Home About EN 奇安信 X 实验室

DNS

Deep Dive into NXDOMAIN Data in China

baijinghua, zhangzaifeng, suqitian, heziqian, ljj, rootkiter, daji

2024年1月29日 • 32 min read

Domain not found here

The Domain Name System (DNS) is an essential protocol in the architecture of today's Internet. It routinely translates domain names into IP addresses and also often handles a multitude of invalid queries. These include requests for non-existent domain names, termed NXDOMAIN. A high volume of such invalid queries can adversely affect online user experience, add strain to network operators, and in certain situations, might even pose a risk to the stability of DNS resolvers at different levels.

Qi An Xin is a Chinese cybersecurity company and one of the major DNS infrastructure service providers in China. It operates the largest public DNS resolution system in the country as well as the largest public PassiveDNS system. This endows us with a broad and clear perspective, enabling nationwide observation of DNS performance and facilitating comprehensive analysis and interpretation of DNS operations. This article leverages the massive recursive resolver data from Qi An Xin's public DNS to conduct an in-depth analysis starting on a particular dataset — NXDOMAIN data. The aim is to deeply understand the domestic DNS system's operational state and thereby enhance the efficiency and security of the entire DNS resolution system.

The article is divided into three parts: The first part compares the similarities and differences in erroneous domain name queries between China and the global DNS system and explores the reasons for these differences. The second part, from the perspective of domain names, comprehensively analyzes the causes and impacts of erroneous domain names. The third part, from a regional perspective, compares the differences in error responses among different provinces in China and their causes. The goal is to more accurately understand and enhance the operation efficiency and security of the domestic DNS system.

> *NXDOMAIN (Non-Existent Domain, rcode=3) is a specific type of response in the Domain Name System, indicating that the queried domain name does not exist in the DNS database. Among the various types of DNS error responses, NXDOMAIN is the most common, accounting for over 90% of all DNS error response types.*

1. Comparison of Root Servers and Recursive Resolver

Before starting the analysis, let's look at the top server in the DNS hierarchy, the root server, in terms of NXDOMAIN responses. For illustration purposes, we will use the L-root server as an example.

1.1 ICANN IHTI Data

The International Corporation for Assigned Names and Numbers (ICANN) has analysed queries for the L-root server to generate the Identifier Technology Health Indicators (ITHI) report. The graph below displays the L-root server metrics from January 2018 to December 2023 from that report. The darker part of the graph indicates the percentage of NXDOMAIN data, which fluctuates between 50% and 80%. Geoff Huston analysed the reasons for the fluctuations in the root server's NXDOMAIN ratio by comparing the trends of Chromoid on the APNIC blog in July 2023. As of December 2023, the percentage of queries answered by NXDOMAIN was 55.27%.

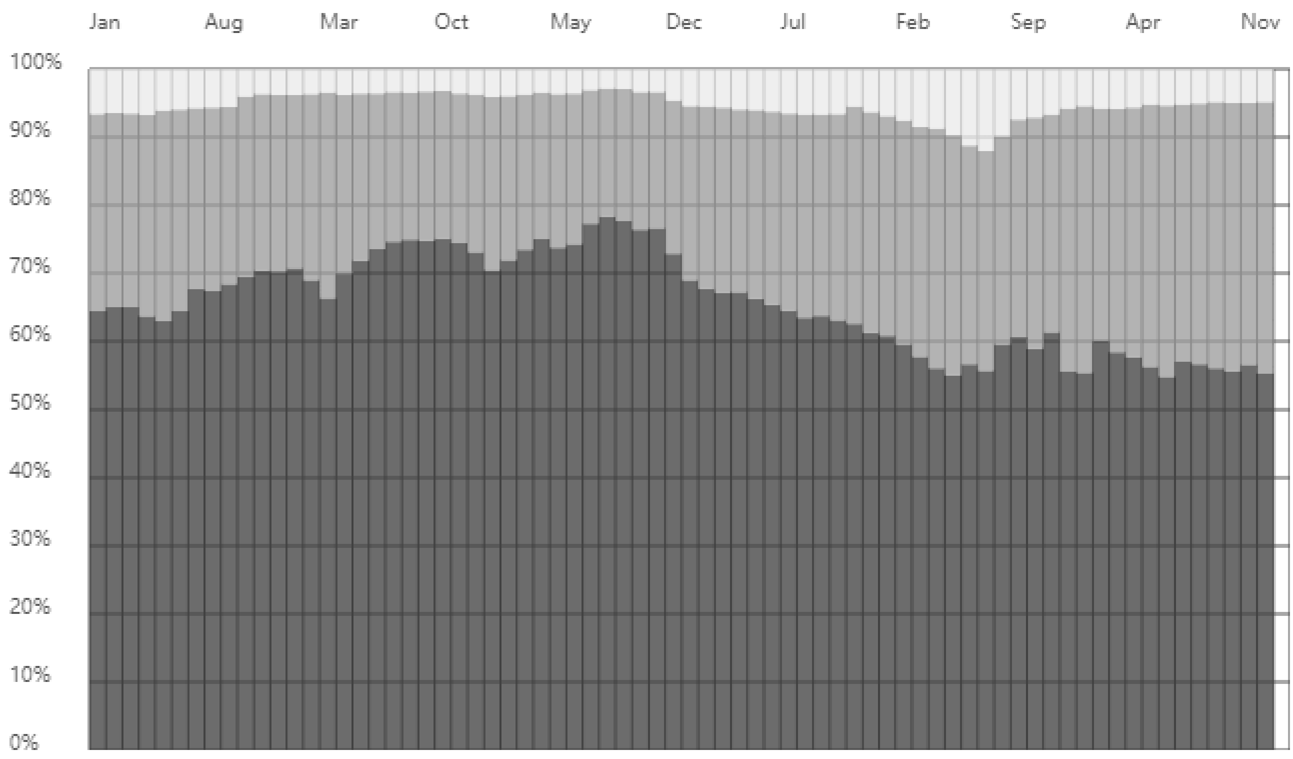

Figure 1 — NXDOMAIN ratio from L-root service. Source.

The table below (Source) shows the top 20 Top-Level Domains (TLDs) in L root NXDOMAIN queries.

RANK	TLD	% OF ALL QUERIES
1	local	7.292%
2	test	1.944%
3	home	1.163%
4	dhcp	1.074%
5	lan	0.999%
6	internal	0.956%
7	ctc	0.778%
8	bbrouter	0.699%
9	arpa	0.692%
10	localhost	0.538%
11	localdomain	0.522%
12	wifi	0.475%

RANK	TLD	% OF ALL QUERIES
13	invalid	0.336%
14	corp	0.336%
15	jpg	0.268%
16	svc	0.238%
17	k8s	0.215%
18	getcacheddhcpresultsforcurrentconfig	0.205%
19	m4a	0.200%
20	openstacklocal	0.193%

1.2 Recursive Resolver Data

The figure below shows the variation of the NXDOMAIN response rate by the Qianxin recursive resolver from July 2021 to December 2023, stabilizing at approximately 14% throughout 2023. Compared to the L-root, it is a quarter of that.This difference primarily stems from their distinct roles and positions within the Domain Name System: Recursive resolvers are at the very bottom layer of the Domain Name System, serving end-users directly, and often encounter repetitive queries. These servers handle requests for various domain levels and optimize resolution speed by caching results. On the other hand, root servers are at the topmost layer, not oriented towards end-users, with relatively fewer repetitive queries. They are mainly responsible for TLD queries, which frequently involve currently unresolvable TLDs, leading to a higher proportion of NXDOMAIN responses.

Figure 2 — NXDOMAIN ratio from QAX recursive resolver service.

The table below shows the list of the top 20 TLDs, which shows a significantly skewed distribution. The `.arpa` domain alone accounts for 30% (a large number of private IP PTR reverse queries, see 3.2). In addition, some commonly used valid TLDs such as `.com` and `.cn` also appear in the list. This is because the NXDOMAIN responses from recursive resolvers are for non-existent domain names, not non-existent TLDs. `.com` has a large number of registrations, and `.cn` (China's ccTLD) is associated with our data collection location.

RANK	TLD	TLD TYPE	% OF ALL QUERIES	% OF NXDOMAIN QUERIES	% OF
1	arpa	ICANN	4.1667%	30.3167%	6.761
2	com	ICANN	3.7407%	27.2168%	33.19
3	cn	ICANN	0.9854%	7.1698%	4.220
4	net	ICANN	0.7240%	5.2681%	4.764
5	org	ICANN	0.6386%	4.6464%	2.035
6	ctc		0.4406%	3.2057%	4.435
7	lan		0.2126%	1.5470%	7.880
8	ru	ICANN	0.1744%	1.2691%	0.199
9	wifi		0.1185%	0.8622%	1.040
10	cnp		0.1115%	0.8114%	0.000
11	localdomain		0.0936%	0.6811%	0.145
12	cc	ICANN	0.0921%	0.6701%	0.151
13	local		0.0820%	0.5968%	2.611
14	xyz	ICANN	0.0754%	0.5487%	0.604
15	top	ICANN	0.0599%	0.4359%	0.388
16	dhcp		0.0561%	0.4082%	0.326
17	3132372e302e302e31		0.0548%	0.3984%	0.000
18	novalocal		0.0545%	0.3962%	0.408
19	rl=http		0.0520%	0.3783%	0.000
20	eu	ICANN	0.0456%	0.3317%	0.055

1.3 Responses from non-ICANN systems

To make analysis easier, we have divided the NXDOMAIN data into two categories based on the TLDs of the query names: queries that belong to the ICANN system and queries that do not. Of the total NXDOMAIN data, the non-ICANN system (i.e. non-standard domain name queries) account for 19%. Their occurrence is due to a variety of reasons, including system or network equipment configuration defects, application bugs, leftover data from test and sandbox environments, user errors and unauthorized tampering with inputs, as well as other unknown reasons leading to invalid access. Attentive readers of this article will notice that some TLDs involve several major manufacturers. We have contacted some of these involved manufacturers through various channels, and they have all responded positively and made improvements. Specific names of the manufacturers are not listed in this article.

Figure 3 — NXDOMAIN Responsategorization by Top-Level Domain.

Below is a list of the top 20 TLDs that do not belong to the ICANN system.

RANK	NON-ICANN TLD	% OF NON-ICANN NXDOMAIN QUERIES	DAILY QUERIES QUANTITY SCA
1	ctc	16.73%	1B+
2	lan	8.07%	1B+
3	wifi	4.50%	100M+
4	cnp	4.23%	100M+
5	localdomain	3.55%	100M+
6	local	3.11%	100M+
7	dhcp	2.13%	100M+
8	3132372e302e302e31	2.08%	100M+
9	novalocal	2.07%	100M+
10	rl=http	1.97%	100M+
11	comp	1.50%	100M+
12	openstacklocal	1.46%	100M+
13	0	1.38%	100M+
14	localhost	1.33%	100M+
15	home	1.13%	100M+
16	***-wlan-controller	1.00%	100M+

RANK	NON-ICANN TLD	% OF NON-ICANN NXDOMAIN QUERIES	DAILY QUERIES QUANTITY SCA
17	url	0.74%	100M+
18	br-lan	0.62%	100M+
19	bbrouter	0.55%	100M+
20	null	0.49%	10M+

Compared to the L-root server, we noticed that many TLDs overlap in the two lists, but their rankings are significantly different. We selected a few TLDs and analysed the possible reasons for their appearance.

(1) Reserved special-use domain names

`.local` and `.localhost` are officially reserved for special purposes. The `.local` domain is specifically reserved for multicast DNS (mDNS) environments. It's used as a unique domain for hostnames in local area networks and can be resolved using the multicast DNS name resolution protocol. However, misconfigurations and other issues might cause these queries to be mistakenly sent to public DNS servers. The `.localhost` domain is reserved for loopback functions, where `.localhost` refers to the current computer accessing the domain.

(2) Common LAN suffixes

`.lan`, `.wifi`, `.localdomain`, `.dhcp`, `.home`, and `.bbrouter` are typically used as default DNS suffixes for internal networks. Their use depends on specific network setups, device types, or organizational preferences.

(3) Cloud environment suffixes

`.openstacklocal` and `.novalocal` are used as specific suffixes in cloud computing environments. `.openstacklocal` is used within private clouds or internal networks managed by the OpenStack cloud computing platform, identifying virtual machines and services. `.novalocal` is particularly associated with OpenStack's Nova compute component and is used for domain names automatically assigned to virtual machines created on OpenStack.

(4) Operator suffix

In the list above, `.ctc` has the highest percentage of non-ICANN TLDs. Statistics show that 92% of the client IP addresses accessing `.ctc` are from a specific operator in China. The default login address for the customized routers of this operator also ends with ctc, leading to the conclusion that the `.ctc` is primarily used by internal networks and services related to the devices of this operator.

Figure 4 — Proportion of .ctc requests from different client sources.

(5) Configuration errors

`.cnp`, `.3132372e302e302e31` (which decodes to `127.0.0.1`, the local loopback address), `.***-wlan-controller`, and `.br-lan` are examples of specific domain names that generate a lot of NXDOMAIN response queries due to network or application setup errors.

- **.cnp**

The `cnp` issue was caused by a misconfiguration of several subdomains of a manufacturer's cloud service, where an extra 'p' was mistakenly added to the correct subdomains, resulting in a large number of NXDOMAIN responses.

.CNP FQDN	% OF .CNP QUERIES	DAILY QUERIES QUANTITY
n-relay-ipc-txc-nj-00.***cloud.com.cnp	62.58%	100M+
n-txc-relay-ipc-nj-01.***cloud.com.cnp	24.13%	100M+
n-txc-relay-ipc-nj-00.***cloud.com.cnp	12.23%	10M+
txc-transmit-ipc-nj.***clouds.com.cnp	0.63%	1M+
n-txc-relay-tumscloud.***cloud.com.cnp	0.40%	1M+
txc-relay-ipc.***cloud.com.cnp	0.01%	10k+

> ***Latest update**: The manufacturer has successfully replicated the issue in their environment and is actively engaged in diagnostic analysis.*

- ***-wlan-controller

`***-wlan-controller`, which refers to a specific brand's wireless LAN controller. From the DNS data, it is observed that this string is being queried extensively as a complete domain name, rather than as a TLD of other domain names.

In the configuration of routers from a specific brand, `***-wlan-controller` is typically used as a domain prefix to aid wireless access points (APs) in dynamically discovering and connecting to the Wireless LAN Controller (AC) via DNS. For instance, in the network environment of this router brand, if the DHCP server's configured AC domain name is ac.example.com, an AP, upon receiving this domain name, automatically prefixes it with the `***-wlan-controller` string, transforming it into `***-wlan-controller.ac.example.com` for DNS resolution. This process helps the AP to determine the IP address of the AC. However, our DNS data indicates that this string is being queried extensively on its own, leading to a significant number of erroneous responses. This suggests the possibility of configuration errors or other issues, resulting in APs being unable to locate the correct AC, which could potentially affect network connectivity and stability.

(6) Mobile application errors

`.asia11`, `.com11`, `.cn11`, `.eu11`, `.asianull`, `.comnull`, `.cnnull`, `.eunull` are primarily due to a large number of query requests generated by an error in a particular application on a specific brand of smartphone. A detailed analysis is as follows,

Our monitoring has revealed that TLDs such as `.asia11`, `.com11`, `.asianull`, and `.comnull` follow a similar pattern. They are all formed by combining common TLDs like `asia` and `com` with suffixes `11` or `null`. The response frequencies for these combinations are also very close, with requests including the `11` suffix accounting for 0.35% and those with the `null` suffix accounting for over 0.14%. The total number of domain requests related to these TLDs reaches up to several hundred million per day, involving client IP numbers in the tens of millions.

RANK	NON-ICANN TLD	% NON-ICANN NXDOMAIN QUERIES	DAILY QUERIES QUANTITY SCALE	DAILY
25	asia11	0.35%	10M+	<1k
26	com11	0.35%	10M+	<1k
27	cn11	0.35%	10M+	<1k
28	eu11	0.35%	10M+	<1k
57	comnull	0.15%	10M+	1k+
58	cnnull	0.14%	10M+	<1k
59	eunull	0.14%	10M+	<1k
61	asianull	0.14%	10M+	<1k

This phenomenon is associated with a specific smartphone manufacturer. The global routing service domain names for this company are as follows:

Figure 5 — global router service domain.

In the process of devices from this manufacturer (smartphones) requesting these official domain names, due to some code bugs, the correct domain names are erroneously appended with suffixes like '11' or 'null'. For example, the original request for `grs.***cloud.com` is altered to `grs.***cloud.com11`. This leads to a large number of NXDOMAIN responses, resulting in abnormally high request volumes for TLDs such as `.asia11`, `.com11`, `.asianull`, and `.comnull`.

The graph below shows that requests for these incorrectly formatted domain names from this manufacturer began to appear as early as June 2022. Domains with the same error suffixes show consistent request frequencies. Domains ending with the '11' suffix (`grs.***cloud.asia11`, `grs.***cloud.cn11`, `grs.***cloud.com11`, `grs.***cloud.eu11`) are represented in the chart by overlapping purple lines, and their request volumes surged starting from July 22, 2022. Similarly, domains ending with the 'null' suffix (`grs.***cloud.asianull`, `grs.***cloud.cnnull`, `grs.***cloud.comnull`, `grs.***cloud.eunull`) are depicted with overlapping blue lines, and their request volumes saw a significant increase beginning November 17, 2022.

Figure 6 — Request counts from QAX recursive resolver service.

Figure 7 — Request counts from QAX recursive resolver service.

We pinpointed the trigger for these types of queries through testing: using this brand of smartphone, when swiping down from the top menu, domain name queries for the routing service with '11' or 'null' suffixes are sent out. The different suffixes are associated with its version.

Figure 8 — Test results of *** smartphone.

Additionally, we have noted that the domain `grs.***cloud.eu` is already in an NXDOMAIN state, yet it still exists in the official support list. In either case, the massive number of this brand devices (smartphones) are generating invalid DNS queries. We recommend that this manufacturer updates its system to correct this issue.

```
; <<>> DiG 9.11.4-P2-RedHat-9.11.4-26.P2.el7_9.13 <<>> grs.***cloud.eu
;; global options: +cmd
;; Got answer:
;; ->>HEADER<<- opcode: QUERY, status: NXDOMAIN, id: 59655
;; flags: qr rd ra; QUERY: 1, ANSWER: 0, AUTHORITY: 1, ADDITIONAL: 1

;; OPT PSEUDOSECTION:
; EDNS: version: 0, flags:; udp: 512
;; QUESTION SECTION:
;grs.***cloud.eu.             IN      A

;; AUTHORITY SECTION:
***cloud.eu.          8       IN      SOA     monkey.dnspod.net. enterprise3dnsadmi

;; Query time: 21 msec
;; SERVER: 10.46.36.5#53(10.46.36.5)
;; WHEN: Wed Dec 13 17:55:59 CST 2023
;; MSG SIZE  rcvd: 129
```

> ***Latest update**: In response to this phenomenon, we provided feedback to the emergency response team of the company, and received a positive response. The conclusions of the analysis were communicated promptly. According to the latest*

> *feedback, the team has identified that the specific cause is due to an application configuration issue. Additionally, the domain status of* `grs.***cloud.eu` *was restored to normal shortly after the feedback was given.*

1.4 Others

We also noticed a small number of queries with TLDs from the 'Opennic,' 'Tor', and 'Namecoin' namespaces. Generally, these TLDs don't leak into the public DNS resolution system. Their appearance may be due to erroneous configurations (such as some test sandboxes) or incorrect usage (like non-Tor browsers attempting to connect to Tor domain names).

TAG	TLD	% OF NON-ICANN NXDOMAIN QUERIES	DAILY QUERIES QUANTITY SCALE	DAIL
OpenNIC	null	0.4914%	10M+	10k+
Tor	onion	0.0054%	1M+	1k+
OpenNIC	o	0.0014%	100k+	1k+
Namecoin	bit	0.0002%	10k+	10+
OpenNIC	oss	0.0002%	10k+	10+
OpenNIC	pirate	0.0001%	10k+	<10
OpenNIC	geek	0.0001%	10k+	10+
OpenNIC	libre	0.0001%	10k+	10+
OpenNIC	gopher	0.0001%	10k+	<10
OpenNIC	bbs	0.0000%	1k+	10+
OpenNIC	parody	0.0000%	1k+	<10
OpenNIC	dyn	0.0000%	1k+	10+
OpenNIC	indy	0.0000%	1k+	<10
OpenNIC	chan	0.0000%	<1k	100+
OpenNIC	oz	0.0000%	<1k	100+
OpenNIC	cyb	0.0000%	<1k	10+
OpenNIC	neo	0.0000%	<1k	10+

2. Prominent domain patterns in NXDOMAIN from QAX recursive resolver

As previously mentioned, there are various reasons for generating of NXDOMAIN domain names. Before proceeding with the analysis in this section, let's first look at one case.

2.1 One case

In September 2022, we received a lead from a root server operation team that they found a large number of DNS queries in the form of MAC addresses from China on their root server. After checking with our monitoring data, we found that there were indeed a large number of such queries in the DNS logs. Starting from August 19, these queries began flooding the DNS system, causing the daily volume of queries on QAX recursive resolver to rise to nearly 2 billion. Further analysis revealed that this anomaly was caused by a feature flaw in a a mobile app of a leading e-commerce platform, which inadvertently led users of the app to initiate DNS queries using the MAC addresses of their current network. The vast user base of the app resulted in a massive influx of these abnormal queries to recursive resolvers and root servers. After identifying the cause, we reported the issue to the relevant departments. The bug was fixed on September 18, and subsequently, these queries disappeared.

Figure 9 — MAC address request counts from QAX recursive resolver service.

Process of association analysis

Through data association analysis, it was discovered that along with DNS requests for domains related to MAC, there are also DNS requests for `dev`, `wlan0`, `lladdr`, `reachable`. Additionally, these DNS requests were found to be associated with the domain names of a leading e-commerce platform.

Figure 10 — Mac class domain DNS requests.

By searching these keywords, it was found that these terms appear when checking the status of Wi-Fi connection hotspots. In the Android system, one can view the real-time status of connected Wi-Fi hotspots using the command 'ip neigh show'.

Figure 11 — IP neigh show command details.

Through analysis utilizing these unique strings, it was found that almost all of the domain name requests were preceded or followed by domain names associated with a specific e-commerce platform, as illustrated in the provided diagram. After conducting tests and validations, it was confirmed that these requests were indeed related to this e-commerce platform. It appears that there was a programmatic flaw in the app of this e-commerce platform when processing the results of Wi-Fi hotspot connection checks. This flaw led to DNS queries being executed for all returned results (`dev`, `wlan0`, `lladdr`, `MAC addresses`), which in turn caused the root servers to receive a large number of DNS query requests in the format of MAC addresses.

Figure 12 — Dev, wlan0, lladdr request before and after details.

2.2 Domain name pattern analysis method

In the first part of our analysis, we focused on the proportion of NXDOMAIN responses and the distribution of TLDs, providing a macro perspective on the NXDOMAIN data. However, this analysis is limited to the highest level of domain name structure and has not delved into more specific domain levels. As the case of this top-level application error reveals, the large number of queries in the MAC address format highlights the need for a more detailed analysis at domain name level. This type of analysis is critical to identifying and understanding the anomalous network behaviour triggered by specific patterns in domain names.

In fact, DNS recursive resolvers also cache the queried domain names that receive NXDOMAIN responses (RFC 2308). When dealing with large numbers of requests for a single domain, servers can make efficient use of the caching mechanism. Once the domain name resolution result is cached, subsequent queries for that domain can be responded to quickly, thus avoiding repeated resolutions. However, when faced with a large number of queries for many different domain names, the efficiency of caching decreases significantly as the server needs to frequently update and maintain caches for a wider range of domain names.

In light of this, this section will conduct a more in-depth analysis from the perspective of domain names, focusing on identifying significant domain name patterns in NXDOMAIN responses to classify and recognize similar domain name forms in the network. This will enable us to more comprehensively understand the characteristics of various domain names in the network and their underlying causes. Such detailed analysis is not only helpful in revealing and understanding abnormal network behavior, but also plays a significant role in optimizing DNS server performance and maintaining overall network security.

Generally, when a large number of NXDOMAIN abnormal responses occur, the corresponding queried domain names often exhibit specific patterns in their structure. Therefore, we first generalize the Rname (FQDN) in NXDOMAIN responses to identify domain names with similar patterns; then we further summarise and generalise these patterns and investigate the reasons for their occurrence.

Specific method:

1. **Decomposing domain names into grams**: all NXDOMAIN domain names are split into segments based on special characters (such as '.', '_', '*', '&', '-') and these segments are referred to as 'grams'. For instance, `djhsa.example-inc.com` would be split into 'djhsa', 'example', 'inc', 'com'.
2. **Statistical analysis and filtering of key grams**: count the frequency of occurrences for all grams. Then, based on a preset threshold, select the

key grams, which are the keywords in NXDOMAIN domain names.

3. **Generalizing domain names into patterns**: generalize each domain name by retaining its key words and special characters. The other parts are generalized into patterns similar to '[a-z]{n}' or '[0-9]{n}' based on their character type and length. For example, `djhsa.example-inc.com` is generalized to `[a-z]{5}.example-inc.com`.
4. **Aggregating, observing, and classifying patterns**: aggregate all generalized patterns, then observe and analyze the frequently occurring ones. Finally, based on the analysis results, refine rules based on the characteristics of these patterns.

2.3 Source analysis of prominent domain name patterns

For almost five months, we continuously monitored and analysed hundreds of millions of unique NXDOMAIN domain names collected daily by Qi An Xin's recursive resolver. The analysis revealed that the reasons for generating large numbers of NXDOMAIN domain names fall into two main reasons: those caused by software applications and those caused by system configurations. Problems caused by applications include Chromoid, Ads, User Tracing, Black Grey App, Blacklist Service, etc.; problems caused by configuration include Device ID, Reverse DNS, FQDNs ending with search suffix, etc. The specific distribution of each type is shown below:

> *During our monitoring period, we noticed that the form of domain names remained relatively stable. Based on this observation, we selected data from a specific date, November 9, 2023, for detailed analysis and illustration. To enhance the readability and visual appeal of the patterns, we replaced common types in the patterns, such as IPv4 addresses, IPv6 addresses, Universally Unique Identifiers(UUIDs), and hash values, with simplified identifiers like 'IPV4', 'IPV6', 'UUID', and 'HASH'.*

Figure 12 — Prominent domain patterns in NXDOMAIN responses (unique by FQDN).

In the chart, the first layer represents all unique NXDOMAIN domain names in the QAX recursive resolver (deduplicated based on FQDN). The second layer displays the proportions of various categories of domain names we have tagged. The third layer then details the specific domain name patterns for different categories and their respective proportions. Below, we attempt to analyze and explain the reasons behind each type and their potential impacts:

(1) Chromoid

Chromoid refers to DNS queries with random letter strings between 7-15 characters in length, sent out by Google Chrome browser to test for network abnormalities like DNS hijacking. After version 87, Chrome modified this detection method, and currently, such types of requests have reduced, accounting for 1.59% of the total. For detailed information about Chromoid domain measurements, please refer to this excellent article by APNIC.

(2) Ads

Ads refer to subdomains related to advertising. The two main patterns are: `UUID-xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.afdback.ptqy.gitv.tv` `UUID-xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.afdback.ppsimg.com`

The term 'afdback' stands for advertisement feedback, indicating that these subdomains are related to the transmission of advertising-related data. The two domains, `gitv.tv` and `ppsimg.com`, are associated with internet television services. The presence of 'ptqy' points to a specific video application on smart internet TVs. Upon analysis, it's evident that these uniquely structured domain names are generated by the television application to acquire information about the DNS server currently used by the user.

As shown in the diagram below, this is the network traffic during the operation of its APK:

Figure 14 — Requests and responses for the afdback domain.

In the first returned JSON string, the 'dns_url' field contains a domain name in a special format, namely `UUID-`

`xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.afdback.ppsimg.com` , which is exactly what we observed in the DNS data. Each request generates a distinct UUID, bypassing the DNS cache, thereby enabling ppsimg.com's authoritative server to receive all corresponding DNS requests. Consequently, this server is able to discern the DNS server details associated with each UUID, effectively identifying the source of each DNS request.

Subsequently, the client application sends an HTTP request with the same UUID (as shown in the second GET in the diagram above), allowing the web server to obtain the IP address corresponding to the user's UUID. Using the UUID as a link, it's possible to correlate the user's external IP with the DNS IP they are using, as shown in the information from the second response.

```
{
  "version": "2020-08-19 v0.0.4",
  "ID": "15f765ac-ebdbd888433e-0abd1306-xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx",
  "SN": "0abd1306",
  "USERIP": "35.221.56.251",
  "USER_AREA": "OVERSEA|US",
  "USERDNSIP": "74.125.18.2",
  "USERDNS_AREA": "GOOGLE |Public_DNS"
}
```

From the above analysis, it is evident that the numerous one-time, non-existent domain names related to the television application's advertisements observed on recursive resolver servers are not intended for actual network access. Instead, their purpose is to collect the IP addresses of the DNS servers currently used in the user's environment through DNS queries. This is done in order to optimize the network performance for accessing their services based on the user's location.

(3) User tracing

User tracing refers to domain names associated with user tracking. The main patterns include `UUID.clo.footprintdns.com` and `v17-[A-Z0-9]{8}.z.irs01.com` , among others. These are primarily related to user visit footprints and online tracking. By querying these tracking-related subdomains, data such as users' web browsing patterns, frequency of visits, and sources of access can be analyzed. This information is used for behavioral analysis, targeted advertising, and other purposes.

(4) Black grey app

Black grey app refers to domain names associated with the black and grey industry. The number of domain names and client sources in this category show a high degree of consistency in their patterns.

Figure 15 — Domain patterns and request details.

If one directly accesses certain second-level domain names within these patterns, such as `608.vip`, it automatically redirects to the download page of a gambling application. The emergence of so many NXDOMAIN subdomains with similar patterns suggests that it might be a case of industry sabotage, using random subdomains to attack the domain names of competitors.

Figure 16 — The webpage of 608.vip.

(5) Blacklist service

Blacklist service refers to subdomains related to blacklist services. From our patterns, we identified various filters used by different service providers for screening spam content. These services include:

- DNSBL (DNS-based Blackhole List) services, such as `IPV4.bl.spamcop.net` and `IPV4.zen.spamhaus.org`, which perform reverse DNS queries on sender IP addresses to determine if they are listed on spam sender blacklists.
- URIBL (Uniform Resource Identifier Blacklist) services, like `[a-z0-9]{8}.uribl.spameatingmonkey.net`, which block spam based on email content (typically domain names or websites).

- HASHBL (Hash-based Blackhole List) services, such as `[a-z0-9]{8}.ebl.msbl.org`, which store blocklist data as one-way encrypted hashes to decide whether to accept, filter, or reject incoming emails.
- ......

These patterns' domain names are specifically used for identifying and blocking spam emails, phishing, and other malicious online activities. Each domain represents a specific service or database, which checks if sender IP addresses, mail servers, website URIs, etc., are marked or have adverse records in these blacklists or threat intelligence sources. This helps to assess the credibility and security threat level of senders, websites, or links, thereby filtering out spam content and preventing attacks.

(6) Device ID

Device ID refers to subdomains related to the device ID, with the primary pattern being `[0-9]{10}.HASH.device.hiwifi.com`. 'HIWIFI' is a wireless router, and these types of subdomains may be used to differentiate and manage devices connected to this model of wireless router. They could also be used for device identity verification or communication for specific purposes.

(7) Reverse DNS

Reverse DNS refers to subdomains used for reverse domain name resolution, with patterns such as `IPV4.in-addr.arpa` and `IPV6.ip6.arpa`. Reverse Domain Name Resolution is a special DNS service that uses PTR records (pointer records) to convert IP addresses back to their corresponding domain names. Simply put, it does the opposite of standard DNS resolution: where standard DNS translates domain names into IP addresses, reverse DNS identifies the corresponding domain names from the IP addresses.

(8) Reserved name: .local

Reserved name: .local represents all domain names with the top-level domain name of `.local`. The domain name `local` is a special domain reserved specifically for multicast DNS environments, which is different from the usage of public DNS (unicast DNS). Here, we categorize it separately with the aim of conducting an in-depth analysis and focusing on its domain name patterns within public DNS, as well as analyzing the reasons for its leakage into public DNS.

The diagram below provides a detailed representation of the distribution of all domain names with `.local` as the TLD in the QAX recursive resolver. The first layer in the diagram shows all domain names with `.local` as the TLD. The second layer displays the proportion of different patterns within these domain names. The third layer then presents the ratio of sub-patterns, along with an example domain name of each sub-pattern.

Figure 16 — .local domain patterns in NXDOMAIN (unique by FQDN).

From the above chart, it is evident that among all domain names with `.local` as the TLD, `UUID.local` accounts for 94%. This indicates that the majority of domain names with the `.local` suffix follow the `UUID.local` format.

The following chart illustrates the development trend of the `UUID.local` pattern domain names. Although we cannot pinpoint the exact time of their initial appearance, the change in their quantity over time can be observed through our public DNS data.

Figure 18 — Trend of UUID.local

In the chart, the red bar graph represents the number of requests, reaching the scale of hundreds of millions per day, corresponding to the vertical axis on the left side of the chart. The green bar graph shows the number of domain names following this pattern, with a scale in the tens of millions per day, also corresponding to the left vertical axis. Meanwhile, the blue curve represents the number of clients initiating requests, with the scale in the millions per day, corresponding to the vertical axis on the right side of the chart.

Upon deeper analysis, it was found that the `UUID.local` pattern emerges as part of a solution to prevent privacy leaks in WebRTC. The widespread occurrence of this pattern on the internet may be associated with a bug in a certain open-source component or compatibility issues with different versions of WebRTC.

The emergence of UUID.local

WebRTC enables web browsers and mobile applications to add real-time audio and video peer-to-peer connections and communication capabilities. This technology allows users to access platforms like Bilibili, Douyin, Kuaishou directly within browsers and mobile apps, utilizing its peer-to-peer connection features for interactive activities such as live streaming connections. However, WebRTC carries the risk of privacy leakage, as it can expose the user's real IP addresses, both public and private.

To address the issue of private IP leakage, there is an IETF draft proposal. Specifically, it involves hiding the IP address using dynamically generated mDNS names. When this solution is enabled, during real-time audio and video peer-to-peer connections and communication with WebRTC, it broadcasts its mDNS name, which is `UUID.local`, within the local network. Here, UUID represents a randomly generated unique identifier.

Figure 19 — Domain broadcasting of UUID.local mode in mDNS.

The destination address for mDNS requests is a fixed multicast address, 224.0.0.251. Normally, this type of traffic should not leak into the public DNS environment.

Analysis of UUID.local leakage

To further pinpoint and analyze the reasons behind the large volume of `UUID.local` occurrences in the public recursive network, we set up our own DNS traffic analysis system in the local network. This system is capable of analyzing DNS traffic from various network devices such as smartphones and PCs in a multi-dimensional manner.

Considering the application scenarios of WebRTC, we hypothesized that it might be related to video websites or the currently popular live streaming platforms. After testing several popular video applications, including Bilibili, Kuaishou, and Douyin, we discovered that the live streaming module of one of these applications, specifically the Android version, triggers DNS requests for `UUID.local`.

From the image below, we can see that the domain names in the requests occurring before the DNS request for `UUID.local` are all related to this application. Particularly, the parts related to STUN indicate the use of a protocol for NAT traversal to establish peer-to-peer communication, which is a crucial technology supporting the WebRTC protocol.

Figure 20 — Domain queries of UUID.local mode in DNS.

Combining DNS traffic analysis with the analysis of the app, we speculate that there are three main possible reasons:

1. **Configuration Error**: When using WebRTC in the live streaming module, this Android app does not correctly construct multicast DNS request addresses (i.e., 224.0.0.251) and continues to use regular DNS addresses, leading to leakage onto the public internet.
2. **Open Source Component Error**: This Android app's live streaming module uses an open-source implementation of WebRTC, which might have certain vulnerabilities, leading to situations where multicast DNS requests are leaked onto the public internet.
3. **Compatibility Issue**: Differences in WebRTC versions between the communicating parties could result in mDNS being announced to the broader network in a unicast form.

Figure 21 — Details of the IETF Draft

It is currently not possible to further pinpoint the specific reason. However, based on the distribution pattern of the `.local`, the latter two possibilities seem more likely. After all, other domain names ending with `.local` are relatively rare on public DNS, aside from the `UUID.local` format. Additionally, data published by the L-root server shows that `.local` accounts for the largest proportion of its erroneous responses. In the case of error requests on 1.1.1.1, `.local` also has the highest share, which may imply that a similar issue exists in global WebRTC usage.

> ***Latest Update***: *We have established contact with the security team of the video application in question. In response to the issues we raised, they conducted an in-depth analysis and have resolved the problem in the latest version of their app. However, it is worth noting that the observed volume of* `UUID.local` *queries on the larger internet has not significantly decreased. This suggests that the issue is not confined to a single company but is more likely a widespread issue caused by a commonly used open-source component.*

(9) FQDNs end with search suffix

FQDNs that end with a search suffix are domain names appended with a search domain suffix. When a user-queried domain name does not exist in DNS, the system automatically tries to append the configured search domain suffix to the domain name and initiates another DNS request. The settings for these search suffixes are typically associated with specific network environments, operating systems, or device types. Structurally, there are no specific restrictions on the search domain suffix, it can be any level of domain name, including top-level domains, second-level domains, third-level domains, etc.

The following list displays some common search domain suffixes found in the pattern:

NO.	SUFFIX
1	smartont.net
2	ctc
3	wifi
4	lan
5	dhcp
6	home
7	huawei.com
8	router.ctc
9	airdream
10	bbrouter
11	tendawifi.com
12	realtek
13	openstacklocal
14	mshome.net
15	wowifi.smartont.net

We will specifically explain the top 5 search domain suffixes, and the Sankey diagram below provides a detailed illustration of the distribution of these search domains. In the diagram, the second layer shows the proportion of domain names for each search domain, while the third layer displays the subcategories within each domain, and the fourth layer shows their specific domain name patterns.

Figure 22 — FQDNs end with search suffix in NXDOMAIN.

We can observe that:

1. The chart primarily showcases the detailed distribution of five search domain suffixes. `.lan` is a commonly used default suffix in local area networks, `.smartont.net` and `.ctc` are related to network operators, `.wifi` is frequently seen in WiFi devices, and `.dhcp` is common in dynamically allocated network environments.
2. In different search domains, many domain name patterns exhibit similarities. For instance, `Chromoid(.suffix)` represents Chromoid domain names with a specific search domain suffix. A similar pattern is also observed in other domain names like `Ads(.suffix)` and `UUID.local(.suffix)`. In this pattern, domain names from the same source form a unique structure by adding their respective specific search domain suffixes. This phenomenon indicates that certain types of domain name patterns are commonly found across search domains.
3. Some domain name patterns appear exclusively within specific search domains. For example, `Device ID(.suffix)` only appears in the `.lan` search domain. This might reflect an intrinsic association between these domain names and specific search domains. Specifically, the presence of `Device ID(.suffix)` solely within the `.lan` search

domain suggests that the default search domain suffix for this router (hiwifi.com) might be `.lan`.

(10) Others

We have completed the attribution analysis for the previous nine category patterns. The table below displays the patterns that are not included in these categories. These unidentified patterns represent the focus of our subsequent attention.

PATTERN	DAILY FQDNS QUANTITY SCALE	FQDN EXAMPLE
[a-z0-9]{8}	1M+	e23c753e
[a-z0-9]{8}.test	1M+	001eab65.test
[a-zA-Z0-9]{17}.mdbook.cn	1M+	50d3d5c8fdb0d5a70f1c4c65d.
[a-z0-9]{4}.kaiyun.com	1M+	6u3b.kaiyun.com
&ver=[0-9]{1}&id=[A-Z0-9]{11}	1M+	&ver=1&id=1782071404875474
[a-z0-9]{24}.jpg	100k+	gkwrimaiw418aaaknwjr6oxv.jpg
[a-z0-9]{4}.baidu.com	100k+	rjd6.baidu.com
[a-z0-9]{4}.spd-inv.net	100k+	kd0i.spd-inv.net
[a-zA-Z0-9]{16}.mdbook.cn	100k+	a0ea7053ca2143937c3d4c65.
[a-z0-9]{4}.baidu2.com	100k+	s4ef.baidu2.com
[a-z0-9]{4}.baidu3.com	100k+	f8lt.baidu3.com
[a-z]{4}.kaiyun.com	100k+	fnud.kaiyun.com
IPV4	100k+	192-168-201-167
inst-[a-z0-9]{29}-[a-z0-9]{16}	100k+	inst-zef83nzhzf96vbz5eccmcc
[a-z]{4}.baidu.com	100k+	lfjr.baidu.com
[a-z0-9]{9}.dyn.telefonica.de	100k+	x4db29bb2.dyn.telefonica.de
[a-z]{4}.spd-inv.net	100k+	iltg.spd-inv.net
[0-9]{7}.b2b.jinanfa.cn	100k+	2233615.b2b.jinanfa.cn
[a-z]{4}.baidu2.com	100k+	lxwi.baidu2.com
[a-z0-9]{11}	100k+	47636zdl709

2.4 Automated Monitoring

To achieve automated tracking of NXDOMAIN response data, we monitor error response data (yes, covering the main types of error rcodes) and have implemented monitoring of various types of errors and anomalies using the method introduced in Section 2.2. Specifically:

1. The NXDOMAIN response data is monitored by dividing the corresponding query domain names into identified patterns and unidentified patterns. Identified patterns refer to domain name patterns that we have observed in historical data and have tagged with specific labels. Unidentified patterns, on the other hand, are either newly emerged or previously unnoticed (generally, patterns with smaller volumes are not noticed).
2. Whether it's identified or unidentified patterns, any significant increase in the error response data they represent is worth analyzing to determine the cause. This is especially true for unidentified patterns, where it's crucial to delve deeper to see if the increase is due to a surge in requests for a newly emerged pattern category. Subsequently, situations where the cause can be determined should be addressed with alerts and appropriate actions.

Using this method, we can achieve timely detection of errors in DNS data, such as incidents where an e-commerce platform inadvertently sends out user MAC addresses.

3. Differences in NXDOMAIN Across Regions

3.1 NXDOMAIN Response Situations in Various Provinces

Due to the variance in the volume of DNS query requests collected from different provinces, directly comparing the absolute number of NXDOMAIN queries does not accurately reflect the real situation. To fairly compare the proportion of NXDOMAIN queries across provinces, we use 'the percentage of NXDOMAIN queries in each province to the total number of queries in that province' as an indicator. By calculating the proportion of NXDOMAIN queries in each province, the impact of different data volumes of DNS queries across provinces can be eliminated, more accurately reflecting the actual situation of DNS resolution quality in each province. (Data from Hong Kong, Macau, and Taiwan regions of China are not included in this analysis due to their relatively small data volume and lack of representativeness.)

The following chart displays a time series graph of the proportion of NXDOMAIN queries in different provinces from August to December 2023. Different colored lines represent different provinces. As can be seen from Graph 1, the proportion of NXDOMAIN queries in most provinces remains within a fluctuation range of (14±5)%. However, it is noteworthy that Jiangxi Province has the highest proportion of NXDOMAIN queries, about 24%, which is significantly different compared to other provinces.

Figure 23 — NXDOMAIN response rates within each province from QAX recursive resolver service.

The detailed proportions for a specific day in the above chart are as follows.

Figure 24 — NXDOMAIN response rates within each province from QAX recursive resolver service.

3.2 NXDOMAIN Analysis in Jiangxi Province

Using the domain name analysis method from Section 2, we classified the NXDOMAIN query domain names originating from Jiangxi Province. It was found that the number of reverse queries, similar to `IPV4.in-addr.arpa`, is exceptionally high, accounting for over 60% of all NXDOMAIN queries in Jiangxi Province.

Figure 25 — Proportion of reverse DNS in NXDOMAIN responses in Jiangxi Province.

Reverse Query Domain Names

For all NXDOMAIN data from Jiangxi Province, we reverse the prefixes of the reverse query domain names to extract the corresponding IP addresses, referred to as query IPs. Upon analyzing these query IPs, it was found that they are predominantly private network addresses like 192.168, accounting for 93% of the total reverse queries. This indicates that these reverse queries mainly originate from internal or local area network requests. When we applied the same method to analyze national data, the proportion of reverse queries for private network addresses also reached 56%, indicating that this situation is not unique to Jiangxi Province.

Figure 26 — Proportion of reverse DNS queries from Private IP as source IP in NXDOMAIN queries.

From an analysis of the client sources initiating these query requests, it was observed that 92% of the requests originated from IPs belonging to a specific telecommunications operator. Further analysis revealed that the majority of these IPs were associated with the operator's home broadband services, amounting to hundreds of thousands in number.

Figure 27 — Proportion of reverse DNS queries by different client sources in NXDOMAIN queries.

From the above analysis, it is clear that the high proportion of NXDOMAIN in Jiangxi Province is primarily due to a large number of reverse DNS queries initiated by internal network IPs, with the querying clients predominantly being home broadband IPs from a specific telecommunications operator. Extending the timeline, we should investigate when this phenomenon began.

The following chart shows the DNS query statistics on the 1st of each month in Jiangxi Province for the years 2022 and 2023.

Figure 28 — NXDOMAIN response rate within Jiangxi Province.

The figure presents two time-series curves: the blue line represents the percentage of NXDOMAIN DNS queries in the province as a proportion of all DNS queries within the province; the red line represents the percentage of reverse DNS queries that received an NXDOMAIN response as a proportion of all DNS queries in the province. Observing the changes in the red line, it is evident that before June 2022, the proportion of reverse DNS queries was very low, consistently below 5%. In July 2022, there was a sharp rise to 21%, followed by a slight decrease but maintaining around 10% (as of March 1, 2023). From February 2023 onwards, the proportion of reverse queries stabilized at around 16%. This indicates that around July 2022, there was a significant change in reverse DNS queries, with the proportion substantially increasing and continuing to remain at a higher level.

Based on the temporal changes in the NXDOMAIN response ratio and previous cases [2][3],it can be speculated that the unusually high percentage of NXDOMAIN response requests in Jiangxi Province may be due to improper configuration of certain network equipment (such as routers or wireless devices) provided by the telecom operator in that region.

> ***Latest Update***: *Monitoring has revealed that after December 27, 2023, the proportion of NXDOMAIN responses in Jiangxi Province has decreased to around 14%, returning to normal levels.*

Conclusion

1. On our recursive resolution server, the proportion of NXDOMAIN responses is around 14%, of which 81% of the NXDOMAIN responses are from domain name queries within the ICANN system, primarily in the .arpa and .com domains. Queries outside the ICANN system account for 19%, with .ctc having the highest proportion.
2. By comparing the most frequently queried non-ICANN top-level domains (invalid TLDs) on both recursive resolvers and root servers, it was found that there are many overlaps. However, the rankings of these TLDs in terms of query volume show significant differences, primarily because the geographical area from which we collect data is relatively concentrated.
3. The reasons for DNS queries resulting in NXDOMAIN responses can be divided into two main categories: those caused by software applications and those caused by system configurations. Application-related issues include blacklist services, dark grey industries, Chromoid, user tracking, advertising tracking, etc., while configuration-related issues include device ID queries, private IP reverse queries, etc.
4. In public DNS, reserved domain names like .local primarily appear in the pattern of `UUID.local`. Analysis reveals that `UUID.local` domain names are used to conceal internal network IP addresses when employing WebRTC technology, and they are broadcast only in multicast DNS environments. The leakage of these domain names to public DNS may be associated with a bug in a specific open-source library.
5. Geographically, the proportion of NXDOMAIN responses in Jiangxi Province is 23%, significantly higher than in other provinces. Analysis indicates that over 60% of the erroneous queries come from PTR queries

for private IPs by clients of a certain telecommunications operator, which may be due to the configuration of local network devices.

6. We have developed an automated monitoring system that enables timely detection of anomalies in error responses from various perspectives, including response ratio, TLD distribution, pattern analysis, and regional clients. Implementing such monitoring on a dataset with such a wide user coverage helps us gain a clearer understanding of the state of DNS network operations. This is significant not only for individuals, businesses, and operators, but also for government regulatory agencies.

What do you think?

0 Responses

👍 Upvote 😝 Funny 😍 Love 😯 Surprised 😤 Angry 😥 Sad

0 Comments 1 Login ▾

G Start the discussion…

LOG IN WITH

OR SIGN UP WITH DISQUS ?

Name

♡ Share **Best** Newest Oldest

RSS