Impact of Social networking sites on Local DNS server

Sandeep Tamrakar
Helsinki University of Technology
stamraka@cc.hut.fi

Abstract

The numbers of online social networking sites are growing rapidly. The social networking sites have redefined the way we interact online. Most of the social networking sties provide customizable personal pages to its members. During customization user may embed contents from different web sites that provide contents in a form of HTML embed codes. Thus a page may contain different contents from several different web sites. As a result a page download may generate hundreds of DNS queries and even if few people visit these social networking sites at the same time and if they are using same Internet service provider, the number of DNS queries sent to local DNS server is quite huge. In many cases the local DNS server are unable to handle such DNS traffic thereby it slows down all its services. Such overloaded DNS server is susceptible to Denial of Service attacks as well. In this paper I tried to mention some of available methods that social networking sites could implement to reduce such DNS references and also tried to mention some methods to improve the carrier network so that it can handle such huge DNS traffics.

KEYWORDS: Domain Name System (DNS), Local name server, Social networking sites, anycast, anycast DNS, improving local name server performance, DNS queries.

1 Introduction

The development of virtual communities like Usenet [11] and Whole Earth 'Lectronic Link (WELL) [12] in 80's was a milestone for the online social networking. Since then several online social networking communities has evolved and developed along with the growth of Internet. In recent years, online social networking has been widely adopted. There are several social networking sites which have millions of users and the membership charge is often free. Facebook [2], MySpace [8], Hi5 [6], Bebo [1], Flixster [3], Friendster [4], LinkedIn [7] are some of the examples of some popular social networking sites. It is estimated that there are currently at least 300 million peoples who uses these social networking sites and the number of members enrolling are still growing [29]. The reason behind this mass adoption is the way the Internet has developed allowing users with wide range of interactive web-base application with rich and varied media contents. People can now share information, music, pictures, discussion, and indeed just about anything that are in digital format, much easily and interactively.

According to Danah m. Boyd and Nicole B. Ellison [16], *social networking sites are web based services that allow users to construct a public profile within a bounded system, maintain a list of other users with whom they share a connection, and view and traverse their list of connections and also those made by others within the system*.

The most important features that social networking sites focus is a user's profile page. Typically, a user profile page contains information describing the user like physical appearance, hobbies, interests, photos, age, location, contact information, "about me"section and other information. These information are provided by the user during the enrollment to the social networking sites or later on. While joining an online social networking site, an individual is asked to fill out forms with lists of questions. The profile page is generated from the answers provided by users to these questions.

Most of the social networking sites such as MySpace, Hi5 allow users to enhance their profile page by modification of their profile's look and feels, and even allow embedding multimedia contents from various sites like YouTube [13], RockYou [10], Glitter-Graphics [5] etc. Other members can view the profile page, add text or multimedia comments on the profile page according to the privacy policy set by the user.

There are hundreds of sites that provide their multimedia contents in a form of HTML codes that can be embedded in other web pages. Users of social networking sites generally add such codes to their profile pages or as comments to other user's profile pages. For example, within a comment input box of a user, another member can add HTML embed code of a video from YouTube as a comment. Thus, a single profile page may contain numbers of pictures, graphics, videos, audio files from different sites. Though this allows users to customize their personal profile page, this feature unintentionally affects Domain Name System (DNS) of carrier and corporate networks [30].

DNS is a globally distributed database which consists of Domain names and their corresponding IP addresses. Primary purpose of DNS is mapping host names to IP address like *www.tkk.fi* to its corresponding IP address *130.233.240.9* [28]. DNS is a domain-based naming scheme in which domain names are maintained in a hierarchical tree structures. Networked computers use IP addresses to locate and connect to each other, but IP addresses is difficult for human use. For example, it is much easier to remember the domain name *www.tkk.fi* than to remember its corresponding IP address *103.233.230.9*. Thus, DNS allows people to connect to remote computers using user-friendly domain names. The mappings in the DNS name space are called resource records.

This paper talks about the impact of social networking sites on the local DNS server. It focuses more on the struc-

ture of the social networking sites pages which is the cause of the problem. Then talks about some solutions that social networking sites could implement to minimize this problem and also some of the methods that carrier network can implement to handle such problem.

The paper is organized as follows: Section 2 describes and mentions about the DNS queries generated by social networking sites pages and various factors that reduces the efficiency of DNS server. Section 3 describes some of the design structures that social networking sites could implement so that they can provide better user interaction while generating less DNS queries. The section 4 describe some methods that carrier and corporate networks could implement to handle such huge DNS traffic and the last section is conclusion.

2 Problem

As the number of broadband subscribers and usage per subscriber has increased, DNS traffic has also grown significantly. Such growths are generally expected and are planned by the Internet service provider during capacity planning. Sometimes viruses, worms and attacks on a network may cause sudden increase in DNS traffic but such huge DNS traffics are temporary [30] which usually lasts of couple of hours. However, the popularity of social networking sites like MySpace, Hi5 etc cause exponential growth on DNS traffic. The growth of such DNS traffic is due to the design structure of such networking sites' pages.

A typical profile page of such social networking sites usually contains personal information, images and comments posted by users and other members of the networking sites. Moreover, these social networking sites allows embedding HTML codes, user may embed videos, images, flash-based contents and MP3 audio files as well. In most of the cases, these additional contents are from other web services which basically provide some HTML embed codes for example YouTube, RockYou etc. Thus, in order to display a profile page a web browser may require number of DNS lookups for all different contents from different websites. More the number of content added in a profile page from different web services, the more the number of DNS queries has to be generated.

Further strain on DNS traffic is added by Content Delivery Networks. Social networking sites uses content delivery networks to provide their content in an optimized way [18, 30]. Content Delivery Network is an overlay network of nodes deployed in different geographical location which caches contents in nodes to provide fast and reliable web services to the client. Content delivery providers like Limelight Network [25] which provides content delivery service to Facebook, use their globally distributed nodes to deliver contents to the end user in a most optimized way. Such providers caches their customer's rich media content in different nodes within the overlay network and then redirects the DNS request from the user to the optimal source for content. The selection of the optimal source could be based on various factors such as performance, load, cost, proximity and available information.

According to study by George Pallis and Athena Vakali [26], most of the popular CDN providers make use of *Un-*

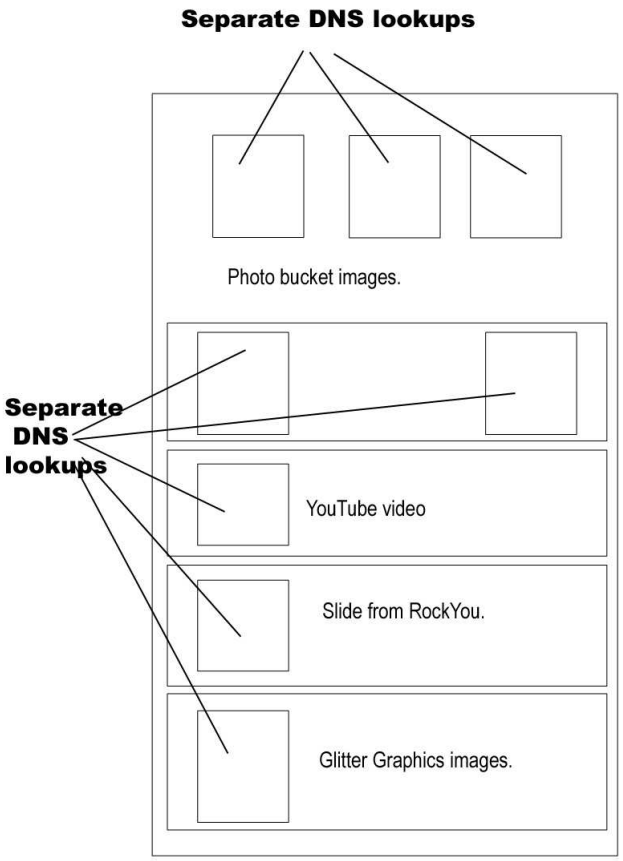

Figure 1: A typical profile page of social networking site

cooperative pull-based content outsourcing. In this type of content outsourcing, clients' requests are redirected to the closest node using DNS redirection. If the cache in that node expires then it again redirects the request to other node or to the original server.

As the content of the social networking sites are generated and maintained by the users of such networks, the content may change any time. In order to provide updated content in an optimized way, the CDN provider uses short Time to Live (TTL) values for DNS data of such caching. Thus, the caching name servers cannot maintain the data in the cache for long time to deliver the content for the requests made for the same page. As a result CDN redirects the request to another nodes or the original server which further adds strain on DNS [18, 26, 30].

To make it clearer - assume that twenty students from a university all want to check out the profile page of their friends to see updated pictures. And assume that they all use internet from the same internet provider or from the university network. Each page download would generate hundreds of DNS lookups. And due to the nature of CDN, the cache data have short TTL, thus the caching name server cannot make use of cached data and the DNS request must go to higher level server to get the right DNS data. Hence, social networking sites not only generate huge amount of DNS traffic, but also reduce the effectiveness of the DNS server's cache and slows the response time. In this way, it impact on the effectiveness of the recursive name server, respon-

sible for the domain name lookups in the DNS on behalf of the client's requests, of the carrier or the Internet service provider.

And as the carrier's DNS server has to perform huge amount of recursive queries on behalf of its clients, the server slows down affecting all the services. Even though the carrier provides higher bandwidth, it slows down the client response time as name resolution process takes longer time. This ultimately slows down the net browsing experiences of user. Moreover, such networks with overloaded DNS are more vulnerable to Denial of Service attacks [18, 30].

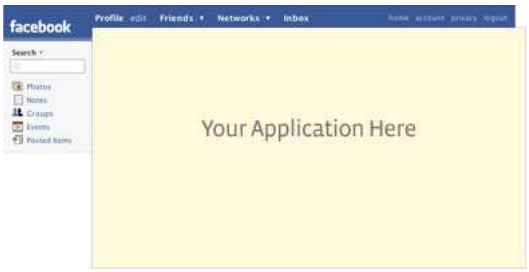

Figure 2: Facebook canvas with application template
source: www.facebook.com

3 Solution

In order to reduce the DNS queries for a single page download, the social networking site could store all the content within the social networking server itself. This way the social networking site can centralize all the data within itself. But in order to do so, it should not allow embedding of any HTML codes from different sites. All the content such as video files, pictures, audio file, text and graphics that user upload should be stored and maintained by the social networking sites. This may create some problem for both the social networking sites as well as users. The social networking sites may have to face copyright issues and privacy issues for contains uploaded by the user. Moreover, sites like YouTube, RockYou, do not provide any source file to download so users will not be able to upload the content from such sites. User will have to create their files themselves or may have to depend on the services their social networking sites provide. This reduces the user interaction and flexibility which are the basis of mass adoption of such social networking sites.

It is also noticeable that allowing HTML codes to embed may lead to security violation. Most of the users of social networking sites are not web designers, malformed Cascading Style Sheet may use full resources of web browser and lead to browser crash. Moreover, malicious JavaScript could be embedded in the page or may import malicious JavaScript from other URL which may cause security violation such as remote download and execution exploits [31].

Another solution could be that such social network sites resolve all the domain names that the page contains into their corresponding IP addresses prior to sending the requested page to the browser. The social networking sites should be able to maintain all the domain names and their corresponding IP addresses that are contained within different pages in their system. Prior to sending the requested page contents to the client browser it should resolve and provide IP addresses instead of domain name for the contents that are embedded in other web sites. In order word the social networking site should also be able to perform the role of DNS server. As the domain names are converted into IP address the client browser will not have to request for name resolution to the DNS server and can load content directly from the respective servers. Though this will create overload to the social networking server, client browser can fetch data directly from the IP addresses and reduces the DNS queries.

There are certain disadvantages to this method. One of the major issues is trust. Should we trust to the name resolution provided by Social networking sites? Is the name resolution process consistent with the resolution provided by authoritative name servers?

The social networking site then will have to keep records of all the domain names that user add to their corresponding pages. So each time a user makes changes in their pages the social networking sites should be able to update the domain names that was added or removed. The social networking site should also maintain an updated list of name resolution table. Any changes made in those domain names should be reflected in social networking site's DNS.

A convenient way of using contents from various web sites and also maintaining them is to provide a common interface through which, various web developers can create services that can be integrated with social networking sites. This way social networking sites can control on the content those users upload on their pages as well as they can keep record of those content providers.

Also social networking site can enhance their features and provide structured templates for social networking pages reducing clumsy contents that are arranged haphazardly. Social networking sites can enforce certain policies so that any change in DNS information of those content providers to be reported to social networking sites so that they can maintain an updated list of domain names. The social networking site will only need to store certain information regarding the content from different sites and user's contents will be stored and managed by corresponding content provider sites.

Facebook seems to be one of the social networking sites which provide an interface to developers. On May 24, 2007 Facebook lunched Facebook Platform providing a framework for software developers to create applications [2]. The framework includes their Application Programming Interface (API) based on (Representational State Transfer) REST [21]. Facebook also provides Query language called Facebook Query Language (FQL) and Facebook Markup Language (FBML). The web services that are developed using these tools can be integrated into Facebook platform in a form of Facebook applications.

Facebook API allows Facebook application to send HTTP GET or POST messages to the Facebook server over Internet. FQL lets application to use SQL- style queries in Facebook server. FBML allows Facebook application to hook into several Facebook integration points such as Profile, News Feed, Mini-Feed etc [2].

Each application that has been developed for Facebook

platform are assigned with a unique identifiers called *api key*. Facebook members can install these applications to their pages by accepting certain terms of services after which user is logged in to the application until user logs out from Facebook. When user logs in to the application, application maintains a session and receives certain information from the Facebook server. All the contents and data user create within these application will be stored in respective server of these applications where as the applications are hosted under *apps.facebook.com*.

This way, social networking sites not only enhance aesthetic of their pages but also can reduce DNS traffics as well as data traffic. The contents within the profile page or any other landing pages are in a form of web applications that are hosted in a particular sub domain. Since all those applications are affiliated with the social networking sites, the social networking sites can maintain the mapping of domain name and IP addresses for all those application servers. Any DNS related information update in application server should be updated in social networking site's DNS manager system. And thus, resolve the domain names into IP addresses prior to sending the requested page to client.

In some cases social networking sites can also store certain cover images or contents of those applications into social networking server itself. And, when user interacts with the application, the actual control is to be transferred to the corresponding server. This not only reduces the DNS traffic but also reduce the unnecessary data traffic as the page download doesn't have to load all the contents from various sites at once. The actual contents will only be loaded when user requests for the contents.

4 Solution by improving performance of carrier network

4.1 DNS queries

When a client needs an IP address for a specific domain name, it queries DNS server to resolve the name. A typically DNS query message contains a fully qualified domain name (FQDN), query type, and a specified class for the DNS domain name [19, 23].

For example: assume a host sends DNS query for *files.mynetwork.com* as in the Fig. 3. The local DNS server checks its records and finds the corresponding IP address then it response back with the IP address as 10.1.1.250.

Now, assume that the host requests name resolution for www.xyz.com, and the host is not in the DNS server's zone data. And also the local DNS doesn't have any records relating to the requested domain name. There are two ways that local DNS server can followed to resolve domain name into its corresponding IP address namely; recursive search or iterative search [19, 23].

When a DNS client issues a recursive query to a local name server, the server attempts to resolve the name completely with full answers or with an error by following the naming hierarchy all the way to the authoritative name server. The client requesting queries receive address information only from the local name server. The local name server requests other DNS servers in the hierarchy on behalf of the client. Whereas upon receiving an iterative query, the local name server first check DNS data within its record, if the local name server doesn't have DNS information within itself then it can simply give a referral to another name server for the client to contact next [20].

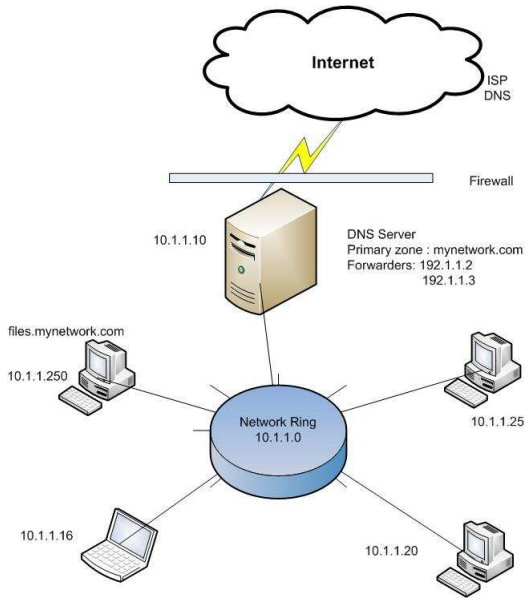

Figure 3: A local DNS server

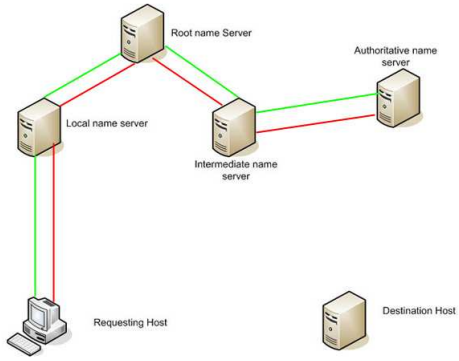

Figure 4: Local DNS server performing recursive DNS query

The processing of recursive query and iterative query is shown in the Fig. 4 and Fig. 5 respectively.The green line indicates the query request where as red line shows the response. In case of recursive query the client always get response from the local name server where as in iterative query the local server refer to other DNS server and the query is again sent to the other DNSs until the query gets resolved.

4.2 OpenDNS

OpenDNS, a DNS service provider advertises its free DNS resolution service as an alternative to using Internet service provider's DNS services. The company offers two recursive name server addresses 206.67.222.222

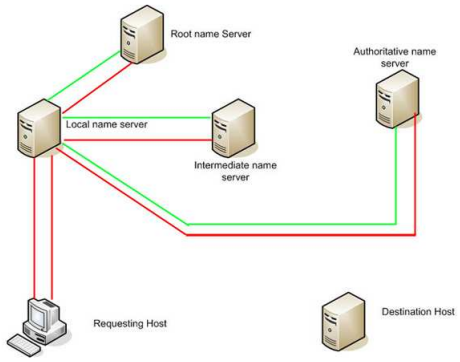

Figure 5: Local DNS server performing iterative DNS query

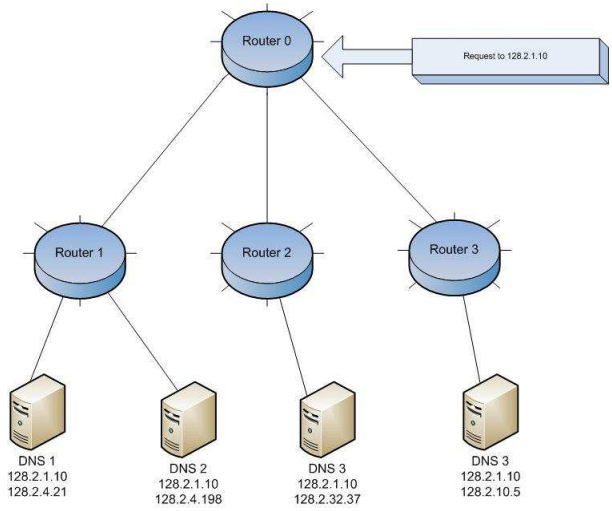

Figure 6: Four DNS server answer request to 128.2.1.10

(resolver1.opendns.com) and 208.67.220.220 (resolver2.opendns.com) for public use. OpenDNS has many name servers strategically placed at various geographical locations and uses Anycast routing technology to direct the DNS queries sent by client to the nearest operational server location. Changing preferred DNS server to high performance DNS server might be an easy solution when the preferred local DNS is not able to handle huge amount of DNS queries [9].

Changing the preferred DNS from local name server to OpenDNS may not be allowed within corporate environment. In most cases user do not have enough permission to change DNS settings. Moreover, if the change in DNS setting is not done correctly then the user will not be able to access internet at all. One of the big issues with Open DNS is that it breaks the hierarchical, distributed structure of DNS. OpenDNS funnels user's request through one centralized server making the user dependent on one company. Sharing DNS requests with third party involves trust and privacy concern as well. OpenDNS can monitor all the DNS queries made by user and also may covertly redirect client's request to different server by returning false results to DNS lookups [14].

4.3 Anycast DNS service

More than one DNS server can be configured using DNS resolver libraries but on most of the operating system the failover mechanism doesn't seem to be effective. According to the study by Kevin Miller [24], if primary configured DNS server fails to response the DNS query it waits until the timeout period expires, then the request is sent to the next server on the list and waits for response until the timeout period again. If all the servers in the list fail to response the requested DNS query, the system returns error information to the client. For the next DNS resolution request the system again starts with the first DNS server on the list regardless of the failure history. Thus, each DNS resolution request will be delayed at least equal to the response timeout from the primary DNS in case of failure of primary DNS.

In his study [24], he purposed 3 ways to improve network performance and increase the reliability of recursive DNS service; namely Anycast addressing, host-based router daemons, and unicast reverse path forwarding. With anycast more than one server are configured with the same anycast IP addresses that provide identical DNS services. Thus, anycast eliminates the dependency upon a single DNS server to resolve request there by increases the reliability of DNS service.

Anycast is an overlay network in which nodes are located at different locations. Each node provides an identical service. When a host requests for service by transmitting a datagram to an anycast address, the network is responsible for delivering the datagram to at least one, preferably the closest, of the servers in the anycast network [17]. Both in IPv4 and IPv6 an anycast datagram is similar to regular IP datagram, the destination address field in the datagram corresponds to an anycast address [27].

Anycast DNS provides DNS services to the same IP address from a number of different geographical or topological locations [22]. A node in DNS anycast is a set of one or more DNS servers with the same anycast IP address. Every DNS query sent to the anycast service IP address is routed precisely to one node, preferably the best node determined by the routing protocol in used [15]. The selection of best node depends upon performance, load, cost, proximity and other available information. It is necessary for a node of an anycast network to indicate to the router that they intend to receive anycast datagram for a specified address. This is achieved by advertising the anycast information to the router that they intent to receive datagram for a specified address using enhanced version of Internet Group Management Protocol (IGMP) [15].There are three method mentioned in the study about using IP anycast for load distribution [15]; each router maintains a regular updates with its neighbors about the reachable hosts. Another method mentioned is to have routing protocol implemented at the end host update the routers about the anycast information. In this case, the host only needs to ensure routers that the anycast address they server are reachable. The third and simple method is to configure the routers statically with the anycast information.

When multiple DNS servers are configured with anycast host-based routing daemon, the dependency upon single

host for name resolution is eliminated and the system's responsiveness to server and network failure recovery can be achieved much faster. In addition to that the reverse path forwarding method self maintains for source address verification which is an important part for protecting the Internet infrastructure [24].

Corporate and carrier network can use this method which is much reliable and efficient to handle huge DNS queries generated by social networking sites.

5 Conclusion

In this paper I have tried to describe those DNS resolution problem that local DNS server faces due to the structure of social networking site pages. I have also tried to suggest some of the methods that social networking sites can adopt. These methods not only reduce the DNS queries but also help define well structured design of their services. Finally I have tried to mention about anycast technique for enhancing the performance of private networks so that they can provide reliable DNS service to their clients. With this anycast DNS, the network can not only handle huge DNS traffic but also provide reliable and fast service to their clients.

References

[1] Bebo. http://www.bebo.com.

[2] Facebook. http://www.facebook.com.

[3] Flixster. http://www.flixster.com.

[4] Friendster. http://www.friendster.com.

[5] Glitter graphics. http://www.glitter-graphics.com.

[6] Hi5. http://www.hi5.com.

[7] Linkedin. http://www.linkedin.com.

[8] Myspace. http://www.myspace.com.

[9] Opendns. http://www.opendns.com.

[10] Rockyou. http://www.rockyou.com.

[11] Usenet. http://www.usenet.com.

[12] Well. http://www.well.com.

[13] Youtube. http://www.youtube.com.

[14] Prevent OpenDNS From Redirecting Google Searches - Fix for Firefox IE Address Bar. 2008. http://www.labnol.org/software/browsers/prevent-opendns-google-redirect%s-firefox-address-bar-ie/2662/.

[15] J. Abley. Hierarchical Anycast for Global Service Distribution. Technical note, ISC, Inc, 2003. http://www.isc.org/pubs/tn/isc-tn-2003-1.html.

[16] D. M. Boyd and N. B. Ellison. Social Network Sites: Defination, History, and Scholarship. Journal 11, Journal of Computer-Mediated Communication, 2007. http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html.

[17] W. M. C. Partridge, T. Mendez. Host Anycasting Service. RFC 1546, BBN, November 1993. http://ietf.org/rfc/rfc1546.txt.

[18] N. W. Carolyn Duffy Marsan. How MySpace is hurting your network. Article, 2007. http://www.networkworld.com/news/2007/062207-myspace.html.

[19] M. Corporation. How DNS query works. Technical manual, Microsoft Corporation, 2005. http://technet2.microsoft.com/windowsserver/en/library/0bcd97e6-b75d-48%ce-83ca-bf470573ebdc1033.mspx?mfr=true.

[20] M. Erik Eckel Network+, MCP+I. Know how to distinguish DNS query types. 2003. http://articles.techrepublic.com.com/5100-1035_11-1058014.html.

[21] R. T. Fielding. Architectural Styles and the Design of Network-based Software Architectures. article, University of California, Irvine, 2000. http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm.

[22] T. Hardie. Distributing Authoritative Name Servers via Shared Unicast Address. RFC 3258, Nominum Inc., April 2002. http://ietf.org/rfc/rfc3258.txt.

[23] IBM. Understanding DNS queries. Technical manual, IBM. http://publib.boulder.ibm.com/infocenter/iseries/v5r3/index.jsp?topic=/%rzakk/rzakkconceptquery.htm.

[24] K. Miller. Three Practical ways to Improve Your Network. Paper, Carnegie Mellon Universtiy, 2003. http://www.usenix.org/event/lisa03/tech/full_papers/miller/miller_html/%index.html.

[25] L. networks. Limeligth Networks Announces Extended Relationship with Facebook. 2003. http://www.limelightnetworks.com/press/2006/01_31_facebook.html.

[26] G. Pallis and A. Vakali. Insight and perspectives for content delivery networks. *Commun. ACM*, 49(1):101–106, 2006.

[27] V. P. Robert Engel and D. Saha. Using Ip Anycast for Load Distribution and Server Location. Technical note, IBM. http://www.zurich.ibm.com/~rha/papers/anycast-gi98.pdf.

[28] A. S. Tanenbaum. *Computer Networks*. Prentice-Hall Inc, 4th edition, 2003.

[29] R. Thomas. Q n a about the future of social networking. private communication, 2008. http://netucation.co.za/qa-about-the-future-of-social-networking/.

[30] T. Tovar. Social Networking and Web 2.0: Create DNS performance Issues for carriers. Article, Nominum Inc, 2007. http://www.convergedigest.com/bp/bp1.asp?ID=466.

[31] D. Tynan. The 25 Worst Web Sites. 2006. http://www.pcworld.com/article/id,127116-page,7-c,sites/article.html.