NORMAN RESEARCH

# Content Distribution Networking and Privacy Discrimination

Or

How I stopped worrying and learned to love the transparency of slow customized POPs

**Authors:** Norman Luhrmann, V

daughter-of-v@protonmail.com / PGP key DF20D21B349C2A9A

**Revision:** 1.0 (2020-06-11) Initial release

## Executive summary

Modern web delivery platforms are dependent on Content Distribution Networking (CDN) to achieve optimal load/response times. This is essential for many kinds of online products due to the heavy impact site/app performance incurs on conversion goals (e.g. sales) [1].

Traditional CDN platforms are inherently intransparent since they potentially utilize advanced routing technologies between POPs (points of presence). This potentially contradicts the data locality requirements imposed by privacy regulations like GDPR or PRC Cybersecurity law. While vendors usually contractually guarantee privacy compliance, in a complex global scale network non-conforming data processing is difficult to rule out.

Additionally, CDN platforms often offer edge caching to further accelerate traffic delivery. These caches are subject to the same privacy regulations with specific requirements to storage locality. Dynamic edge population mechanisms often lack the policy control required to ensure that compliance is guaranteed. Again, due to the complexities of a global cache architecture it is difficult to ensure this requirement even if common control mechanisms like CDN/cache region filtering enable simple policy adaptation.

As a conclusion the authors see the requirement for better CDN standardisation (comparable to the ISO cloud initiative), acceptance of full contractual privacy risk transfer or the provisioning of fully-owned custom CDN infrastructure in consequence.

Disclaimer: not part of this paper are privacy concerns related to CDN-side security issues like the widely observed Cloudbleed vulnerability [2] or a data protection regulation assessment of the mentioned CDN impacts which might be negligible in specific applications of use case and law.

## CDN traffic shape

Figure 1 shows an exemplary CDN network utilizing modern optimization techniques like adaptive congestion avoidance and tiered caching to handle requests in a flexible fashion.

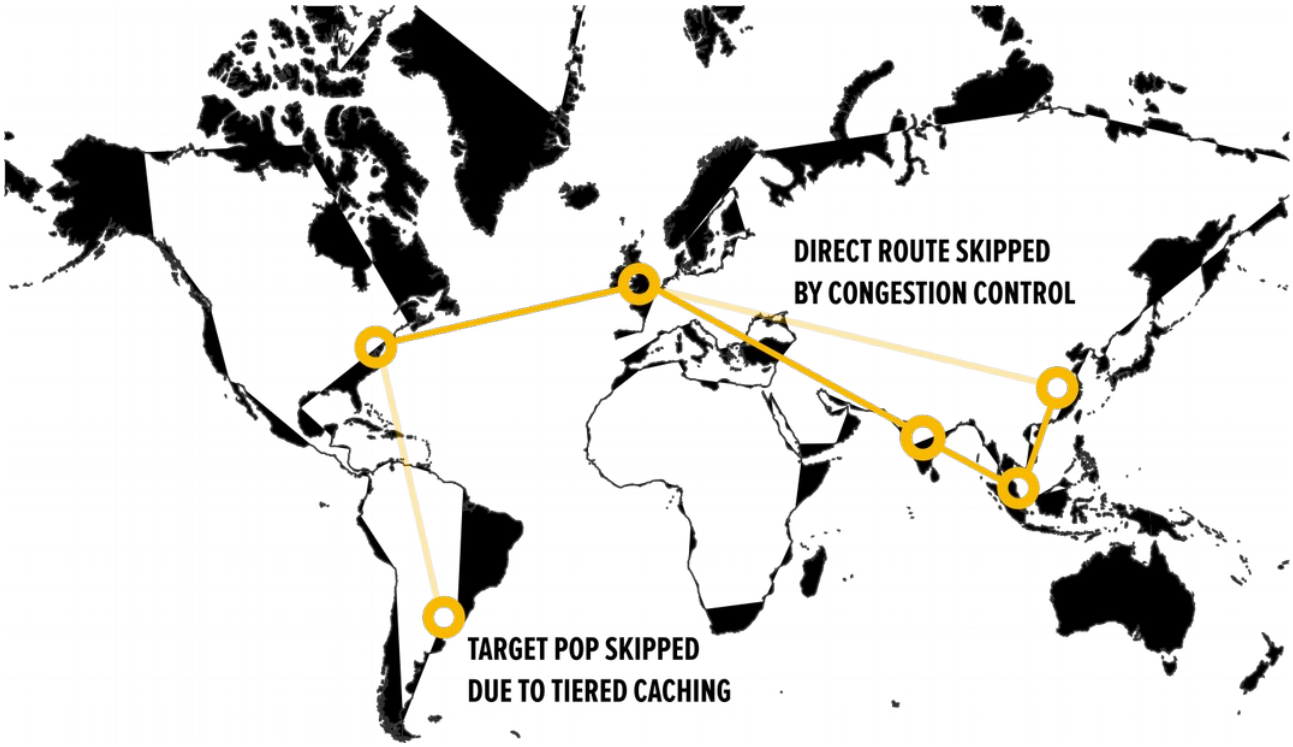


Fig. 1: Non-standard routing behavior of modern CDN

In the case of congestion control the network is configured to detect network bottlenecks in real-time and provide route optimization to avoid suboptimal paths. This can lead to network paths being traversed which could implicate regulatory issues depending on the locality of the alternative route hops.

The second case shows a request that is served by an intermediate hop due to tiered caching having populated enough data to satisfy the request without forwarding to the target POP. Again regulatory concerns might occur.

In general we identify three primary privacy-related route cases:

1. A request which satisfies source side policy demands is routed via proxy regions which are potentially incompatible.
2. A request which satisfies destination side policy demands is never routed into the destination.
3. An intra-policy-domain request which inherently satisfies source and destination side policy demands leaves the policy region.

An example for the last case is show in Fig. 2.


Fig. 2: Non-standard routing behavior for intra-region traffic

# Edge caching mechanisms

Example 2 in Fig. 1 has shown that aggressive caching allows to satisfy requests inside the CDN before reaching the final POP or destination host.

Tiered caching mechanisms further optimize the cache strategy to facilitate the huge amounts of data transmitted due to high traffic and/or the huge amount of cache space requirements due to high traffic variance.

In tiered caching some/all intra-route hops are potential cache locations for requests. This often happens unknowingly to CDN users who are concerned with source and destination locality and behavior.

Simple edge cache implementations should be considered safe from a privacy risk management perspective since they form an obvious component in system assessment that should obviously fulfil regulatory demands.

The tiered caching scenario needs more consideration and currently vendors seem to not fully apply policy mechanisms necessary for compliance.

# Privacy regulation requirements

Data protection legislative usually is concerned with the activities processing and storage of data.

In regards to CDNs the transport behavior needs to be aligned with processing requirements while the caching layer needs to be aligned with both processing and storage requirements.

Usually both processing and storage of data observe basic principles like data locality, user consent or security by design.

## Extended net neutrality definition

Traditionally net neutrality is concerned with the cost of operations and performance of the transport layer of the internet. The common (recently attacked) agreement is that the cost of moving data over infrastructure is deemed part of the owners responsibility and every data packet should be considered equal on public networks.

In the context of this paper we propose an extended net neutrality definition that also concerns itself with the transparency of transport layer activity. A network can no longer be considered neutral if it is opaque to analysis of traffic behaviour and intermediate processing impacts.

Privacy regulations require the possibility to assess data processing and storage activities all along the paths packets take between the user and service/application provider. In the experience of the authors CDNs do no longer inhibit this transparency, and sometimes require black box analysis to at least in part assess that compliance requirements are met (e.g. via latency experimentation).

## Conflict of interest

While it is in the best interest of both the user and the service/application provider to reach high quality and performance of delivery, the complexities of the current diverse international regulation landscape demand high efforts in decision making and assessment for the applicability of advanced CDN features like the ones mentioned above.

## Conclusion

Without a proper internationally accepted standard for content distribution (akin to ISO 35.210 Cloud Computing [3]) and/or better harmonisation of regulations (e.g. COPPA, GDPR, PRC Cybersecurity law, etc.) CDNs pose a potentially high operating risk in more complex deployments.

While regulations usually consider technically required storage/processing activities as safe to do, even then data locality, consent and similar topics are to be applied for proper privacy control.

The authors currently do not see a way to incorporate a strict interpretation of more extensive privacy regulations and modern CDN architectures without a high degree of interaction between CDN and application/service provider.

As it stands only contractually bound risk transfer or the provisioning of a custom CDN with fully-owned POPs provide means to rectify these inherent privacy risks.

## References

[1] https://www.cloudflare.com/learning/performance/more/website-performance-conversion-rates/

[2] https://bugs.chromium.org/p/project-zero/issues/detail?id=1139

[3] https://www.iso.org/ics/35.210/x/