

REQUEST ROUTING IN A NETWORKED ENVIRONMENT

BACKGROUND

[0001] Generally described, computing devices and communication networks can be utilized to exchange information. In a common application, a computing device can request content from another computing device via the communication network. For example, a user at a personal computing device can utilize a software browser application to request a Web page from a server computing device via the Internet. In such embodiments, the user computing device can be referred to as a client computing device and the server computing device can be referred to as a content provider.

[0002] Content providers are generally motivated to provide requested content to client computing devices often with consideration of efficient transmission of the requested content to the client computing device and/or consideration of a cost associated with the transmission of the content. For larger scale implementations, a content provider may receive content requests from a high volume of client computing devices which can place a strain on the content provider's computing resources. Additionally, the content requested by the client computing devices may have a number of components, which can further place additional strain on the content provider's computing resources.

[0003] With reference to an illustrative example, a requested Web page, or original content, may be associated with a number of additional resources, such as images or videos, which are to be displayed with the Web page. In one specific embodiment, the additional resources of the Web page are identified by a number of embedded resource identifiers, such as uniform resource locators ("URLs"). In turn, software on the client computing devices typically processes embedded resource identifiers to generate requests for the content. Often, the resource identifiers associated with the embedded resources reference a computing device associated with the content provider such that the client computing device would transmit the request for the additional resources to the referenced content provider computing device. Accordingly, in order to satisfy a content request, the content provider(s) (or any service provider on behalf of the content provider(s)) would provide client computing devices data associated with the Web page and/or the data associated with the embedded resources.

[0004] Content providers are generally motivated to provide requested content to client computing devices often with consideration of efficient transmission of the requested content to the client computing device and/or consideration of a cost associated with the transmission of the content. Accordingly, content providers often consider factors such as latency of delivery of requested content in processing client computing device requests in order to meet service level agreements or to generally improve the quality of delivery service.

DESCRIPTION OF THE DRAWINGS

[0005] The foregoing aspects and many of the attendant advantages of this invention will become more readily appreciated as the same become better understood by reference to the following detailed description, when taken in conjunction with the accompanying drawings, wherein:

[0006] FIGURE 1 is a block diagram illustrative of content delivery environment including a number of client computing devices, a content provider, a network storage provider, and a content delivery network service provider;

[0007] FIGURE 2 is a block diagram of the content delivery environment of FIGURE 1 illustrating the registration of a content provider with a network storage provider;

[0008] FIGURE 3 is a block diagram of the content delivery environment of FIGURE 1 illustrating the registration of a content provider with a network storage provider;

[0009] FIGURE 4A-4C are block diagrams of the content delivery environment of FIGURE 1 illustrating one embodiment of the processing of resource requests generated by client computing devices;

[0010] FIGURE 5 is a flow diagram illustrative of a content provider registration processing routine implemented by a service provider;

[0011] FIGURE 6 is a flow diagram illustrative of a request routing processing routine implemented by a service provider; and

[0012] FIGURE 7 is an illustrative user interface displaying information for use in registering request routing services with a service provider.

DETAILED DESCRIPTION

[0013] Generally described, the present disclosure is directed to managing requesting routing functionality associated with resource requests for one or more resources associated with

a content provider. Specifically, aspects of the disclosure will be described with regard to the management and processing request routing functionality by a service provider, such as a content delivery network (“CDN”) service provider, on behalf of a content provider. Illustratively, the request routing functionality can correspond to the processing of domain name service (“DNS”) requests for resources by computing devices and the resolution of the DNS requests by the identification of a network address of a computing device that will provide the requested resources. Unlike traditional CDN service provider implementation, the processing of DNS requests by the service provider can be separated from the storage of content by content provider or on behalf of the content provider. Additionally, the processing of the DNS requests by the service provider can also be separated from the delivery of the content by the content provider or on behalf of the content provider.

[0014] Although various aspects of the disclosure will be described with regard to illustrative examples and embodiments, one skilled in the art will appreciate that the disclosed embodiments and examples should not be construed as limiting. For example, the present disclosure may be described with regard to request routing services provided by a service provider, such as a CDN service provider, that may provide additional services and functionality including network-based storage services, caching services, and content delivery services. However, one skilled in the relevant art will appreciate that a service provider need not provide all, or any, of the additional services or functionality that may be associated with some service providers, such as a CDN service provider.

[0015] FIGURE 1 is a block diagram illustrative of content delivery environment 100 for the managing registration of content provider with a service provider, such as a CDN service provider, and subsequent processing of at least a portion of content requests on behalf of the content provider. As illustrated in FIGURE 1, the content delivery environment 100 includes a number of client computing devices 102 (generally referred to as clients) for requesting content from a content provider, a network storage provider 110, and/or a CDN service provider 106. In an illustrative embodiment, the client computing devices 102 can correspond to a wide variety of computing devices including personal computing devices, laptop computing devices, hand-held computing devices, terminal computing devices, mobile devices, wireless devices, various electronic devices and appliances and the like. In an illustrative embodiment, the client

computing devices 102 include necessary hardware and software components for establishing communications over a communication network 108, such as a wide area network or local area network. For example, the client computing devices 102 may be equipped with networking equipment and browser software applications that facilitate communications via the Internet or an intranet.

[0016] Although not illustrated in FIGURE 1, each client computing device 102 utilizes some type of local DNS resolver component, such as a DNS Name server, that generates the DNS queries attributed to the client computing device. In one embodiment, the local DNS resolver component may be provide by an enterprise network to which the client computing device 102 belongs. In another embodiment, the local DNS resolver component may be provided by an Internet Service Provider (ISP) that provides the communication network connection to the client computing device 102.

[0017] The content delivery environment 100 can also include a content provider 104 in communication with the one or more client computing devices 102 via the communication network 108. The content provider 104 illustrated in FIGURE 1 corresponds to a logical association of one or more computing devices associated with a content provider. Specifically, the content provider 104 can include a web server component 112 corresponding to one or more server computing devices for obtaining and processing requests for content (such as Web pages) from the client computing devices 102. The content provider 104 can further include an origin server component 114 and associated storage component 116 corresponding to one or more computing devices for obtaining and processing requests for network resources. One skilled in the relevant art will appreciate that the content provider 104 can be associated with various additional computing resources, such additional computing devices for administration of content and resources and the like. Additionally, although the origin server component 114 and associated storage component 116 are logically associated with the content provider 104, the origin server component 114 and associated storage components 116 may be geographically distributed throughout the communication network 108 in a manner to best serve various demographics of client computing devices 102.

[0018] As further illustrated in FIGURE 1, the content provider 104 can be associated with one or more DNS name server components 118 that are operative to receive DNS queries

related to registered domain names associated with the content provider. The one or more DNS name servers can be authoritative to resolve client computing device DNS queries corresponding to the registered domain names of the content provider 104. A DNS name server component is considered to be authoritative to a DNS query if the DNS name server can resolve the query by providing a responsive IP address. As will be explained in greater detail below, in accordance with illustrative embodiments, at least a portion of the request routing functionality provided by the DNS name server components 118 will be provided by a service provider.

[0019] With continued reference to FIGURE 1, the content delivery environment 100 can further include a service provider 106 in communication with the one or more client computing devices 102, the content provider 104, and the network storage provider 110 via the communication network 108. The service provider 106 illustrated in FIGURE 1 corresponds to a logical association of one or more computing devices associated with a service provider. Specifically, the service provider 106 can include a number of Point of Presence (“POP”) locations 120, 126, 132 that correspond to nodes on the communication network 108. Each POP 120, 126, 132 includes a DNS component 122, 128, 134 made up of a number of DNS server computing devices for resolving DNS queries from the client computers 102. Each POP 120, 126, 132 also optionally includes a resource cache component 124, 130, 136 made up of a number of cache server computing devices for storing resources from content providers or network storage providers and transmitting various requested resources to various client computers. The DNS components 122, 128, 134 and the resource cache components 124, 130, 136 may further include additional software and/or hardware components that facilitate communications including, but not limited to, load balancing or load sharing software/hardware components.

[0020] In an illustrative embodiment, the DNS component 122, 128, 134 and resource cache component 124, 130, 136 are considered to be logically grouped, regardless of whether the components, or portions of the components, are physically separate. Additionally, although the POPs 120, 126, 132 are illustrated in FIGURE 1 as logically associated with the CDN service provider 106, the POPs will be geographically distributed throughout the communication network 108 in a manner to best serve various demographics of client computing devices 102. Additionally, one skilled in the relevant art will appreciate that the service provider 106 can be

associated with various additional computing resources, such additional computing devices for administration of content and resources, and the like.

[0021] With further continued reference to FIGURE 1, the content delivery environment 100 can also include a network storage provider 110 in communication with the one or more client computing devices 102, the service provider 106, and the content provider 104 via the communication network 108. The network storage provider 110 illustrated in FIGURE 1 also corresponds to a logical association of one or more computing devices associated with a network storage provider. Specifically, the network storage provider 110 can include a number of network storage provider Point of Presence (“NSP POP”) locations 138, 142 that correspond to nodes on the communication network 108. Each NSP POP 138, 142 includes a storage component 140, 144 made up of a number of storage devices for storing resources from content providers which will be processed by the network storage provider 110 and transmitted to various client computers. The storage components 140, 144 may further include additional software and/or hardware components that facilitate communications including, but not limited to, load balancing or load sharing software/hardware components.

[0022] In an illustrative embodiment, the storage components 140, 144 are considered to be logically grouped, regardless of whether the components, or portions of the components, are physically separate. Additionally, although the NSP POPs 138, 142 are illustrated in FIGURE 1 as logically associated with the network storage provider 110, the NSP POPs will be geographically distributed throughout the communication network 108 in a manner to best serve various demographics of client computing devices 102. Additionally, one skilled in the relevant art will appreciate that the network storage provider 110 can be associated with various additional computing resources, such additional computing devices for administration of content and resources, DNS name servers, and the like. For example, the network storage provider 110 can be associated with one or more DNS name server components that are operative to receive DNS queries related to registered domain names associated with the network storage provider 110. The one or more DNS name servers can be authoritative to resolve client computing device DNS queries corresponding to the registered domain names of the network storage provider 110. As similarly set forth above, a DNS name server component is considered

to be authoritative to a DNS query if the DNS name server can resolve the query by providing a responsive IP address.

[0023] Even further, one skilled in the relevant art will appreciate that the components of the network storage provider 110 and components of the service provider 106 can be managed by the same or different entities. One skilled in the relevant art will also appreciate that the components and configurations provided in FIGURE 1 are illustrative in nature. Accordingly, additional or alternative components and/or configurations, especially regarding the additional components, systems, and subsystems for facilitating communications may be utilized. Specifically, one skilled in the relevant art will appreciate the network storage provider 110 may be omitted from the content delivery environment 100.

[0024] With reference now to FIGURES 2-4C, the interaction between various components of the content delivery environment 100 of FIGURE 1 will be illustrated. For purposes of the example, however, the illustration has been simplified such that many of the components utilized to facilitate communications are not shown. One skilled in the relevant art will appreciate that such components can be utilized and that additional interactions would accordingly occur without departing from the spirit and scope of the present disclosure.

[0025] With reference to FIGURE 2, an illustrative interaction for the optional registration of a content provider 104 with the network storage provider 110 for hosting content on behalf of the content provider 104 will be described. As illustrated in FIGURE 2, the storage provider content registration process begins with registration of the content provider 104 with the network storage provider 110. In an illustrative embodiment, the content provider 104 utilizes a registration application program interface (“API”) to register with the network storage provider 110 such that the network storage provider 110 can provide content on behalf of the content provider 104. The registration API can include the identification of the origin server 114 of the content provider 104 that may provide requested resources to the network storage provider 110. In addition or alternatively, the registration API can include the content to be stored by the network storage provider 110 on behalf of the content provider 104. In one embodiment, the network storage provider 110 may act as an origin server for the content provider 104.

[0026] One skilled in the relevant art will appreciate that upon storage of the content by the network storage provider 110, the content provider 104 can begin to direct requests for content from client computing devices 102 to the network storage provider 110. Specifically, in accordance with DNS routing principles, a client computing device request corresponding to a resource identifier would eventually be directed toward a storage component 140, 140, 144 of a NSP POP 138, 142 associated with the network storage provider 110. The request routing of a client computing device request will be described in greater detail below.

[0027] With continued reference to FIGURE 2, upon receiving the registration API, the network storage provider 110 obtains and processes the content provider registration information. In an illustrative embodiment, the network storage provider 110 can then generate additional information that will be used by the client computing devices 102 as part of the content requests. The additional information can include, without limitation, content provider identifiers, such as content provider identification codes, storage provider identifiers, such as storage provider identification codes, executable code for processing resource identifiers, such as script-based instructions, and the like. One skilled in the relevant art will appreciate that various types of additional information may be generated by the network storage provider 110 and that the additional information may be embodied in any one of a variety of formats.

[0028] In one embodiment, the network storage provider 110 returns an identification of applicable domains for the network storage provider (unless it has been previously provided) and any additional information to the content provider 104. In turn, the content provider 104 can then process the stored content with content provider specific information. In one example, as illustrated in FIGURE 2, the content provider 104 translates resource identifiers originally directed toward a domain of the origin server 114 to a domain corresponding to the network storage provider 110. The translated URLs are embedded into requested content in a manner such that DNS queries for the translated URLs will resolve to a DNS server corresponding to the network storage provider 110 and not a DNS server corresponding to the content provider 104.

[0029] Generally, the identification of the resources originally directed to the content provider 104 will be in the form of a resource identifier that can be processed by the client computing device 102, such as through a browser software application. In an illustrative embodiment, the resource identifiers can be in the form of a uniform resource locator (“URL”).

Because the resource identifiers are included in the requested content directed to the content provider, the resource identifiers can be referred to generally as the “content provider URL.” For purposes of an illustrative example, the content provider URL can identify a domain of the content provider 104 (e.g., contentprovider.com), a name of the resource to be requested (e.g., “resource.xxx”) and a path where the resource will be found (e.g., “path”). In this illustrative example, the content provider URL has the form of:

http://www.contentprovider.com/path/resource.xxx

[0030] In an alternative embodiment, the resource identifiers provided by the content provider 104 can correspond to the network storage provider 110, or network storage providers, that are hosting content on behalf of the content provider. These resource identifiers can be generally referred to as network storage provider URLs. The network storage provider URLs identify a domain of the network storage provider 110 (e.g., “storageprovider.com”), the same name of the resource to be requested (e.g., “resource.xxx”) and the same path where the resource will be found (e.g., “path”). Additionally, the network storage provider URL can include additional processing information (e.g., “additional information”) such as request routing information or identification information, such as a content provider identifier. Illustratively, the network storage provider URL would have the form of:

http://additional information.storageprovider.com/path/resource.xxx

[0031] In another embodiment, the information associated with the network storage provider 110 is included in the network storage provider URL, such as through prepending or other techniques, such that the network storage provider URL can maintain all of the information associated with the original URL. In this embodiment, the network storage provider URL would have the form of:

http://additional information.storageprovider.com/www.contentprovider.com
/path/resource.xxx

[0032] With reference now to FIGURE 3, an illustrative interaction for registration, by the content provider 104, to utilize the request routing services provided by the service provider 106 will be described. As illustrated in FIGURE 3, the request routing service registration process provided by the service provider 106 begins with registration of the content provider 104 with the service provider 106. In an illustrative embodiment, the content provider 104 utilizes a registration application program interface (“API”) to register with the service provider 106 such that the service provider 106 can provide request routing services on behalf of the content provider 104. The registration API includes the identification of the domains for which the service provider 106 will be authoritative (e.g., “contentprovider.com”), the identification of the storage component 140, 144 of the network storage provider 110 or origin servers 114 that will provide requested resources to the client computing devices 102. The registration API can also include additional information including request routing information, identification information, or other information that will be used to resolve client computing device DNS requests on behalf of the content provider 104, as will be explained below. An illustrative screen display for use in a manual configuration of the DNS request routing services will be described with regard to FIGURE 7.

[0033] The service provider 106 obtains the registration API and processes the information. In one aspect, the service provider 106 can generate the necessary request processing rules or alternative identifiers that may be utilized in the resolution of client computing device DNS queries. In another aspect, the service provider 106 can cause the registration of its DNS nameserver components for the relevant domains specified by the content provider 104. The service provider 104 can then send a response to the content provider 104, such as a confirmation.

[0034] Illustratively, upon the optional identification of appropriate storage component 140, 144, 148 of the network storage provider 110 and the registration for request routing functionality with the service provider 106, the content provider 104 can, in one embodiment as will be further described below in reference to FIGURES 4A-4C and 4B, begin to process DNS requests for content generated on behalf of the client computing devices 102. Specifically, in accordance with DNS routing principles, a client computing device DNS query corresponding to a resource identifier would eventually be resolved by identifying a network

address corresponding to either the origin server component 114 and associated storage component 116 or storage component 140, 144, 148 of the network storage provider 110 by a DNS nameserver associated with the service provider 106.

[0035] With reference first to FIGURE 4A, the interaction begins with the generation of a request for content from the client computing device 102, such as through a browser software application. The first request for content is often referred to as the original resource request, which often can correspond to a request for a Web page (or other network resource). As illustrated in FIGURE 4A, the original resource request is received and processed by the content provider 104. In accordance with an illustrative embodiment, the request for content can be in accordance with common network protocols, such as the hypertext transfer protocol (“HTTP”). Upon receipt of the content request, the content provider 104 identifies the appropriate responsive content. In an illustrative embodiment, the requested content can correspond to a Web page that is displayed on the client computing device 102 via the processing of information, such as hypertext markup language (“HTML”), extensible markup language (“XML”), and the like. The content provider 104, such as via a Web server component 112, can return the requested resource and include a number of embedded resources in the original request resources. For example, a Web page may include references to a number of embedded image and data files to be rendered by the client computing device 102.

[0036] With reference now to FIGURE 4B, upon receipt of the requested content, the client computing device 102, such as through a browser software application, begins processing any of the markup code included in the content and attempts to acquire the resources identified by the embedded resource identifiers (e.g., the embedded, modified URLs). Accordingly, the first step in acquiring the content correspond to the issuance, by the client computing device 102 (through its local DNS resolver), a DNS query for the resource identifier (e.g., the Original URL or network storage provider URL) that results in the identification of a DNS server authoritative to the “.” and the “com” portions of the URL. After partially resolving the modified URL according to the “.” and “com” portions of the URL, the client computing device 102 then issues another DNS query for the URL that results in “.contentprovider” or “.storageprovider” portion of the URL. The issuance of DNS queries corresponding to the “.” and the “com” portions of a

URL, such as the Original URL or network storage provider URL, are well known and have not been illustrated.

[0037] In an illustrative embodiment, the identification of the identification of a DNS server authoritative to the “contentprovider” corresponds to an IP address of a DNS server associated with the service provider 106. In one embodiment, the IP address is a specific network address unique to DNS server component(s) of a specific POP associated with the service provider 106. In another embodiment, the IP address can be shared by one or more POPs associated with the service provider 106, which may be geographically or logically distributed. In this embodiment, a DNS query to the shared IP address utilizes a one-to-many network routing schema, such as anycast, such a specific POP will receive the request as a function of network topology. For example, in an anycast implementation, a DNS query issued by a client computing device 102 to a shared IP address will arrive at a DNS server component of the service provider 106 logically having the shortest network topology distance, often referred to as network hops, from the client computing device. The network topology distance does not necessarily correspond to geographic distance. However, in some embodiments, the network topology distance can be inferred to be the shortest network distance between a client computing device 102 and a service provider POP.

[0038] With continued reference to FIGURE 4B, once one of the DNS servers in the service provider 106 receives the request, the specific DNS server attempts to resolve the request. In an illustrative embodiment, a specific DNS server can resolve the DNS query by identifying an IP address of a network storage provider storage component 142, 148, 154 (FIGURE 1) or the origin server component 114 and associated storage component 116 (FIGURE 1) that will process the request for the requested resource. The identified IP address of a network storage provider storage component 142, 148, 154 (FIGURE 1) or the origin server component 114 and associated storage component 116 (FIGURE 1) will be generically referred to as the “storage component” or the “selected storage component.” Illustratively, the service provider 106 can utilize information provided by the content provider 104, information maintained by the service provider 106 and information provided by third parties to resolve client computing device DNS queries.

[0039] In an illustrative embodiment, the service provider 106 will utilize cost information, at least in part, to resolve a DNS request and identify an IP address of one or more storage components that will process (or can process) the request for the requested resource (e.g., resolve the client computing device DNS query). The resolution of a DNS query as a function of cost information may depend on the cost information maintained by the server provider 106 or cost information transmitted as part of the DNS query (e.g., cost information included in the URL). In one example, the cost information may designate that the content provider 104 has requested that the service provider 106 select the storage component that will process the request for the requested resource associated with the lowest current cost to the content provider 104 to provide the requested resource. Accordingly, the service provider 106 could obtain cost information from the content provider 104 and utilize the cost information in resolving the DNS query. In another example, the service provider 106 can utilized cost information associated with an attributed cost to the service provider to deliver the requested resource if an associated component of the service provider is also providing the requested resource on behalf of the content provider 104.

[0040] In another embodiment, the service provider 106 can attempt to resolve the DNS query according to geographic criteria. The geographic criteria can correspond to geographic-based regional service plans contracted between the service-provider 106 and the content provider 104 or a storage component and the content provider 104. Accordingly, a client computing device 102 DNS query received in a region not corresponding to the content provider's regional plan may be better processed by a storage component in region corresponding to the content provider's regional plan. In this example, the DNS server component 118 may also obtain geographic information from the client computing device 102 directly (such as information provided by the client computing device or ISP) or indirectly (such as inferred through a client computing device's IP address).

[0041] In a further embodiment, the service provider 106 can attempt resolve the DNS query according to service level criteria. The service level criteria can correspond to service or performance metrics contracted between the service provider 106 and the content provider 104 or between a storage component and the content provider 104. Examples of performance metrics can include thresholds for latencies of data transmission associated with

transmissions to the client computing devices 102, total data provided on behalf of the content provider 104, error rates for data transmissions, and the like.

[0042] In still a further embodiment, the service provider 106 can attempt to resolve the DNS query according to network performance criteria. The network performance criteria can correspond to measurements of network performance for transmitting data from the service provider 106 to the client computing device 102 or storage components to the client computing devices. Examples of network performance metrics can include network data transfer latencies (measured by the client computing device 102 or the service provider 106, network data error rates, and the like.

[0043] As an alternative to selecting a storage component (e.g., resolving the DNS query), the CDN service provider 106 can maintain sets of various alternative resource identifiers. The alternative resource identifiers can be provided by the CDN service provider 106 to the client computing device 102 such that a subsequent DNS query on the alternative resource identifier will resolve to a different DNS server component within the CDN service provider's network. In such an embodiment, the receiving DNS nameserver may utilize alternative resource identifiers in the form of one or more canonical name ("CNAME") records to provide additional request routing information. In one embodiment, each CNAME record identifies a domain of the service provider (e.g., "serviceprovider.com" or "serviceprovider-1.com"). As will be explained in greater detail below, the domain in the CNAME does not need to be the same domain found in original URL or network storage provider URL. Still further, the domain in the CNAME may be unique to the content provider or a common domain utilized by the service provider on behalf of a number of content providers. Additionally, each CNAME record includes additional information, such as request routing information, (e.g., "request routing information"). An illustrative CNAME record can have the form of:

CNAME request_routing_information.serviceprovider.com

[0045] In a manner similar to the information described above, in an illustrative embodiment, each CNAME record includes additional information utilized by a receiving DNS name server for processing the DNS query. In an illustrative embodiment, the additional

information can include information or criteria utilized by the service provider 106 during the request routing process. The additional information included in the CNAME can be the same cost information provided in the original URL or network storage provider URL or additional/alternative costs information obtained by the service provider 106 and previously provided by the content provider 104. Illustratively, the additional information can include cost information, service level information, network performance information, client identifiers, content provider identifiers, or other identifiers used to look up request additional request routing information. An illustrative CNAME record can have the form of:

CNAME request_routing_information.cost_information.cdnprovider.com

[0046] In an illustrative embodiment, the CNAME records are generated and provided by the DNS servers to direct a more appropriate DNS server of the service provider 106. As used in accordance with the present disclosure, appropriateness can be defined in any manner by the service provider 106 for a variety of purposes. In an illustrative embodiment, the service provider 106 will utilize the additional information, at least in part, to identify the more appropriate DNS server of the service provider 106.

[0047] In one illustrative example, the additional information may designate the content provider 104 has requested that the cost associated with the providing the requested resource be maintained below one or more cost thresholds or cost tiers. Accordingly, the service provider 106 could obtain cost information for at least a portion of the POPs and select one or more DNS servers associated with a cost at or below the cost thresholds. The service provider 106 could then utilize other request routing criteria to select from the selected DNS server (if more than one DNS server is identified) or selected in accordance with other selections methodologies (e.g., random, round robin, etc.).

[0048] In another illustrative example, the service provider 106 can attempt to direct a DNS query to DNS servers according to geographic criteria. The geographic criteria can correspond to geographic-based regional service plans contracted between the service-provider 106 and the content provider 104 in which various service provider 106 POPs are grouped into geographic regions. Accordingly, a client computing device 102 DNS query

received in a region not corresponding to the content provider's regional plan may be better processed by a DNS server in region corresponding to the content provider's regional plan. In this example, the DNS server component 118 may also obtain geographic information from the client directly (such as information provided by the client computing device or ISP) or indirectly (such as inferred through a client computing device's IP address).

[0049] In a further illustrative example, the service provider 106 can attempt to direct a DNS query to DNS servers according to service level criteria. The service level criteria can correspond to service or performance metrics contracted between the service provider 106 and the content provider 104. Examples of performance metrics can include latencies of data transmission between the service provider POPs and the client computing devices 102, total data provided on behalf of the content provider 104 by the service provider POPs, error rates for data transmissions, and the like.

[0050] In still a further example, the service provider 106 can attempt to direct a DNS query to DNS servers according to network performance criteria. The network performance criteria can correspond to measurements of network performance for transmitting data from the CDN service provider POPs to the client computing device 102. Examples of network performance metrics can include network data transfer latencies (measured by the client computing device or the service provider 106, network data error rates, and the like.

[0051] In accordance with an illustrative embodiment, the DNS server maintains a data store that defines CNAME records for various URLs. If a DNS query corresponding to a particular URL matches an entry in the data store, the DNS server component 118 returns a CNAME record as defined in the data store. In an illustrative embodiment, the data store can include multiple CNAME records corresponding to a particular original URL. The multiple CNAME records would define a set of potential candidates that can be returned to the client computing device. In such an embodiment, the DNS server component 118, either directly or via a network-based service, can implement additional logic in selecting an appropriate CNAME from a set of possible of CNAMEs. In an illustrative embodiment, each DNS server component 118, 124, 130 maintains the same data stores that define CNAME records, which can be managed centrally by the service provider 106. Alternatively, each DNS server

component 118, 124, 130 can have a POP specific data stores that define CNAME records, which can be managed centrally by the service provider 106 or locally at the POP 116, 122, 128.

[0052] With reference now to FIGURE 4C, upon receipt of the successful resolution of the DNS query (e.g., resulting in the returning of an IP address to the client computing device 102), the client computing device 102 transmits embedded resource requests to the IP address. In turn, the receiving storage component (either the network storage provider 110 or origin server component 114 and associated storage component 116) can process the request by providing the requested content. The requested content can be transmitted to the requesting client computing device 102 via the communication network 108.

[0053] With reference now to FIGURE 5, one embodiment of a routine 500 implemented by the service provider 106 for managing registration of content provider 104 will be described. One skilled in the relevant art will appreciate that actions/steps outlined for routine 500 may be implemented by one or many computing devices/components that are associated with the service provider 106. Accordingly, routine 500 has been logically associated as being generally performed by the service provider 106, and thus the following illustrative embodiments should not be construed as limiting.

[0054] At block 502, the service provider 106 obtains a request for providing DNS request routing services on behalf of a content provider 104. Illustratively, the request for providing DNS request routing services may be facilitated through a registration API in which the content provider specifies information necessary for the service provider 106 to begin hosting DNS nameserver functionality on behalf of the content provider. The transmission of the registration API (and associated information) may be an automatic process corresponding to an exchange of information between computing devices without need for administrator, or other human interaction. Alternatively, the transmission of the registration API (and associated information) may be an automatic process may be a manual, or semi-manual, process in which an administrator specifies at least a portion of the information necessary for the service provider 106 to begin hosting DNS nameserver functionality on behalf of the content provider. An illustrative screen display for the manual specification of at least a portion of the information necessary for the service provider 106 to begin hosting DNS nameserver functionality on behalf of the content provider will be described with regard to FIGURE 7.

[0055] At block 504, the service provider 106 obtains the identification of the content provider domains and the storage components, cost information, request routing information or other information that will be utilized by the service provider in processing and resolving the DNS queries. Illustratively, the service provider 106 can obtain at least a portion of the information from the registration API. Additionally, the service provider 106 can request additional information from the content provider 104 or other third parties, such as network storage provider 110. At block 506, the service provider 106 generates request routing rules and information utilized in resolving DNS queries or providing CNAMEs. For example, the service provider 106 may generate, update, or modify a table of CNAMEs in accordance with the registration information. At block 508, the service provider 106 registers its DNS nameservers with the appropriate entities such that the service provider 106 can begin receiving DNS queries directed toward the specified content provider domains. The registration of IP addresses for DNS nameserver components is well known and will not be described in greater detail. At block 510, the routine 500 terminates.

[0056] With reference now to FIGURE 6, a request routing processing routine 600 implemented by the service provider 106 will be described. One skilled in the relevant art will appreciate that actions/steps outlined for routine 600 may be implemented by one or many computing devices/components that are associated with the service provider 106. Accordingly, routine 600 has been logically associated as being performed by the service provider 106.

[0057] At block 602, the service provider 106 obtains a DNS query corresponding to resource identifier, such as at a “receiving DNS server.” In an embodiment, the one of the DNS servers associated with the service provider 106 may directly receive the DNS query transmitted by a client computing device 102 or on behalf of a client computing device 102. Alternatively, the service provider may receive information associated with a DNS query via an API transmitted by a computing device receiving the DNS query transmitted by a client computing device 102 or on behalf of a client computing device 102.

[0058] As previously discussed, the resource identifier can be a URL that has been embedded in content requested by the client computing device 102 and previously provided by the content provider 104. Alternatively, the resource identifier can also correspond to a CNAME provided by a content provider DNS server in response to a DNS query previously received from

the client computing device 102. At block 604, the receiving DNS server obtains content provider additional request routing information. As described above, in an illustrative embodiment, the additional request routing information may be maintained by the service provider 106 or included, at least in part, in the URL or CNAME. Such additional request routing information may specify that the service provider 106 should utilize additional request routing information in attempting to resolve the DNS query. As previously discussed, the additional request routing information can include cost information, service level information, geographic information, network performance information, and the like. The additional request routing information may be in the form of one or more thresholds (maximum, intermediate or minimum), ranges, averages, logic statements, declarative criteria, and the like.

[0059] At decision block 606, a test is conducted to determine whether the current DNS server is authoritative to resolve the DNS query. In one illustrative embodiment, the DNS server can determine whether it is authoritative to resolve the DNS query if there are no CNAME records corresponding to the received resource identifier. For example, based on the receiving DNS server may maintain one or more CNAMEs that define various cost alternatives for request routing processing. In this embodiment, the receiving DNS can utilize the additional request routing information obtained in block 604 to select the appropriate CNAME. Alternatively, the receiving DNS may select a CNAME without requiring additional information from the URL or CNAME. In one example, the additional request routing information may designate that the content provider 104 has requested that the service provider 106 select the DNS server of the service provider 106 associated with the lowest current cost to provide the requested resource. Accordingly, the service provider 106 could obtain additional request routing information for at least a portion of the POPs and select a CNAME corresponding to a DNS server associated with the lowest cost.

[0060] In another example, the additional request routing information may designate the content provider 104 has requested that the cost associated with the providing the requested resource be maintained below one or more cost thresholds or cost tiers so long as a maximum average latency threshold is maintained. Accordingly, the service provider 106 could obtain cost information and latency information associated for at least a portion of the POPs and select a CNAME corresponding to a DNS server associated with a cost satisfying the specified cost

threshold. Alternative or additional methodologies may also be practiced to determine whether the DNS server is authoritative.

[0061] If the current DNS server is authoritative (including a determination that the same DNS server will be authoritative for subsequent DNS queries), the current DNS server resolves the DNS query by returning the IP address of cache server component at block 608. In a non-limiting manner, a number of methodologies for selecting an appropriate resource cache component have been previously discussed. Additionally, the service provider 106 can obtain real time or semi-real time information regarding that status of the various available storage components, such as a current load, availability, resource utilization, health information, and the like. Still further, as described above, the IP address may correspond to a specific cache server of a resource cache component or generally to group of cache servers. The identification of the IP address may be provided to the client computing device 102 responsive to the received DNS query. Alternatively, if the DNS query was received as an API, the identification of the IP address may be provided via an API.

[0062] Alternatively, if at decision block 606, the DNS server is not authoritative, at block 610, the DNS server component selects and transmits an alternative resource identifier. As described above, the DNS server component can utilize a data store to identify an appropriate CNAME as a function of the current DNS query. Additionally, the DNS server component can also implement additional logical processing to select from a set of potential CNAMES. Illustratively, the identification of the selected CNAMES may be provided to the client computing device 102 responsive to the received DNS query. Alternatively, if the DNS query was received as an API, the identification of the CNAME may be provided via an API. At block 612, different DNS server components 118, 124, 130 receive a DNS query corresponding to the CNAME. The routine 600 then returns to decision block 606 and continues to repeat as appropriate.

[0063] With reference now to FIGURE 7, one illustrative embodiment of a user interface or screen display 700 in which information for registering a content provider 104 for request routing functionality with a service provider 106 will now be described. In this illustrative embodiment, the screen display 700 corresponds to a Web page or software

application display illustrating specification of request routing parameters to processing DNS queries for content provider resources.

[0064] The screen display 700 includes a first portion identifying the embedded resources, such as an embedded resource the storage components will provide, such as name 702, an embedded resource type 704, and a size of the embedded resource 706. The identifying information may be manually entered, or automatically filled in the service provider 106 or network storage provider 110. The screen display 700 includes as second portion for identifying the location of the storage components that will provided the resources, such as components 708 and 710. The second portion can also include various cost information, such as cost codes, or other request routing information that specifies criteria utilized by the service provider 106 to select among storage components.

[0065] With continued reference to FIGURE 7, the screen display 700 includes two additional sections 712, 714 for specifying additional information utilized in request routing, such as regional or geographic information and service level information. The screen display 700 also includes controls 716 for causing the submission of the registration information. One skilled in the relevant art will appreciate that the screen display 700 is illustrative in nature and that additional or alternative interface display components may be utilized. Additionally, additional or alternative screen displays may be implemented based on the association/relationship of the content provider 104 with other service providers, such as a network storage provider 110 or a CDN service provider.

[0066] It will be appreciated by those skilled in the art and others that all of the functions described in this disclosure may be embodied in software executed by one or more processors of the disclosed components and mobile communication devices. The software may be persistently stored in any type of non-volatile storage.

[0067] Conditional language, such as, among others, "can," "could," "might," or "may," unless specifically stated otherwise, or otherwise understood within the context as used, is generally intended to convey that certain embodiments include, while other embodiments do not include, certain features, elements and/or steps. Thus, such conditional language is not generally intended to imply that features, elements and/or steps are in any way required for one or more embodiments or that one or more embodiments necessarily include logic for deciding, with or

without user input or prompting, whether these features, elements and/or steps are included or are to be performed in any particular embodiment.

[0068] Any process descriptions, elements, or blocks in the flow diagrams described herein and/or depicted in the attached figures should be understood as potentially representing modules, segments, or portions of code which include one or more executable instructions for implementing specific logical functions or steps in the process. Alternate implementations are included within the scope of the embodiments described herein in which elements or functions may be deleted, executed out of order from that shown or discussed, including substantially concurrently or in reverse order, depending on the functionality involved, as would be understood by those skilled in the art. It will further be appreciated that the data and/or components described above may be stored on a computer-readable medium and loaded into memory of the computing device using a drive mechanism associated with a computer readable storing the computer executable components such as a CD-ROM, DVD-ROM, or network interface further, the component and/or data can be included in a single device or distributed in any manner. Accordingly, general purpose computing devices may be configured to implement the processes, algorithms, and methodology of the present disclosure with the processing and/or execution of the various data and/or components described above.

[0069] It should be emphasized that many variations and modifications may be made to the above-described embodiments, the elements of which are to be understood as being among other acceptable examples. All such modifications and variations are intended to be included herein within the scope of this disclosure and protected by the following claims.

Clause 1. A computer-implemented method for managing resource requests comprising:
obtaining a request routing inquiry from a content provider at a service provider, the request routing inquiry corresponds to a request for request routing functionality by the service provider on behalf of the content provider;

causing generation, by the service provider, of a user interface for request routing functionality by the service provider on behalf of the content provider;

obtaining registration information at the service provider for registering the one or more domains associated with the content provider with the service provider, the registration

information including a request by the content provider to request DNS query processing functionality by the service provider;

obtaining a DNS query from a client computing device at a first DNS server, wherein the DNS query corresponds to a requested resource associated with an original resource identifier provided by a content provider, wherein the original resource identifier corresponds to the one or more domains associated with the content provider and wherein the first DNS server corresponds to the service provider that is different from the content provider;

determining whether the first DNS server is authoritative to the DNS query as a function of additional information associated with the content provider, wherein at least a portion of the additional information is included in the DNS query;

obtaining an alternative resource identifier based on information associated with the client computing device and the content provider if the DNS server is determined not to be authoritative, wherein the alternative resource identifier includes request routing information not included in the original resource identifier and wherein the alternative resource identifier is selected at least in part on the additional information associated with the content provider;

transmitting the alternative resource identifier to the client computing device;

obtaining a subsequent DNS query from the client computing device at a second DNS server, wherein the subsequent DNS query corresponds to the alternative resource identifier provided by the service provider and wherein the second DNS server corresponds to the service provider that is different from the content provider;

determining whether the second DNS server is authoritative to the subsequent DNS query; and

selecting a cache component for providing content associated with the original resource request if the second DNS server is authoritative to the second DNS query; and

transmitting information identifying the selected cache component if the second DNS server is determined to be authoritative.

Clause 2. The method as recited in Clause 1, wherein the original resource identifier corresponds to a uniform resource locator provided by the content provider.

Clause 3. The method as recited in Clause 1, wherein the alternative resource identifier corresponds to a canonical name record identifier.

Clause 4. The method as recited in Clause 1 further comprising parsing the original resource identifier to obtain the cost information.

Clause 5. The method as recited in Clause 1, wherein the one or more resources correspond to one or more embedded resources.

Clause 6. The method as recited in Clause 1, wherein processing data corresponding to the registration of the one or more resources with the service provider comprises notifying the content provider of the registration of the one or more resources with the service provider.

Clause 7. The method as recited in Clause 1 further comprising:

obtaining a second alternative resource identifier based on information associated with the client computing device and the content provider if the second DNS server is determined not to be authoritative to the subsequent DNS query, wherein the second alternative resource identifier includes request routing information not included in the original resource identifier or the alternative resource identifier;

transmitting the alternative resource identifier to the client computing device;

obtaining a second subsequent DNS query from the client computing device at a third DNS server, wherein the second subsequent DNS query corresponds to the second alternative resource identifier provided by the service provider and wherein the third DNS server corresponds to the service provider that is different from the content provider;

determining whether the third DNS server is authoritative to the second subsequent DNS query; and

selecting a cache component for providing content associated with the original resource request if the third DNS server is authoritative to the second subsequent DNS query; and

transmitting information identifying the selected cache component if the third DNS server is determined to be authoritative to the second subsequent DNS query.

Clause 8. A computer implemented method for managing request routing functionality comprising:

obtaining registration information at a service provider for registering one or more content providers with the service provider, the registration corresponding to hosting at least of portion of request routing functionality associated with one or more resources provided by the content provider; and

obtaining a DNS query from a client computing device at a first DNS server, wherein the DNS query corresponds to a requested resource associated with a resource identifier and wherein the first DNS server corresponds to the service provider; and

providing the client computing device with at least one of an identifier corresponding to a storage component for the requested resource based on supplemental request routing information not included in the resource request and an alternative resource identifier query to resolve to a domain corresponding to the service provider and the supplemental request routing information not included in the first resource identifier.

Clause 9. The method as recited in Clause 8, wherein the one or more resources correspond to one or more embedded resources.

Clause 10. The method as recited in Clause 8, wherein the one or more resources correspond to a domain associated with the content provider

Clause 11. The method as recited in Clause 8, wherein the storage component corresponds to at least one of the content provider, the service provider and a network storage provider.

Clause 12. The method as recited in Clause 8, wherein the request routing information comprises service level plan information.

Clause 13. The method as recited in Clause 8, wherein the request routing information comprises geographic selection information.

Clause 14. The method as recited in Clause 8, wherein the request routing information comprises cost information associated with one or more storage components.

Clause 15. The method as recited in Clause 8 further comprising causing generation, by the service provider, of a user interface for registering the one or more domains associated with the content provider with the service provider.

Clause 16. A system for request routing comprising:

an interface component for obtaining registration information at a service provider for registering one or more content providers with the service provider, the registration corresponding to hosting at least of portion of request routing functionality associated with one or more resources provided by the content provider

a first network point of presence associated with a service provider, wherein the first network point of presence includes a DNS server component that receives a DNS query from a client computing device, wherein the DNS query corresponds to a requested resource associated with a first resource identifier, and wherein the DNS server in the first network point of presence is operable to:

determine that the first resource identifier is associated with an alternative resource identifier as a function of request routing criteria associated with a content provider; and

transmit an alternative resource identifier to the client computing device, wherein the alternative resource identifier includes information for causing a DNS query to resolve to a domain corresponding to the service provider;

a second network point of presence associated with a service provider, wherein the second network point of presence includes a DNS server component that receives a subsequent DNS query from a client computing device, wherein the subsequent DNS query corresponds to a requested resource associated with the alternative resource identifier, and wherein the DNS server in the second network point of presence is operable to:

resolve the subsequent DNS query to identify a cache component for providing content associated with the original resource request; and

transmit information identifying the identified cache component to the client computing device.

Clause 17. The system as recited in Clause 16, wherein the original resource identifier includes at least a portion of the request routing criteria.

Clause 18. The system as recited in Clause 16, wherein the DNS server in the first network point of presence is further operable to obtain at least a portion of the cost information.

Clause 19. The system as recited in Clause 16, wherein the alternative resource identifier corresponds to a canonical name record identifier.

Clause 20. The system as recited in Clause 16, wherein the request routing criteria includes regional service information and wherein the alternative resource identifier includes information identifying a regional service plan provided by the service provider.

Clause 21. The system as recited in Clause 20, wherein the DNS server in the first network point of presence is operable to obtain the information corresponding to a geographic location of the client computing device.

Clause 22. The system as recited in Clause 16, wherein information identifying the identified cache component to the client computing device includes a network address of a cache component associated with the content provider.

Clause 23. The system as recited in Clause 16, wherein information identifying the identified cache component to the client computing device includes a network address of a cache component associated with a third party service provider.

Clause 24. The system as recited in Clause 16, wherein the request routing criteria includes cost information and wherein the alternative resource identifier includes information identifying a cost plan provided by the service provider.

Clause 25. The system as recited in Clause 16, wherein the request routing criteria includes network performance information and wherein the alternative resource identifier includes information identifying a network performance plan provided by the service provider.

WHAT IS CLAIMED IS:

1. A computer-implemented method for managing request routing functionality comprising:

obtaining registration information at a service provider for registering one or more content providers with the service provider, the registration corresponding to hosting at least of portion of request routing functionality associated with one or more resources provided by the content provider; and

obtaining a DNS query from a client computing device at a first DNS server, wherein the DNS query corresponds to a requested resource associated with a resource identifier and wherein the first DNS server corresponds to the service provider; and

providing the client computing device with at least one of an identifier corresponding to a storage component for the requested resource based on supplemental request routing information not included in the resource request and an alternative resource identifier query to resolve to a domain corresponding to the service provider and the supplemental request routing information not included in the first resource identifier.

2. The method as recited in Claim 1, wherein the one or more resources correspond to one or more embedded resources.

3. The method as recited in Claim 1, wherein the one or more resources correspond to a domain associated with the content provider

4. The method as recited in Claim 1, wherein the storage component corresponds to at least one of the content provider, the service provider and a network storage provider.

5. The method as recited in Claim 1, wherein the request routing information comprises at least one of service level plan information and geographic selection information.

6. The method as recited in Claim 1, wherein the request routing information comprises cost information associated with one or more storage components.

7. The method as recited in Claim 1 further comprising causing generation, by the service provider, of a user interface for registering the one or more domains associated with the content provider with the service provider.

8. A system for request routing comprising:

an interface component for obtaining registration information at a service provider

for registering one or more content providers with the service provider, the registration corresponding to hosting at least of portion of request routing functionality associated with one or more resources provided by the content provider

a first network point of presence associated with a service provider, wherein the first network point of presence includes a DNS server component that receives a DNS query from a client computing device, wherein the DNS query corresponds to a requested resource associated with a first resource identifier, and wherein the DNS server in the first network point of presence is operable to:

determine that the first resource identifier is associated with an alternative resource identifier as a function of request routing criteria associated with a content provider; and

transmit an alternative resource identifier to the client computing device, wherein the alternative resource identifier includes information for causing a DNS query to resolve to a domain corresponding to the service provider;

a second network point of presence associated with a service provider, wherein the second network point of presence includes a DNS server component that receives a subsequent DNS query from a client computing device, wherein the subsequent DNS query corresponds to a requested resource associated with the alternative resource identifier, and wherein the DNS server in the second network point of presence is operable to:

resolve the subsequent DNS query to identify a cache component for providing content associated with the original resource request; and

transmit information identifying the identified cache component to the client computing device.

9. The system as recited in Claim 8, wherein the original resource identifier includes at least a portion of the request routing criteria.

10. The system as recited in Claim 8, wherein the DNS server in the first network point of presence is further operable to obtain at least a portion of the cost information.

11. The system as recited in Claim 8, wherein the alternative resource identifier corresponds to a canonical name record identifier.

12. The system as recited in Claim 8, wherein the request routing criteria includes regional service information and wherein the alternative resource identifier includes information identifying a regional service plan provided by the service provider.

13. The system as recited in Claim 8, wherein information identifying the identified cache component to the client computing device includes a network address of a cache component associated with the content provider.

14. The system as recited in Claim 8, wherein information identifying the identified cache component to the client computing device includes a network address of a cache component associated with a third party service provider.

15. The system as recited in Claim 8, wherein the request routing criteria includes cost information and wherein the alternative resource identifier includes information identifying a cost plan provided by the service provider.

16. The system as recited in Claim 8, wherein the request routing criteria includes network performance information and wherein the alternative resource identifier includes information identifying a network performance plan provided by the service provider.