scripting facilities here
    if counter.ev then event.cancel(counter.ev)

```

(continues on next page)

---

<sup>1</sup> Mandatory symbol.



(continued from previous page)

```

    event.recurrent(conf.interval, function ()
        print(counter.total - counter.last, 'q/s')
        counter.last = counter.total
    end)
end
return counter

```

**Tip:** The API functions may return an integer value just like in other languages, but they may also return a coroutine that will be continued asynchronously. A good use case for this approach is a deferred initialization, e.g. loading a chunks of data or waiting for I/O.

```

function counter.init(module)
    counter.total = 0
    counter.last = 0
    counter.failed = 0
    return coroutine.create(function ()
        for line in io.lines('/etc/hosts') do
            load(module, line)
            coroutine.yield()
        end
    end)
end
end

```

The created module can be then loaded just like any other module, except it isn't very useful since it doesn't provide any layer to capture events. The Lua module can however provide a processing layer, just *like its C counterpart*.

```

-- Notice it isn't a function, but a table of functions
counter.layer = {
    begin = function (state, data)
        counter.total = counter.total + 1
        return state
    end,
    finish = function (state, req, answer)
        if state == kres.FAIL then
            counter.failed = counter.failed + 1
        end
        return state
    end
}

```

There is currently an additional “feature” in comparison to C layer functions: the consume, produce and checkout functions do not get called at all if `state == kres.FAIL` (note that finish does get called nevertheless).

Since the modules are like any other Lua modules, you can interact with them through the CLI and any interface.

**Tip:** The module can be placed anywhere in the Lua search path, in the working directory or in the MODULESDIR.

## 5.4 Writing a module in C

As almost all the functions are optional, the minimal module looks like this:

```
#include "lib/module.h"
/* Convenience macro to declare module API. */
KR_MODULE_EXPORT(mymodule);
```

Let's define an observer thread for the module as well. It's going to be stub for the sake of brevity, but you can for example create a condition, and notify the thread from query processing by declaring module layer (see the [Writing layers](#)).

```
static void* observe(void *arg)
{
    /* ... do some observing ... */
}

int mymodule_init(struct kr_module *module)
{
    /* Create a thread and start it in the background. */
    pthread_t thr_id;
    int ret = pthread_create(&thr_id, NULL, &observe, NULL);
    if (ret != 0) {
        return kr_error(errno);
    }

    /* Keep it in the thread */
    module->data = thr_id;
    return kr_ok();
}

int mymodule_deinit(struct kr_module *module)
{
    /* ... signalize cancellation ... */
    void *res = NULL;
    pthread_t thr_id = (pthread_t) module->data;
    int ret = pthread_join(thr_id, res);
    if (ret != 0) {
        return kr_error(errno);
    }

    return kr_ok();
}
```

This example shows how a module can run in the background, this enables you to, for example, observe and publish data about query resolution.

## 5.5 Writing a module in Go

The Go modules use `CGO` to interface C resolver library, there are no native bindings yet. Second issue is that layers are declared as a structure of function pointers, which are **not present in Go**, the workaround is to declare them in `CGO` header. Each module must be the `main` package, here's a minimal example:

```

package main

/*
#include "lib/module.h"
*/
import "C"
import "unsafe"

/* Mandatory functions */

//export mymodule_api
func mymodule_api() C.uint32_t {
    return C.KR_MODULE_API
}
func main() {}

```

**Warning:** Do not forget to prefix function declarations with `//export symbol_name`, as only these will be exported in module.

In order to integrate with query processing, you have to declare a helper function with function pointers to the layer implementation. Since the code prefacing `import "C"` is expanded in headers, you need the *static inline* trick to avoid multiple declarations. Here's how the preface looks like:

```

/*
#include "lib/layer.h"
#include "lib/module.h"
// Need a forward declaration of the function signature
int finish(kr_layer_t *);
// Workaround for layers composition
static inline const kr_layer_api_t *_layer(void)
{
    static const kr_layer_api_t api = {
        .finish = &finish
    };
    return &api;
}
*/
import "C"
import "unsafe"

```

Now we can add the implementations for the `finish` layer and finalize the module:

```

//export finish
func finish(ctx *C.kr_layer_t) C.int {
    // Since the context is unsafe.Pointer, we need to cast it
    var param *C.struct_kr_request = (*C.struct_kr_request)(ctx.data)
    // Now we can use the C API as well
    fmt.Printf("[go] resolved %d queries\n", C.list_size(&param.rplan.resolved))
    return 0
}

//export mymodule_layer
func mymodule_layer(module *C.struct_kr_module) *C.kr_layer_api_t {
    // Wrapping the inline trampoline function
    return C._layer()
}

```

(continues on next page)

}

See the [CGO](#) for more information about type conversions and interoperability between the C/Go.

### 5.5.1 Gotchas

- `main()` function is mandatory in each module, otherwise it won't compile.
- Module layer function implementation must be done in C during `import "C"`, as Go doesn't support pointers to functions.
- The library doesn't have a Go-ified bindings yet, so interacting with it requires CGO shims, namely structure traversal and type conversions (strings, numbers).
- Other modules can be called through C call `C.kr_module_call(kr_context, module_name, module_property, input)`

## 5.6 Configuring modules

There is a callback `X_config()` that you can implement, see `hints` module.

## 5.7 Exposing C/Go module properties

A module can offer NULL-terminated list of *properties*, each property is essentially a callable with free-form JSON input/output. JSON was chosen as an interchangeable format that doesn't require any schema beforehand, so you can do two things - query the module properties from external applications or between modules (i.e. *statistics* module can query *cache* module for memory usage). JSON was chosen not because it's the most efficient protocol, but because it's easy to read and write and interface to outside world.

---

**Note:** The `void *env` is a generic module interface. Since we're implementing daemon modules, the pointer can be cast to `struct engine*`. This is guaranteed by the implemented API version (see [Writing a module in C](#)).

---

Here's an example how a module can expose its property:

```
char* get_size(void *env, struct kr_module *m,
              const char *args)
{
    /* Get cache from engine. */
    struct engine *engine = env;
    struct kr_cache *cache = &engine->resolver.cache;
    /* Read item count */
    int count = (cache->api)->count(cache->db);
    char *result = NULL;
    asprintf(&result, "{ \"result\": %d }", count);

    return result;
}

struct kr_prop *cache_props(void)
{
```

(continues on next page)

(continued from previous page)

```
static struct kr_prop prop_list[] = {
    /* Callback, Name, Description */
    {&get_size, "get_size", "Return number of records."},
    {NULL, NULL, NULL}
};
return prop_list;
}

KR_MODULE_EXPORT(cache)
```

Once you load the module, you can call the module property from the interactive console. *Note* — the JSON output will be transparently converted to Lua tables.

```
$ kresd
...
[system] started in interactive mode, type 'help()'
> modules.load('cached')
> cached.get_size()
[size] => 53
```

*Note* — this relies on function pointers, so the same `static inline` trick as for the `Layer()` is required for C/Go.

### 5.7.1 Special properties

If the module declares properties `get` or `set`, they can be used in the Lua interpreter as regular tables.



## 6.1 Knot Resolver 3.1.0 (2018-11-02)

### 6.1.1 Incompatible changes

- `hints.use_nodata(true)` by default; that's what most users want
- `libknot >= 2.7.2` is required

### 6.1.2 Improvements

- `cache`: handle out-of-space SIGBUS slightly better (#197)
- `daemon`: improve TCP timeout handling (!686)

### 6.1.3 Bugfixes

- `cache.clear('name')`: fix some edge cases in API (#401)
- fix error handling from TLS writes (!669)
- avoid SERVFAILs due to certain kind of NS dependency cycles (#374)

## 6.2 Knot Resolver 3.0.0 (2018-08-20)

### 6.2.1 Incompatible changes

- `cache`: fail lua operations if cache isn't open yet (!639) By default cache is opened *after* reading the configuration, and older versions were silently ignoring cache operations. Valid configuration must open cache using `cache.open()` or `cache.size =` before executing cache operations like `cache.clear()`.

- libknot >= 2.7.1 is required, which brings also larger API changes
- in case you wrote custom Lua modules, please consult <https://knot-resolver.readthedocs.io/en/latest/lib.html#incompatible-changes-since-3-0-0>
- in case you wrote custom C modules, please see compile against Knot DNS 2.7 and adjust your module according to messages from C compiler
- DNS cookie module (RFC 7873) is not available in this release, it will be later reworked to reflect development in IEFT dnsop working group
- version module was permanently removed because it was not really used by users; if you want to receive notifications about new releases please subscribe to <https://lists.nic.cz/cgi-bin/mailman/listinfo/knot-resolver-announce>

### 6.2.2 Bugfixes

- fix multi-process race condition in trust anchor maintenance (!643)
- ta\_sentinel: also consider static trust anchors not managed via RFC 5011

### 6.2.3 Improvements

- reorder\_RR() implementation is brought back
- bring in performance improvements provided by libknot 2.7
- cache.clear() has a new, more powerful API
- cache documentation was improved
- old name “Knot DNS Resolver” is replaced by unambiguous “Knot Resolver” to prevent confusion with “Knot DNS” authoritative server

## 6.3 Knot Resolver 2.4.1 (2018-08-02)

### 6.3.1 Security

- fix CVE-2018-10920: Improper input validation bug in DNS resolver component (security!7, security!9)

### 6.3.2 Bugfixes

- cache: fix TTL overflow in packet due to min\_ttl (#388, security!8)
- TLS session resumption: avoid bad scheduling of rotation (#385)
- HTTP module: fix a regression in 2.4.0 which broke custom certs (!632)
- cache: NSEC3 negative cache even without NS record (#384) This fixes lower hit rate in NSEC3 zones (since 2.4.0).
- minor TCP and TLS fixes (!623, !624, !626)



## 6.4 Knot Resolver 2.4.0 (2018-07-03)

### 6.4.1 Incompatible changes

- minimal libknot version is now 2.6.7 to pull in latest fixes (#366)

### 6.4.2 Security

- fix a rare case of zones incorrectly downgraded to insecure status (!576)

### 6.4.3 New features

- TLS session resumption (RFC 5077), both server and client (!585, #105) (disabled when compiling with gnutls < 3.5)
- TLS\_FORWARD policy uses system CA certificate store by default (!568)
- aggressive caching for NSEC3 zones (!600)
- optional protection from DNS Rebinding attack (module rebinding, !608)
- module bogus\_log to log DNSSEC bogus queries without verbose logging (!613)

### 6.4.4 Bugfixes

- prefill: fix ability to read certificate bundle (!578)
- avoid turning off qname minimization in some cases, e.g. co.uk. (#339)
- fix validation of explicit wildcard queries (#274)
- dns64 module: more properties from the RFC implemented (incl. bug #375)

### 6.4.5 Improvements

- systemd: multiple enabled kresd instances can now be started using kresd.target
- ta\_sentinel: switch to version 14 of the RFC draft (!596)
- support for glibc systems with a non-Linux kernel (!588)
- support per-request variables for Lua modules (!533)
- support custom HTTP endpoints for Lua modules (!527)

## 6.5 Knot Resolver 2.3.0 (2018-04-23)

### 6.5.1 Security

- fix CVE-2018-1110: denial of service triggered by malformed DNS messages (!550, !558, security!2, security!4)
- increase resilience against slow loris attack (security!5)

## 6.5.2 New features

- new policy.REFUSE to reply REFUSED to clients

## 6.5.3 Bugfixes

- validation: fix SERVFAIL in case of CNAME to NXDOMAIN in a single zone (!538)
- validation: fix SERVFAIL for DS . query (!544)
- lib/resolve: don't send unnecessary queries to parent zone (!513)
- iterate: fix validation for zones where parent and child share NS (!543)
- TLS: improve error handling and documentation (!536, !555, !559)

## 6.5.4 Improvements

- prefill: new module to periodically import root zone into cache (replacement for RFC 7706, !511)
- network\_listen\_fd: always create end point for supervisor supplied file descriptor
- use CPPFLAGS build environment variable if set (!547)

## 6.6 Knot Resolver 2.2.0 (2018-03-28)

### 6.6.1 New features

- cache server unavailability to prevent flooding unreachable servers (Please note that caching algorithm needs further optimization and will change in further versions but we need to gather operational experience first.)

### 6.6.2 Bugfixes

- don't magically -D\_FORTIFY\_SOURCE=2 in some cases
- allow large responses for outbound over TCP
- fix crash with RR sets with over 255 records

## 6.7 Knot Resolver 2.1.1 (2018-02-23)

### 6.7.1 Bugfixes

- when iterating, avoid unnecessary queries for NS in insecure parent. This problem worsened in 2.0.0. (#246)
- prevent UDP packet leaks when using TLS forwarding
- fix the hints module also on some other systems, e.g. Gentoo.

---

## 6.8 Knot Resolver 2.1.0 (2018-02-16)

### 6.8.1 Incompatible changes

- stats: remove tracking of expiring records (predict uses another way)
- systemd: re-use a single kresd.socket and kresd-tls.socket
- ta\_sentinel: implement protocol draft-ietf-dnsop-kskroll-sentinel-01 (our draft-ietf-dnsop-kskroll-sentinel-00 implementation had inverted logic)
- libknot: require version 2.6.4 or newer to get bugfixes for DNS-over-TLS

### 6.8.2 Bugfixes

- detect\_time\_jump module: don't clear cache on suspend-resume (#284)
- stats module: fix stats.list() returning nothing, regressed in 2.0.0
- policy.TLS\_FORWARD: refusal when configuring with multiple IPs (#306)
- cache: fix broken refresh of insecure records that were about to expire
- fix the hints module on some systems, e.g. Fedora (came back on 2.0.0)
- build with older gnutls (conditionally disable features)
- fix the predict module to work with insecure records & cleanup code

## 6.9 Knot Resolver 2.0.0 (2018-01-31)

### 6.9.1 Incompatible changes

- systemd: change unit files to allow running multiple instances, deployments with single instance now must use *kresd@1.service* instead of *kresd.service*; see *kresd.systemd(7)* for details
- systemd: the directory for cache is now */var/cache/knot-resolver*
- unify default directory and user to *knot-resolver*
- directory with trust anchor file specified by *-k* option must be writeable
- policy module is now loaded by default to enforce RFC 6761; see documentation for *policy.PASS* if you use locally-served DNS zones
- drop support for alternative cache backends memcached, redis, and for Lua bindings for some specific cache operations
- REORDER\_RR option is not implemented (temporarily)

### 6.9.2 New features

- aggressive caching of validated records (RFC 8198) for NSEC zones; thanks to ICANN for sponsoring this work.

- forwarding over TLS, authenticated by SPKI pin or certificate. `policy.TLS_FORWARD` pipelines queries out-of-order over shared TLS connection Beware: Some resolvers do not support out-of-order query processing. TLS forwarding to such resolvers will lead to slower resolution or failures.
- trust anchors: you may specify a read-only file via `-K` or `--keyfile-ro`
- trust anchors: at build-time you may set `KEYFILE_DEFAULT` (read-only)
- `ta_sentinel` module implements draft `ietf-dnsop-kskroll-sentinel-00`, enabled by default
- `serve_stale` module is prototype, subject to change
- extended API for Lua modules

### 6.9.3 Bugfixes

- fix build on osx - regressed in 1.5.3 (different linker option name)

## 6.10 Knot Resolver 1.5.3 (2018-01-23)

### 6.10.1 Bugfixes

- fix the hints module on some systems, e.g. Fedora. Symptom: *undefined symbol: engine\_hint\_root\_file*

## 6.11 Knot Resolver 1.5.2 (2018-01-22)

### 6.11.1 Security

- fix CVE-2018-1000002: insufficient DNSSEC validation, allowing attackers to deny existence of some data by forging packets. Some combinations pointed out in RFC 6840 sections 4.1 and 4.3 were not taken into account.

### 6.11.2 Bugfixes

- `memcached`: fix fallout from module rename in 1.5.1

## 6.12 Knot Resolver 1.5.1 (2017-12-12)

### 6.12.1 Incompatible changes

- `script supervisor.py` was removed, please migrate to a real process manager
- module `ketc` was renamed to `etcd` for consistency
- module `kmemcached` was renamed to `memcached` for consistency

## 6.12.2 Bugfixes

- fix SIGPIPE crashes (#271)
- tests: work around out-of-space for platforms with larger memory pages
- lua: fix mistakes in bindings affecting 1.4.0 and 1.5.0 (and 1.99.1-alpha), potentially causing problems in dns64 and workarounds modules
- predict module: various fixes (!399)

## 6.12.3 Improvements

- add priming module to implement RFC 8109, enabled by default (#220)
- add modules helping with system time problems, enabled by default; for details see documentation of detect\_time\_skew and detect\_time\_jump

## 6.13 Knot Resolver 1.5.0 (2017-11-02)

### 6.13.1 Bugfixes

- fix loading modules on Darwin

### 6.13.2 Improvements

- new module ta\_signal\_query supporting Signaling Trust Anchor Knowledge using Keytag Query (RFC 8145 section 5); it is enabled by default
- attempt validation for more records but require it for fewer of them (e.g. avoids SERVFAIL when server adds extra records but omits RRSIGs)

## 6.14 Knot Resolver 1.99.1-alpha (2017-10-26)

This is an experimental release meant for testing aggressive caching. It contains some regressions and might (theoretically) be even vulnerable. The current focus is to minimize queries into the root zone.

### 6.14.1 Improvements

- negative answers from validated NSEC (NXDOMAIN, NODATA)
- verbose log is very chatty around cache operations (maybe too much)

### 6.14.2 Regressions

- dropped support for alternative cache backends and for some specific cache operations
- **caching doesn't yet work for various cases:**
  - **negative answers without NSEC (i.e. with NSEC3 or insecure)**

- \* +cd queries (needs other internal changes)
- \* positive wildcard answers
- **spurious SERVFAIL on specific combinations of cached records, printing:** <= bad keys, broken trust chain
- make check
- a few Deckard tests are broken, probably due to some problems above
- also unknown ones?

## 6.15 Knot Resolver 1.4.0 (2017-09-22)

### 6.15.1 Incompatible changes

- lua: query flag-sets are no longer represented as plain integers. `kres.query.*` no longer works, and `kr_query_t` lost trivial methods `'hasflag'` and `'resolved'`. You can instead write code like `qry.flags.NO_0X20 = true`.

### 6.15.2 Bugfixes

- fix exiting one of multiple forks (#150)
- cache: change the way of using LMDB transactions. That in particular fixes some cases of using too much space with multiple kresd forks (#240).

### 6.15.3 Improvements

- `policy.suffix`: update the aho-corasick code (#200)
- root hints are now loaded from a zonefile; exposed as `hints.root_file()`. You can override the path by defining `ROOTHINTS` during compilation.
- `policy.FORWARD`: work around resolvers adding unsigned NS records (#248)
- reduce unneeded records previously put into authority in wildcarded answers

## 6.16 Knot Resolver 1.3.3 (2017-08-09)

### 6.16.1 Security

- Fix a critical DNSSEC flaw. Signatures might be accepted as valid even if the signed data was not in bailiwick of the DNSKEY used to sign it, assuming the trust chain to that DNSKEY was valid.

### 6.16.2 Bugfixes

- `iterate`: skip RRSIGs with bad label count instead of immediate SERVFAIL
- `utils`: fix possible incorrect seeding of the random generator
- `modules/http`: fix compatibility with the Prometheus text format

### 6.16.3 Improvements

- policy: implement remaining special-use domain names from RFC6761 (#205), and make these rules apply only if no other non-chain rule applies

## 6.17 Knot Resolver 1.3.2 (2017-07-28)

### 6.17.1 Security

- fix possible opportunities to use insecure data from cache as keys for validation

### 6.17.2 Bugfixes

- daemon: check existence of config file even if rundir isn't specified
- policy.FORWARD and STUB: use RTT tracking to choose servers (#125, #208)
- dns64: fix CNAME problems (#203) It still won't work with policy.STUB.
- **hints: better interpretation of hosts-like files (#204)** also, error out if a bad entry is encountered in the file
- dnssec: handle unknown DNSKEY/DS algorithms (#210)
- predict: fix the module, broken since 1.2.0 (#154)

### 6.17.3 Improvements

- embedded LMDB fallback: update 0.9.18 -> 0.9.21

## 6.18 Knot Resolver 1.3.1 (2017-06-23)

### 6.18.1 Bugfixes

- modules/http: fix finding the static files (bug from 1.3.0)
- policy.FORWARD: fix some cases of CNAMEs obstructing search for zone cuts

## 6.19 Knot Resolver 1.3.0 (2017-06-13)

### 6.19.1 Security

- Refactor handling of AD flag and security status of resource records. In some cases it was possible for secure domains to get cached as insecure, even for a TLD, leading to disabled validation. It also fixes answering with non-authoritative data about nameservers.

## 6.19.2 Improvements

- major feature: support for forwarding with validation (#112). The old policy.FORWARD action now does that; the previous non-validating mode is still available as policy.STUB except that also uses caching (#122).
- command line: specify ports via @ but still support # for compatibility
- policy: recognize 100.64.0.0/10 as local addresses
- layer/iterate: *do* retry repeatedly if REFUSED, as we can't yet easily retry with other NSs while avoiding retrying with those who REFUSED
- modules: allow changing the directory where modules are found, and do not search the default library path anymore.

## 6.19.3 Bugfixes

- validate: fix insufficient caching for some cases (relatively rare)
- avoid putting "duplicate" record-sets into the answer (#198)

## 6.20 Knot Resolver 1.2.6 (2017-04-24)

### 6.20.1 Security

- dnssec: don't set AD flag for NODATA answers if wildcard non-existence is not guaranteed due to opt-out in NSEC3

### 6.20.2 Improvements

- layer/iterate: don't retry repeatedly if REFUSED

### 6.20.3 Bugfixes

- lib/nsrep: revert some changes to NS reputation tracking that caused severe problems to some users of 1.2.5 (#178 and #179)
- dnssec: fix verification of wildcarded non-singleton RRsets
- dnssec: allow wildcards located directly under the root
- layer/rrcache: avoid putting answer records into queries in some cases

## 6.21 Knot Resolver 1.2.5 (2017-04-05)

### 6.21.1 Security

- layer/validate: clear AD if closest enclosing proof has opt-outed NSEC3 (#169)
- layer/validate: check if NSEC3 records in wildcard expansion proof has an opt-out
- dnssec/nsec: missed wildcard no-data answers validation has been implemented



## 6.21.2 Improvements

- modules/dnstap: a DNSTAP support module (Contributed by Vicky Shrestha)
- modules/workarounds: a module adding workarounds for known DNS protocol violators
- layer/iterate: fix logging of glue addresses
- kr\_bitcmp: allow bits=0 and consequently 0.0.0.0/0 matches in view and renumber modules.
- modules/padding: Improve default padding of responses (Contributed by Daniel Kahn Gillmor)
- New kresc client utility (experimental; don't rely on the API yet)

## 6.21.3 Bugfixes

- trust anchors: Improve trust anchors storage format (#167)
- trust anchors: support non-root TAs, one domain per file
- policy.DENY: set AA flag and clear AD flag
- lib/resolve: avoid unnecessary DS queries
- lib/nsrep: don't treat servers with NOIP4 + NOIP6 flags as timeouted
- layer/iterate: During packet classification (answer vs. referral) don't analyze AUTHORITY section in authoritative answer if ANSWER section contains records that have been requested

## 6.22 Knot Resolver 1.2.4 (2017-03-09)

### 6.22.1 Security

- Knot Resolver 1.2.0 and higher could return AD flag for insecure answer if the daemon received answer with invalid RRSIG several times in a row.

### 6.22.2 Improvements

- modules/policy: allow QTRACE policy to be chained with other policies
- hints.add\_hosts(path): a new property
- module: document the API and simplify the code
- policy.MIRROR: support IPv6 link-local addresses
- policy.FORWARD: support IPv6 link-local addresses
- add net.outgoing\_{v4,v6} to allow specifying address to use for connections

### 6.22.3 Bugfixes

- layer/iterate: some improvements in cname chain unrolling
- layer/validate: fix duplicate records in AUTHORITY section in case of WC expansion proof
- lua: do *not* truncate cache size to unsigned

- forwarding mode: correctly forward +cd flag
- fix a potential memory leak
- don't treat answers that contain DS non-existence proof as insecure
- don't store NSEC3 and their signatures in the cache
- layer/iterate: when processing delegations, check if qname is at or below new authority

## **6.23 Knot Resolver 1.2.3 (2017-02-23)**

### **6.23.1 Bugfixes**

- Disable storing GLUE records into the cache even in the (non-default) QUERY\_PERMISSIVE mode
- iterate: skip answer RRs that don't match the query
- layer/iterate: some additional processing for referrals
- lib/resolve: zonecut fetching error was fixed

## **6.24 Knot Resolver 1.2.2 (2017-02-10)**

### **6.24.1 Bugfixes:**

- Fix -k argument processing to avoid out-of-bounds memory accesses
- lib/resolve: fix zonecut fetching for explicit DS queries
- hints: more NULL checks
- Fix TA bootstrapping for multiple TAs in the IANA XML file

### **6.24.2 Testing:**

- Update tests to run tests with and without QNAME minimization

## **6.25 Knot Resolver 1.2.1 (2017-02-01)**

### **6.25.1 Security:**

- Under certain conditions, a cached negative answer from a CD query would be reused to construct response for non-CD queries, resulting in Insecure status instead of Bogus. Only 1.2.0 release was affected.

### **6.25.2 Documentation**

- Update the typo in the documentation: The query trace policy is named policy.QTRACE (and not policy.TRACE)

### 6.25.3 Bugfixes:

- lua: make the map command check its arguments

## 6.26 Knot Resolver 1.2.0 (2017-01-24)

### 6.26.1 Security:

- In a policy.FORWARD() mode, the AD flag was being always set by mistake. It is now cleared, as the policy.FORWARD() doesn't do DNSSEC validation yet.

### 6.26.2 Improvements:

- The DNSSEC Validation has been refactored, fixing many resolving failures.
- Add module *version* that checks for updates and CVEs periodically.
- Support RFC7830: EDNS(0) padding in responses over TLS.
- Support CD flag on incoming requests.
- hints module: previously */etc/hosts* was loaded by default, but not anymore. Users can now actually avoid loading any file.
- DNS over TLS now creates ephemeral certs.
- Configurable cache.{min,max}\_ttl option, with max\_ttl defaulting to 6 days.
- Option to reorder RRs in the response.
- New policy.QTRACE policy to print packet contents

### 6.26.3 Bugfixes:

- Trust Anchor configuration is now more robust.
- Correctly answer NOTIMPL for meta-types and non-IN RR classes.
- Free TCP buffer on cancelled connection.
- Fix crash in hints module on empty hints file, and fix non-lowercase hints.

### 6.26.4 Miscellaneous:

- It now requires knot >= 2.3.1 to link successfully.
- The API+ABI for modules changed slightly.
- New LRU implementation.

## 6.27 Knot Resolver 1.1.1 (2016-08-24)

### 6.27.1 Bugfixes:

- Fix 0x20 randomization with retransmit
- Fix pass-through for the stub mode
- Fix the root hints IPv6 addresses
- Fix dst addr for retries over TCP

### 6.27.2 Improvements:

- Track RTT of all tried servers for faster retransmit
- DAF: Allow forwarding to custom port
- systemd: Read EnvironmentFile and user \$KRESD\_ARGS
- systemd: Update systemd units to be named after daemon

## 6.28 Knot Resolver 1.1.0 (2016-08-12)

### 6.28.1 Improvements:

- RFC7873 DNS Cookies
- RFC7858 DNS over TLS
- HTTP/2 web interface, RESTful API
- Metrics exported in Prometheus
- DNS firewall module
- Explicit CNAME target fetching in strict mode
- Query minimisation improvements
- Improved integration with systemd

## 6.29 Knot Resolver 1.0.0 (2016-05-30)

### 6.29.1 Initial release:

- The first initial release

# CHAPTER 7

---

## Indices and tables

---

- `genindex`
- `modindex`
- `search`



**A**

array\_next\_count (C function), 95  
array\_std\_free (C function), 95  
array\_std\_reserve (C function), 95  
array\_t (C function), 101

**C**

cache.backends (C function), 16  
cache.clear (C function), 18  
cache.close (C function), 16  
cache.count (C function), 16  
cache.get (C function), 18  
cache.max\_ttl (C function), 17  
cache.min\_ttl (C function), 17  
cache.ns\_tout (C function), 18  
cache.open (C function), 15  
cache.stats (C function), 17  
cache\_peek (C function), 76  
cache\_stash (C function), 76

**E**

environment variable  
    cache.current\_size, 16  
    cache.current\_storage, 16  
    cache.size, 16  
    cache.storage, 16  
    env(table), 7  
    net.ipv4=truelfalse, 11  
    net.ipv6=truelfalse, 11  
    trust\_anchors.hold\_down\_time=30\*day, 14  
    trust\_anchors.keep\_removed=0, 14  
    trust\_anchors.refresh\_time=nil, 14  
    worker.count, 22  
    worker.id, 22  
    worker.pid, 22  
event.after (C function), 19  
event.cancel (C function), 20  
event.recurrent (C function), 20  
event.reschedule (C function), 20

event.socket (C function), 20

**H**

hints.add\_hosts (C function), 28  
hints.config (C function), 28  
hints.del (C function), 29  
hints.get (C function), 28  
hints.root (C function), 29  
hints.root\_file (C function), 29  
hints.set (C function), 29  
hints.use\_nodata (C function), 30  
hostname (C function), 7

**K**

KEY\_COVERING\_RRSIG (C function), 90  
KEY\_FLAG\_RANK (C function), 90  
knot\_dname\_lf2wire (C function), 91  
kr\_bitcmp (C function), 90  
kr\_cache\_clear (C function), 77  
kr\_cache\_close (C function), 77  
kr\_cache\_closest\_apex (C function), 78  
kr\_cache\_insert\_rr (C function), 77  
kr\_cache\_is\_open (C function), 77  
kr\_cache\_make\_checkpoint (C function), 77  
kr\_cache\_match (C function), 78  
kr\_cache\_materialize (C function), 77  
kr\_cache\_open (C function), 76  
kr\_cache\_peek\_exact (C function), 77  
kr\_cache\_remove (C function), 77  
kr\_cache\_remove\_subtree (C function), 78  
kr\_cache\_sync (C function), 77  
kr\_cache\_ttl (C function), 77  
kr\_dname\_lf (C function), 91  
kr\_dname\_text (C function), 91  
kr\_error (C function), 93  
kr\_family\_len (C function), 90  
kr\_inaddr (C function), 89  
kr\_inaddr\_family (C function), 89  
kr\_inaddr\_len (C function), 89

kr\_inaddr\_port (C function), 90  
kr\_inaddr\_set\_port (C function), 90  
kr\_inaddr\_str (C function), 90  
kr\_memreserve (C function), 89  
kr\_module\_call (C function), 91  
kr\_module\_embedded (C function), 87  
kr\_module\_load (C function), 87  
kr\_module\_unload (C function), 87  
kr\_now (C function), 91  
kr\_nsrep\_copy\_set (C function), 82  
kr\_nsrep\_elect (C function), 81  
kr\_nsrep\_elect\_addr (C function), 81  
kr\_nsrep\_set (C function), 81  
kr\_nsrep\_sort (C function), 82  
kr\_nsrep\_update\_rep (C function), 82  
kr\_nsrep\_update\_rtt (C function), 82  
kr\_pkt\_clear\_payload (C function), 89  
kr\_pkt\_make\_auth\_header (C function), 89  
kr\_pkt\_put (C function), 89  
kr\_pkt\_qclass (C function), 92  
kr\_pkt\_qtype (C function), 92  
kr\_pkt\_recycle (C function), 89  
kr\_pkt\_text (C function), 91  
KR\_PRINTF (C function), 89  
kr\_qflags\_clear (C function), 71  
kr\_qflags\_set (C function), 71  
kr\_rand\_reseed (C function), 89  
kr\_rand\_uint (C function), 89  
kr\_rank\_check (C function), 67  
kr\_rank\_set (C function), 67  
kr\_rank\_test (C function), 67  
kr\_ranked\_rrarray\_add (C function), 91  
kr\_ranked\_rrarray\_set\_wire (C function), 91  
kr\_resolve\_begin (C function), 67  
kr\_resolve\_checkout (C function), 68  
kr\_resolve\_consume (C function), 68  
kr\_resolve\_finish (C function), 69  
kr\_resolve\_plan (C function), 69  
kr\_resolve\_pool (C function), 69  
kr\_resolve\_produce (C function), 68  
kr\_rplan\_deinit (C function), 72  
kr\_rplan\_empty (C function), 72  
kr\_rplan\_find\_resolved (C function), 73  
kr\_rplan\_init (C function), 71  
kr\_rplan\_last (C function), 73  
kr\_rplan\_pop (C function), 72  
kr\_rplan\_push (C function), 72  
kr\_rplan\_push\_empty (C function), 72  
kr\_rplan\_resolved (C function), 73  
kr\_rplan\_satisfies (C function), 73  
kr\_rrkey (C function), 91  
kr\_rrset\_init (C function), 92  
kr\_rrset\_text (C function), 91  
kr\_rrset\_type\_maysig (C function), 91

kr\_rrsig\_sig\_expiration (C function), 92  
kr\_rrsig\_sig\_inception (C function), 92  
kr\_rrsig\_type\_covered (C function), 92  
kr\_rrtype\_text (C function), 91  
kr\_sockaddr\_cmp (C function), 90  
kr\_sockaddr\_len (C function), 90  
kr\_straddr (C function), 91  
kr\_straddr\_family (C function), 90  
kr\_straddr\_join (C function), 90  
kr\_straddr\_socket (C function), 90  
kr\_straddr\_split (C function), 90  
kr\_straddr\_subnet (C function), 90  
kr\_strcatdup (C function), 89  
kr\_unpack\_cache\_key (C function), 78  
kr\_verbose\_set (C function), 89  
kr\_zonecut\_add (C function), 84  
kr\_zonecut\_copy (C function), 84  
kr\_zonecut\_copy\_trust (C function), 84  
kr\_zonecut\_deinit (C function), 83  
kr\_zonecut\_del (C function), 84  
kr\_zonecut\_del\_all (C function), 84  
kr\_zonecut\_find (C function), 85  
kr\_zonecut\_find\_cached (C function), 85  
kr\_zonecut\_init (C function), 83  
kr\_zonecut\_is\_empty (C function), 85  
kr\_zonecut\_set (C function), 83  
kr\_zonecut\_set\_sbelt (C function), 85

## L

lru\_t (C function), 81  
lua\_push\_printf (C function), 91

## M

map (C function), 21  
map\_clear (C function), 98  
map\_contains (C function), 97  
map\_del (C function), 98  
map\_get (C function), 98  
map\_make (C function), 97  
map\_set (C function), 98  
map\_walk\_prefixed (C function), 98  
mode (C function), 8  
moduledir (C function), 8  
modules.list (C function), 15  
modules.load (C function), 15  
modules.unload (C function), 15

## N

net.bufsize (C function), 12  
net.close (C function), 11  
net.interfaces (C function), 12  
net.list (C function), 11  
net.listen (C function), 11  
net.outgoing\_v4 (C function), 12



net.tcp\_pipeline (C function), 12  
 net.tls (C function), 13  
 net.tls\_padding (C function), 13  
 net.tls\_sticket\_secret (C function), 13  
 net.tls\_sticket\_secret\_file (C function), 13

## P

pack\_clone (C function), 101  
 pack\_last (C function), 101  
 pack\_obj\_del (C function), 101  
 pack\_obj\_find (C function), 101  
 pack\_obj\_len (C function), 101  
 pack\_obj\_next (C function), 101  
 pack\_obj\_push (C function), 101  
 pack\_obj\_val (C function), 101  
 package\_version (C function), 10  
 policy.add (C function), 35  
 policy.del (C function), 35  
 policy.rpz (C function), 36  
 policy.suffix\_common (C function), 36  
 policy.todnames (C function), 36  
 predict.config (C function), 38

## R

reorder\_RR (C function), 8  
 resolve (C function), 9

### RFC

RFC 1035, 52  
 RFC 3986, 16  
 RFC 5011, 14, 22, 23, 50  
 RFC 5077, 13  
 RFC 6147, 48  
 RFC 6761, 36  
 RFC 6761#section-6, 28  
 RFC 7646, 22  
 RFC 7706, 51  
 RFC 7858, 33  
 RFC 8109, 51  
 RFC 8145#section-5, 50  
 RFC 8198, 51

round\_power (C function), 103

## S

stats.clear\_frequent (C function), 31  
 stats.frequent (C function), 31  
 stats.get (C function), 31  
 stats.list (C function), 31  
 stats.set (C function), 31  
 stats.upstreams (C function), 31

## T

time\_diff (C function), 89  
 trie\_apply (C function), 104

trie\_clear (C function), 104  
 trie\_create (C function), 104  
 trie\_del (C function), 105  
 trie\_del\_first (C function), 105  
 trie\_free (C function), 104  
 trie\_get\_first (C function), 104  
 trie\_get\_ins (C function), 104  
 trie\_get\_leq (C function), 104  
 trie\_get\_try (C function), 104  
 trie\_it\_begin (C function), 105  
 trie\_it\_finished (C function), 105  
 trie\_it\_free (C function), 105  
 trie\_it\_key (C function), 105  
 trie\_it\_next (C function), 105  
 trie\_it\_val (C function), 105  
 trie\_weight (C function), 104  
 trust\_anchors.add (C function), 14  
 trust\_anchors.add\_file (C function), 13  
 trust\_anchors.config (C function), 13  
 trust\_anchors.set\_insecure (C function), 14

## U

user (C function), 8

## V

verbose (C function), 8  
 view:addr (C function), 37  
 view:tsig (C function), 37

## W

worker.coroutine (C function), 21  
 worker.sleep (C function), 21  
 worker.stats (C function), 22