Internet Engineering Task Force
Internet-Draft
Intended status: Experimental
Expires: December 27, 2020

M. Sivaraman
Akira Systems Private Limited
Liu
Infoblox

June 25, 2020

The DNS thundering herd problem
draft-muks-dnsop-dns-thundering-herd-00

Abstract

This document describes an observed regular pattern of spikes in queries that affects caching resolvers, and recommends software mitigations for it.

Status of This Memo

Copyright Notice

Table of Contents

1. Problem Description

Typically, DNS caching resolvers prepare answers for multiple clients from a single cached RRset [RFC1034]. Depending on when in time the clients make their queries, caching resolvers reply with lower and lower valued TTLs, before the cached RRset from which answers are prepared expires. Clients themselves may cache and use their copies of RRsets until the TTL that the resolver replied with expires. A key property is that all these copies of answers, and the cached answer from which they are prepared, expire at the same absolute time.

As an example, consider the following query sequence received by a resolver from 10 clients all querying for a popular www.example.org./A RRset. We use this example to illustrate two kinds of spikes in queries.

Client	Query time (seconds since epoch)	Answer RRset TTL	Notes
C1	1591441620	600	Answer was not found in cache. Resolver performs a resolution and caches authoritative answer with TTL=600.
C2	1591441626	594	Answered from cache.
C3	1591441713	507	Answered from cache.
C4	1591441780	440	Answered from cache.
C5	1591441866	354	Answered from cache.
C6	1591442006	214	Answered from cache.
C7	1591442070	150	Answered from cache.
C8	1591442070	150	Answered from cache.
C9	1591442213	7	Answered from cache.
C3	1591442220	600	Previously cached answer had expired in the resolver's cache. So the resolver performs a fresh resolution and caches authoritative answer with TTL=600.
C5	1591442220	600	Ditto if not joined with previous.
C2	1591442220	600	Ditto if not joined with previous.
C6	1591442220	600	Ditto if not joined with previous.
C1	1591442221	599	Answered from cache.
C9	1591442221	599	Answered from cache.
C4	1591442221	599	Answered from cache.
C8	1591442221	599	Answered from cache.
C7	1591442221	599	Answered from cache.
C10	1591442227	593	Answered from cache.
C7	1591442820	600	Previously cached answer had expired in the resolver's cache. So the resolver performs a fresh resolution and caches authoritative answer with TTL=600.
C4	1591442820	600	Ditto if not joined with previous.
C1	1591442820	600	Ditto if not joined with previous.
C2	1591442820	600	Ditto if not joined with previous.
C10	1591442820	600	Ditto if not joined with previous.
C8	1591442820	600	Ditto if not joined with previous.
C3	1591442821	599	Answered from cache.
C9	1591442821	599	Answered from cache.
C5	1591442821	599	Answered from cache.
C6	1591442821	599	Answered from cache.

2. Requirements Notation

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. Mitigations

3.1. Combine identical queries to upstream nameservers

At a resolver, when multiple queries have arrived together asking the same question and there is no existing unexpired cached answer, DNS resolutions have to be performed to answer these queries. De-duplication of these multiple resolutions into a single DNS resolution by the resolver is RECOMMENDED where possible.

If such de-duplication is not performed, the client queries will effectively be forwarded 1:1 by the resolver to upstream nameservers, and they will significantly increase the upstream nameservers' query rate in spikes. Some nameserver operators may have deployed measures such as response rate limiting [RRL] and other IP-address based rate limiting, which may cause them to deny service to the resolver due to the query spikes of identical queries.

3.2. Include noise in response TTLs from caching resolvers

Caching resolvers are permitted to lower the TTLs of RRsets in their answers as they please [RFC2181]. This can be used to distribute the time at which RRset copies received by clients expire from a single absolute time to a time interval. However, this has to be done with some consideration such that the thundering herd doesn't re-converge at the expiry time of the cached RRset that is used to generate answers to the clients.

TBD.

3.3. Other mitigations

With very low authoritative RRset TTLs (such as under 60s) for popular questions, the frequency of the thundering herd increases and including noise in response TTLs is less effective because the maximum TTL to work with is low. In other words, there is a shorter interval over which the thundering herd can be distributed by adding noise. Some implementations permit an operator to set a minimum TTL value such that authoritative RRset TTLs with lower values are increased and clamped to the minimum TTL value. This breaks

currently accepted DNS protocol, and hence this document does not make any recommendation about it.

4. Security Considerations

There are no security considerations.

5. IANA considerations

There are no IANA considerations.

6. Acknowledgements

This document was prepared from thundering herd client query patterns noticed at resolvers of ISPs and large institutions, which resulted in traffic spikes that caused performance issues and lookup failures. The authors acknowledge the contribution of Ramesh Damodaran who participated in analysis of these patterns.

7. References

7.1. Normative references

[RFC1034] Mockapetris, P., "Domain names - concepts and facilities", STD 13, RFC 1034, DOI 10.17487/RFC1034, November 1987, <https://www.rfc-editor.org/info/rfc1034>.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC2181] Elz, R. and R. Bush, "Clarifications to the DNS Specification", RFC 2181, DOI 10.17487/RFC2181, July 1997, <https://www.rfc-editor.org/info/rfc2181>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

7.2. Informative references

[RRL] Vixie, P. and V. Schryver, "DNS Response Rate Limiting (DNS RRL)", 2012, <https://ftp.isc.org/isc/pubs/tn/isc-tn-2012-1.txt>.

Appendix A. Change history (to be removed before publication)

- o draft-muks-dnsop-dns-thundering-herd-00
  * Initial draft.

Authors' Addresses

Mukund Sivaraman
Akira Systems Private Limited
1 Coleman Street, #05-05 The Adelphi
Singapore 179803
SG

Email: muks@akira.org
URI: https://akira.org/

Cricket Liu
Infoblox
3111 Coronado Drive
Santa Clara 95054
US

Email: cricket@infoblox.com
URI: http://www.infoblox.com/