



- (51) **International Patent Classification:**
G06F 21/00 (2013.01)
- (21) **International Application Number:**
PCT/US2014/042711
- (22) **International Filing Date:**
17 June 2014 (17.06.2014)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (71) **Applicant:** HEWLETT-PACKARD DEVELOPMENT COMPANY, L. P. [US/US]; 11445 Compaq Center Drive West, Houston, Texas 77070 (US).
- (72) **Inventor:** MANADHATA, Pratyusa K.; 5 Vaughn Drive, Suite 301, Princeton, New Jersey 08540 (US).
- (74) **Agent:** LEMMON, Marcus B.; Hewlett-Packard Company, Intellectual Property Administration, 3404 E Harmony Road, Mail Stop 35, Fortt Collins, Colorado 80528-9599 (US).
- (81) **Designated States** (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM,

DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

- (84) **Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- *as to the identity of the inventor (Rule 4.17(i))*

Published:

- *with international search report (Art. 21(3))*

(54) **Title:** DNS BASED INFECTION SCORES

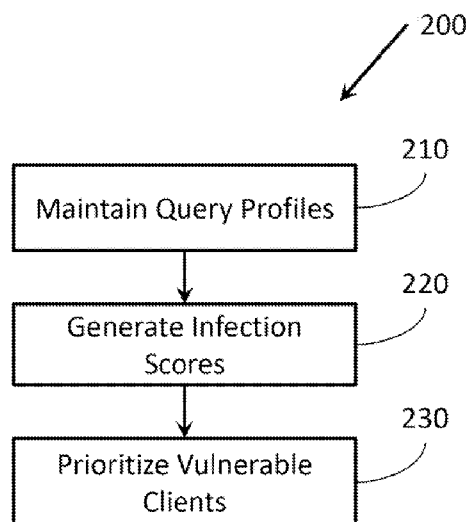


Figure 2

(57) **Abstract:** Systems and methods associated with domain name system (DNS) based infection scores. One example method includes maintaining query profiles for members of a set of clients in a network. The query profiles may be maintained based on DNS queries sent from the members of the set of clients, and on DNS responses received by the members of the set of clients. The method also includes generating infection scores for the members of the set of clients based on their respective query profiles. The method also includes prioritizing a vulnerable member of the set of clients for remedial action. The vulnerable member may be prioritized based on infection scores of members of the set of clients.

DNS BASED INFECTION SCORES

BACKGROUND

[0001] The domain name system (DNS) is used to translate web addresses (e.g., www.[example].com) into internet protocol (IP) addresses (e.g., 15.201.225.10). For example, when a client seeks to reach a website, the client will send a DNS request identifying the website by its web address to a DNS server. The DNS server will then lookup the web address in a table, and if the address is found in the table, the DNS server will respond with a corresponding IP address. DNS is used in internet communications, including malicious traffic (e.g., traffic related to attacks on enterprises).

BRIEF DESCRIPTION OF THE DRAWINGS

[0002] The present application may be more fully appreciated in connection with the following detailed description taken in conjunction with the accompanying drawings.

[0003] FIG. 1 illustrates an example system associated with DNS based infection scores.

[0004] FIG. 2 illustrates an example flowchart of example operations associated with DNS based infection scores.

[0005] FIG. 3 illustrates an example domain name system (DNS) based infection scoring system.

[0006] Fig. 4 illustrates another example DNS based infection scoring system associated with DNS based infection scores.

[0007] FIG. 5 illustrates an example computing environment in which example systems and methods, and equivalents, may operate.

DETAILED DESCRIPTION

[0008] Systems and methods associated with domain name system (DNS) based infection scores are described. Figure 1 illustrates an example system 100 associated with DNS based infection scores. System 100 may be a system in an enterprise used to identify which of several clients (e.g., client 112, client 114, and client 116) should be prioritized for remediation.

[0009] By way of illustration, clients 112, 114, and 116 may belong to a set of clients used by a company in various aspects of the company's business. The clients may include, for example, servers, personal computers assigned to users, and so forth. At any given time, many members of the set of clients may be infected with malware or otherwise compromised from a security perspective. However, due to limited resources of the company's information technology (IT) department, the IT department may be able to remediate issues on a limited number of clients in any given time period. Consequently, the IT department may desire to prioritize for remediation, clients that most need the remediation. Many factors may be taken into account for identifying which clients should be prioritized for remediation including, for example, the severity of infections of various clients, the data and/or user of a client, and so forth.

[0010] System 100 includes an observation logic 130. Observation logic may inspect DNS packets (e.g., DNS request packets 122). DNS packets 122 may be retrieved from a data stream 190 between a client (e.g., client 112) and a DNS server 199. Figure 1 illustrates two data streams 190, including a DNS request stream of DNS request packets sent from clients to DNS server 199 and a DNS response stream of DNS response packets sent from DNS server 199 to clients. To prevent slowdown of traffic in the network, the packet may be copied from a data stream 190 to system 100 by, for example, a network tap 195. In this example, client 112 sent DNS request packets 122 requesting internet protocol (IP) addresses associated with 4 domain names. Additionally, each of these domains may be associated with zero, one, or several internet protocol (IP) addresses. Though DNS request packets having single domain names are shown, there may be situations

where a DNS request packet contains several domain names. In this case, data may be tracked for each individual domain contained in a DNS request packet, and for groupings of IP addresses provided in DNS response packets.

[0011] In this example, the domain names are categorized as being associated with whitelisted domains (e.g., [good1].com, [good2].com), blacklisted domains (e.g., [bad1].com), and grey domains (e.g., [grey1].com). IP addresses may also be similarly categorized. Whitelisted domains and IP addresses may be known, safe, or assumed safe domains and IP addresses that have a low likelihood of being associated with a malicious event (e.g., a malware attack). Whitelisted domains and IP addresses may be, for example, selected based on a known or assumed list of safe domains and IP addresses, manually selected by an administrator, automatically updated over time based on network traffic, and so forth.

[0012] Blacklisted domains and IP addresses may include domains and IP addresses that are known or assumed to be associated with malicious attacks. For example, some malware will communicate with command and control servers to obtain instructions, provide stolen personal information, and so forth. If a client is requesting DNS information associated with a blacklisted domain or receives a DNS response containing a blacklisted IP address, it is likely that the client is infected with an associated malware which can be removed to improve security of the client. Blacklisted domains and IP addresses may be, for example, selected based on a known list of unsafe domains and IP addresses, manually selected by an administrator, automatically updated over time based on network traffic, and so forth.

[0013] Grey domains and IP addresses may be, for example, domains and IP addresses that are neither whitelisted nor blacklisted. Because essentially anyone can create a website, and these websites may change ownership over time, it may be difficult to maintain a fully accurate list of which domains and IP addresses are safe and which are unsafe. However, because most legitimate traffic is directed at a small number of domains and IP addresses, these domains and IP addresses may be whitelisted, domains and IP addresses known to be associated with

malicious events may be blacklisted, and the remaining domains and IP addresses may be treated as unknown, grey domains.

[0014] Because grey domains and IP addresses could be associated with either legitimate traffic or malicious events, a small number of accesses to grey domains may be treated as normal for most clients. However, large numbers of accesses to grey domains may indicate, for example, a misconfiguration of the client, a user of the client accessing inappropriate content, an as of yet unidentified malware on the client, and so forth. In these cases, there may be remedial actions that can be taken to reduce the number of grey domains and/or IP addresses being accessed. These actions may include, for example, reconfiguring a misconfigured client to stop mistakenly accessing grey domains, educating a user, studying and removing a malware, and so forth.

[0015] In evaluating how likely a client (e.g., client 112) is infected with a malware, observation logic 130 may inspect DNS request packets (e.g., DNS request packets 122) and DNS response packets and maintain a query profile associated with the client based on contents of the DNS request packets and DNS response packets. In this example, profile 142 is associated with client 112, profile 144 is associated with client 114, and profile 146 is associated with client 116. As mentioned above, DNS request packets 122, which were sent from client 112, contain 2 whitelisted domains, 1 blacklisted domain, and one grey domain. Consequently, Observation logic 130 may update profile 142 based on the numbers of types domains in DNS request packets 122.

[0016] Generally, when a client makes many DNS queries associated with blacklisted domains, it is likely that the client is infected by some form of malware. When a client primarily makes DNS queries associated with whitelisted domains, it is likely that the client is secure. When a client makes numerous DNS queries associated with grey domains, the client may be treated as suspicious. Suspicious clients may require further investigation to identify whether the client requires some remedial action.

[0017] Though profiles 142, 144, and 146 are only illustrated as tracking numbers of whitelisted, blacklisted, and grey domains from DNS request packets, many other metrics may be tracked within the query profiles. In addition to tracking numbers of various types of domains requested in DNS request packets, query profiles (e.g., profile 142) may track numbers of types of IP addresses received by clients (e.g., client 112) from DNS servers (e.g., DNS server 199) in DNS response packets. These types of IP address may include, for example, blacklisted IP addresses, whitelisted IP addresses, grey IP addresses, and so forth. As with domain names in DNS requests, receiving DNS responses containing blacklisted IP addresses may indicate a client is infected, receiving DNS responses containing whitelisted IP addresses may indicate a client is secure, and receiving DNS responses containing grey IP addresses may indicate a client is suspicious.

[0018] Further, numbers of IP addresses received in DNS responses may also be tracked in profiles (e.g., profile 142). By way of illustration, profiles may track how many DNS responses contain zero IP addresses (indicating a domain requested was invalid), how many DNS responses contain a single IP address, and how many DNS responses contain multiple IP addresses.

[0019] A large number of DNS responses containing no IP addresses, which are designated by a NXDOMAIN response in the DNS response packet, indicates a high likelihood that the client receiving the NXDOMAIN responses is infected. This may be because, for example some malware algorithmically generate domain names to contact command and control servers, and many of these may fail to resolve. Legitimate websites, on the other end of the spectrum, may respond with multiple IP addresses to facilitate better failover and/or load balancing. Thus, query profiles are likely associated with a healthy client when a query profile indicates a client has received mostly responses with multiple IP addresses.

[0020] Responses that contain only a single IP address may be treated as suspicious. Smaller and less popular websites (e.g., serving a niche interest group) may have a single IP address because load-balancing is unnecessary for smaller websites. It is also possible that a DNS request resolving to a single IP address is

associated with a malicious website. For example, a phishing site, hosted on a compromised personal computer, may resolve to a single IP address. In one example, it may be possible to predict whether a DNS request resolving to a single IP address is associated with a malicious attack by attempting to re-request the IP address at a later point in time. Because malicious sites are often cleaned up and/or detected and shut down relatively quickly, mappings between domain names and IP addresses associated with these malicious sites may be short lived and/or change regularly. Consequently, if an IP address associated with a domain changes after a predetermined period of time, it may be more likely that the client has accessed a malicious domain.

[0021] Derived metrics may also be stored in query profiles. Derived metrics may include, for example, what percentage of requested domains are associated with blacklisted domains and what percentage of received IP addresses are blacklisted IP addresses. This is because a client accessing 80 malicious domains out of 100 queries is likely more in need of remediation than a client accessing 100 malicious domains out of 1000 queries. Additionally, malicious domains and IP addresses may be weighted so that certain types of malicious events are prioritized for remediation. For example, a client having the Zeus worm on it may be considered more in need of remediation than a client having undesirable tracking cookies.

[0022] Periodically, the query profiles (e.g., profile 142, profile 144, and profile 146) may be run through an infection score generation logic 150. Infection score generation logic 150 may generate infection scores for each of the clients based on their respective query profiles. In one example, the scores may be generated based on control data 155. Control data 155 may contain data previously retrieved from a set of machines known to be infected and a set of machines known to be free of malware. By comparing query profiles to the control data, clients may be classified based on their levels of infection.

[0023] In one example, query profiles and profiles from the control data may be converted to feature vectors to facilitate comparison of the query profiles and the

control data. Some techniques for comparing the feature vectors may include, for example, decision trees, logistic regression, support vector machines, and so forth. Depending on the comparison technique applied, different scoring functions may be useful. For example, when using a decision tree to classify clients, a class distribution of a decision node may serve as a scoring function. When using support vector machines, distance of a vector from classification hyperplanes may serve as a scoring function. For regression analysis, the square root of an error metric may serve as the scoring function. In some examples, it may be possible to perform multiple types of classifications which may be useful when the control data contains information regarding multiple types of malware. This may facilitate assigning clients multiple labels based on their query profiles, which can allow clients to be classified as having multiple types of malware. Knowing what types of malware have infected a client may be used for generating infection scores for clients, prioritizing clients for remediation, and recommending a remedial action to take on a client.

[0024] In another example, applying clustering techniques to the query profiles may facilitate generating infection scores for the clients. Clusters of control data may be generated, including infected clusters and benign clusters. A query profile may then be compared to the cluster by measuring a distance between a vector generated from the query profile and the cluster. This distance may be measured from, for example, the cluster's centroid, the average distance from all points in the cluster, and so forth. This distance may serve as the scoring function.

[0025] Though it may be possible to update a client's infection score by comparing the client's query profile to the control data every time the query profile is updated due to detecting a new DNS request packet or a new DNS response packet associated with the client, this may be inefficient due to performance reasons. Consequently, a time period may be selected for periodically updating infection scores of clients. The time period may be for example, a minute, thirty minutes, an hour, eight hours, a day, and so forth. Selecting a long time period may facilitate building stronger query profiles for the clients. Selecting a shorter time period may reduce the likelihood that a profile may become polluted by, for example, a short burst of malicious traffic followed by a long period of normal use.

[0026] Once infection scores have been generated by infection score generation logic 150, the infection scores may then be provided to a prioritization logic 160. Prioritization logic 160 may use the infection scores to rank which client should be prioritized for remediation. The ranking may also be based on how important the various machines are to a company implementing system 100. By way of illustration, a client holding sensitive financial data or belonging to the CEO of the company may be considered more important than a client belonging to a salesperson. In an alternative example, a proxy server may aggregate numerous DNS queries, including queries associated with malicious domains. However, because these queries are likely actually associated with clients other than the proxy server, the proxy server may be given a low priority. Prioritization logic 160 may also take into account, for example, past infection states of clients by examining and/or weighting previous infection scores for the clients.

[0027] In this example, a prioritization logic 160 may be illustrated. Client 112, which was given an infection score of 30% by infection score generation logic 150, has a value of 100, client 114 has an infection score of 40% and a value of 1000, and client 116 has an infection score of 80% and a value of 300. In this example, the priorities are ranked by multiplying infection scores by client value, and consequently client 114 has the highest priority for remediation, followed by client 116, and client 112. Thus, in this example, where the infection score generation logic is generating a score indicating a likelihood of infection, client 114 is prioritized for remediation over client 116 because client 114 has a higher value than client 116 despite the fact that client 116 is much more likely to be infected than client 114.

[0028] It is appreciated that, in the following description, numerous specific details are set forth to provide a thorough understanding of the examples. However, it is appreciated that the examples may be practiced without limitation to these specific details. In other instances, some methods and structures may not be described in detail to avoid unnecessarily obscuring the description of the examples. Also, the examples may be used in combination with each other.

[0029] Figure 2 illustrates a method 200 associated DNS based infection scores. Method 200 may be, for example, embodied on a non-transitory computer-readable medium storing computer-executable instructions. The instructions, when executed by a computer, may cause the computer to perform method 200.

[0030] Method 200 includes maintaining query profiles at 210 for members of a set of clients in a network. The query profiles may be maintained based on domain name system (DNS) queries sent from members of the set of clients. For example, the query profiles may track, for their respective members of the set of clients, quantities of types of domain names requested in DNS queries sent from the members of the set of clients. The types of domain names may include, for example, blacklisted domains, whitelisted domains, grey domains, and so forth.

[0031] The query profiles may also be maintained based on DNS responses received by the members of the set of clients. For example, the query profiles may track, for their respective members of the set of clients, quantities of types of internet protocol (IP) addresses received in DNS responses received by the members of the set of clients. The types of IP addresses received may include, for example, blacklisted IP addresses, whitelisted IP addresses, grey IP addresses, and so forth.

[0032] The query profiles may also track, for their respective members of the set of clients, quantities of DNS responses received that contain no IP addresses, quantities of DNS responses received that contain a single IP address, and quantities of DNS responses received that contain multiple IP addresses, and so forth. Here a DNS response refers not to an entire DNS response packet but to a response to an individual DNS request associated with a domain name, as many DNS requests and/or responses may be included in a DNS packet.

[0033] The query profiles may also track, for their respective members of the set of clients, derived features of the DNS queries and DNS responses. The derived features may include, for example, a percentage of blacklisted domains requested, a percentage of blacklisted IP addresses received, a weighted score of blacklisted domains requested, a weighted score of blacklisted IP addresses received, and so forth. Weighted scores may be generated based on a metric defining how

dangerous certain blacklisted IP addresses are. As some malware may be more likely to compromise, modify, and/or erase sensitive data, these malware may be considered more threatening than a malware that mildly deteriorates client performance. Thus, IP addresses and domains on the blacklist may be weighted to reflect their respective dangerousness to various clients. Tracking weighted scores of blacklisted domains and blacklisted IP addresses may therefore include factoring in the weights of the blacklisted IP addresses and domains into the respective metrics.

[0034] Method 200 also includes generating infection scores for members of the set of clients. The infection scores may be generated for members of the set of clients based on their respective query profiles at 220. Generating infection scores may comprise generating feature vectors for members of the set of clients based on their respective query profiles. Generating infection scores may also include comparing the feature vectors to a control set of feature vectors having known infection states. The feature vectors may be compared to the control set of feature vectors using, for example, a decision tree, a logistic regression, clustering, a support vector machine, and so forth.

[0035] The comparison may be based on a scoring function that facilitates ranking infection states of members of the set of clients. For example, for a decision tree, a class distribution at various decision nodes may serve as the scoring function. For a support vector machine, the distance of feature vectors from classification hyperplanes may serve as the scoring function. For logistic regressions or other regression techniques, the square root of an error value may serve as the scoring function. For clustering, a distance between a vector and an attribute of a cluster (e.g., a point, a vector, an average) may server as a scoring function.

[0036] Method 200 also includes prioritizing a vulnerable member of the set of clients for a remedial action at 230. The prioritization may be based on infection scores of members of the set of clients. In a case where the infection score serves as the priority score, the higher the infection score of a client, the higher the priority that client will be given for remediation.

[0037] Prioritizing the vulnerable member of the set of clients may also be based on contents of the members of the set of clients, user identities associated with members of the set of clients, past infection and/or priority states of members of the set of clients and so forth. Prioritizing the vulnerable member based on the contents of the clients and on user identities associated with the clients may ensure that clients having sensitive data and clients assigned to important persons (e.g., a CEO) are more likely to be remediated than less important clients (e.g., a proxy server, a client assigned to a salesperson). In this scenario, clients may therefore be assigned a value V based on their contents, user identities, and so forth. A prioritization value P for a client may be generated by multiplying the client's infection score I by the client's value V . Thus, in some examples, prioritization value P may essentially be a weighted infection score, where the infection score of a client is weighted by a value assigned to the infection score.

[0038] Prioritizing the vulnerable client based on past priority scores may be useful because clients that have moderate infection scores over a long period of time may not be otherwise prioritized for a remedial action. Similarly, clients that become infected in the past may become infected again despite remedial action due to, for example, a remedial action being taken that is not sufficient to remedy the infection, or user behavior that should be corrected. Consequently, a client's priority may take into account the priority score of the client associated with the current time period, as well priority scores from the last N time periods, where N is defined by a policy. To ensure that clients that have high priority scores recently are more likely to be prioritized for remedial action, priority scores may be decayed over time. By way of illustration, where P_k represents an priority score from k time periods ago, current priority P taking decay into account may be estimated according to:

$$1. \quad P = \frac{P_N}{2^N} + \frac{P_{N-1}}{2^{N-1}} + \dots + \frac{P_1}{2^1} + \frac{P_0}{2^0}$$

[0039] The remedial action may be, for example, removing malware from the vulnerable member of the set of clients. This remedial action may be appropriate

when there is an easily removable malware on the vulnerable client. The remedial action may be installing a security application on the vulnerable member of the set of clients. The security application may be for example, a firewall, an antivirus application, an ad-blocker, and so forth. The remedial action may include restoring the vulnerable member of the set of clients to a prior state. This may be required when there is no simple fix to the current state of the vulnerable member of the set of clients but prior states of the vulnerable member were not infected. In even more extreme cases, the remedial action may require replacing the vulnerable member of the set of clients.

[0040] Figure 3 illustrates a domain name system (DNS) based infection scoring system 300. DNS based infection scoring system 300 may be incorporated into, for example, an enterprise security management system. DNS based infection scoring system 300 includes a data store 310. Data store 310 may store domain name system (DNS) query profiles 320. DNS query profiles 320 may be associated with members of a set of clients 399. The set of clients may include clients protected by DNS based infection scoring system 300.

[0041] DNS based infection scoring system 300 also includes an observation logic 330. Observation logic 330 may perform real time updating of DNS query profiles 320 based on DNS queries 390 and DNS responses 391. DNS queries 390 may be queries that are sent from clients 399 to, for example, a DNS server 398. DNS responses 391 may be responses provided by DNS server 398 to clients 399 in response to DNS queries 390.

[0042] DNS based infection scoring system 300 also includes an infection score generation logic 340. Infection score generation logic 340 may generate weighted infection scores for members of the set of clients 399. The infection scores may be generated for members of the set of clients 399 based on their respective DNS query profiles 320.

[0043] DNS based infection scoring system 300 also includes a prioritization logic 350. Prioritization logic 350 may cause a remedial action to be performed on a member of the set of clients 399. The member of the set of clients 399 on which the

remedial action is performed may be a member of the set of clients having a highest weighted infection score. In one example, prioritization logic 350 may cause the remedial action to be performed by initiating a logic (not shown) to perform the remedial action. In another example, prioritization logic 350 may cause the remedial action to be performed by providing an alert to an administrator identifying the member of the set of clients and the remedial action. Providing the alert may be appropriate instead of initiating a logic to perform the remedial action when the remedial action requires physical access to the machine. There may be other scenarios where it is also desirable to involve an administrator. For example, if the remedial action includes educating a user of the client regarding safe usage behavior, involving a person may make it more likely that educating the user is successful.

[0044] Figure 4 illustrates a DNS based infection scoring system 400 associated with DNS based infection scores. DNS based infection scoring system 400 contains several items similar to those described with reference to DNS based infection scoring system 300 (Figure 3). For example, DNS based infection scoring system 400 includes a data store 410 containing DNS query profiles 420 for a set of clients (not shown), an observation logic 430, an infection score generation logic 440, and a prioritization logic 450.

[0045] DNS based infection scoring system 400 also includes a control data set 460. Control data set 460 may be used by infection score generation logic 440 when infection score generation logic 440 generates infection scores for members of the set of clients. Specifically, infection score generation logic 440 may compare DNS query profiles 420 to control data set 460 to generate infection scores. DNS query profiles 420 may be compared to control data set 460 using, for example, a decision tree, logistic regression, clustering, a support vector machine, and so forth.

[0046] DNS based infection scoring system 400 also includes a learning logic 470. Learning logic 470 may modify control data set 460 over time based on the infection scores. By way of illustration, over time, learning logic 470 may detect certain patterns in infection scores based on certain trends in query profiles.

Consequently, if these patterns are recognized as being associated with benign activity, control data set 460 may be updated to reduce the likelihood that benign activity is prioritized for remedial action. Conversely, if patterns in infection scores and query profiles begin to indicate that a certain type of activity is associated with a malicious event, control data set 460 may be updated to increase the likelihood that clients performing the certain type of activity are prioritized for remedial action. The learning logic may also incorporate, for example, input from administrators dispatched to perform remedial actions, data from external logics that seek to track and identify new malicious events, the query profiles and so forth.

[0047] FIG. 5 illustrates an example computing environment in which example systems and methods, and equivalents, may operate. The example computing device may be a computer 500 that includes a processor 510 and a memory 520 connected by a bus 530. The computer 500 includes a DNS based infection score generation logic 540. In different examples, DNS based infection score generation logic 540 may be implemented as a non-transitory computer-readable medium storing computer-executable instructions in hardware, software, firmware, an application specific integrated circuit, and/or combinations thereof.

[0048] The instructions, when executed by computer 500, may cause computer 500 to inspect domain name system (DNS) packets set from a client and received by a client for attributes. The attributes may indicate a suspicion level of the client. The client may be a member of a set of clients in a network protected by computer 500. The instructions may also cause computer 500 to collect the attributes over time into a query profile associated with the client. The instructions may also cause computer 500 to compare the query profile against control profiles to generate an infection score for the client. The instructions may also cause computer 500 to schedule the client for a remedial action when then infection score of the client has a high infection score when compared against infection scores of other members of the set of clients. Whether a client has a high infection score may take into account sensitivity levels of members of the set of clients. The sensitivity levels may be related to, for example, content stored on the client, who uses the client, and so forth.

[0049] The instructions may also be presented to computer 500 as data 550 and/or process 560 that are temporarily stored in memory 520 and then executed by processor 510. The processor 510 may be a variety of various processors including dual microprocessor and other multi-processor architectures. Memory 520 may include volatile memory (e.g., read only memory) and/or non-volatile memory (e.g., random access memory). Memory 520 may also be, for example, a magnetic disk drive, a solid state disk drive, a floppy disk drive, a tape drive, a flash memory card, an optical disk, and so on. Thus, memory 520 may store process 560 and/or data 550. Computer 500 may also be associated with other devices including other computers, peripherals, and so forth in numerous configurations (not shown).

[0050] It is appreciated that the previous description of the disclosed examples is provided to enable any person skilled in the art to make or use the present disclosure. Various modifications to these examples will be readily apparent to those skilled in the art, and the generic principles defined herein may be applied to other examples without departing from the spirit or scope of the disclosure. Thus, the present disclosure is not intended to be limited to the examples shown herein but is to be accorded the widest scope consistent with the principles and novel features disclosed herein.

WHAT IS CLAIMED IS:

1. A non-transitory computer-readable medium storing computer-executable instructions that when executed by a computer cause the computer to:
 - maintain query profiles for members of a set of clients in a network based on domain name system (DNS) queries sent from the members of the set of clients and DNS responses received by the members of the set of clients;
 - generate infection scores for the members of the set of clients based on their respective query profiles; and
 - prioritize a vulnerable member of the set of clients for remedial action based on infection scores of members of the set of clients.
2. The non-transitory computer-readable medium of claim 1, where the query profiles track, for their respective members of the set of clients, quantities of types of domain names requested in DNS queries sent from the members of the set of clients, and where the types of domain names include blacklisted domains, whitelisted domains, and grey domains.
3. The non-transitory computer-readable medium of claim 1, where the query profiles track, for their respective members of the set of clients, quantities of types of internet protocol (IP) addresses received in DNS responses received by the members of the set of clients, and where the types of IP addresses received include blacklisted IP addresses, whitelisted IP addresses, and grey IP addresses.
4. The non-transitory computer-readable medium of claim 1, where the query profiles track, for their respective members of the set of clients, quantities of DNS responses received that contain no IP addresses, quantities of DNS responses received that contain a single IP address, and quantities of DNS responses received that contain multiple IP addresses.

5. The non-transitory computer-readable medium of claim 1, where the query profiles track, for their respective members of the set of clients, derived features of the DNS queries and DNS responses, where the derived features include one or more of, a percentage of blacklisted domains requested, a percentage of blacklisted IP addresses received, a weighted score of blacklisted domains requested, and a weighted score of blacklisted IP addresses received.

6. The non-transitory computer-readable medium of claim 1, where the instructions that cause the computer to generate infection scores comprise instructions that cause the computer to:

generate feature vectors for members of the set of clients based on their respective query profiles; and

compare the feature vectors to a control set of feature vectors having known infections states.

7. The non-transitory computer-readable medium of claim 6, where the feature vectors are compared to the control set of feature vectors using one or more of, a decision tree, a logistic regression, clustering, and a support vector machine.

8. The non-transitory computer-readable medium of claim 1, where prioritizing the vulnerable member of the set of clients is also based on one or more of, contents of the members of the set of clients, user identities associated with members of the set of clients, and past infection states of members of the set of clients.

9. The non-transitory computer-readable medium of claim 1, where the remedial action comprises one or more of, removing a malware from the vulnerable member of the set of clients, installing a security application on the vulnerable member of the set of clients, restoring the vulnerable member of the set of clients to a prior state, and replacing the vulnerable member of the set of clients.

10. An domain name system (DNS) based infection scoring system, comprising:
- a data store to store DNS query profiles associated with members of a set of clients protected by the DNS based infection scoring system;
 - an observation logic to perform real time updating of the DNS query profiles based on DNS queries sent from members of the set of clients and DNS responses received by the members of the set of clients;
 - an infection score generation logic to generate weighted infection scores for members of the set of clients based on their respective DNS query profiles; and
 - a prioritization logic to cause a remedial action to be performed on a member of the set of clients having a highest weighted infection score.
11. The DNS based infection scoring system of claim 10, where the prioritization logic causes the remedial action to be performed by one or more of, initiating a logic to perform the remedial action, and providing an alert to an administrator identifying the member of the set of clients and the remedial action.
12. The DNS based infection scoring system of claim 10, further comprising a control data set against which DNS query profiles are compared when generating infection scores for members of the set of clients.
13. The DNS based infection scoring system of claim 12, further comprising a learning logic to modify the control data set over time based on the infection scores.
14. The DNS based infection scoring system of claim 13, where the learning logic also modifies the control data set over time based on one or more of, input from administrators, data from external logics, and the query profiles.

15. A method, comprising:

inspecting domain name system (DNS) packets sent from a client and received by the client for attributes that indicate a suspicion level of the client, where the client is a member of a set of clients in a network protected by the computer;

collecting the attributes over time into a query profile associated the client;

comparing the query profile against control profiles to generate an infection score for the client; and

scheduling the client for a remedial action when the infection score of the client has a high infection score, when compared against infection scores of other members of the set of clients, taking into account sensitivity levels of the members of the set of clients.

1/5

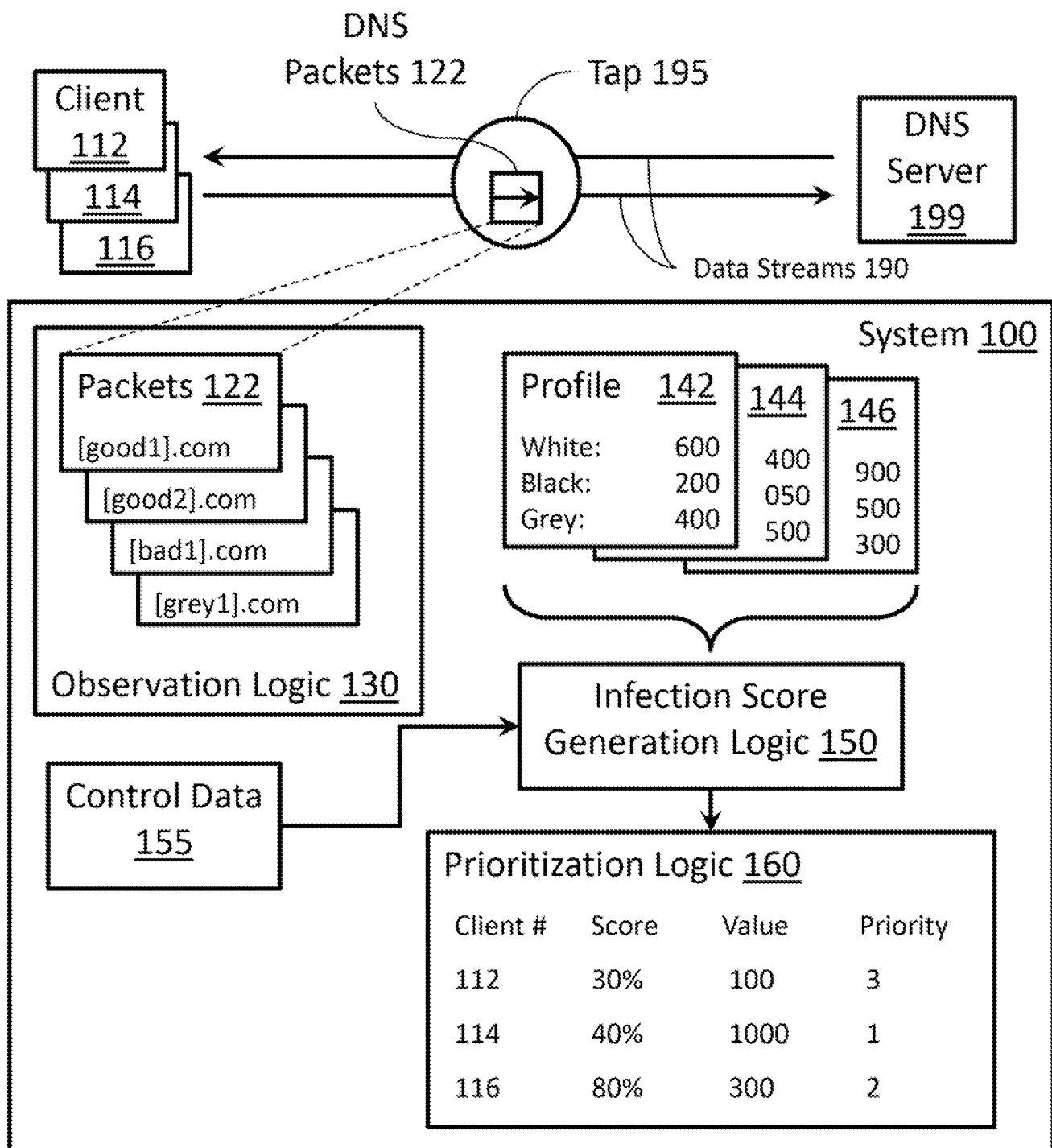


Figure 1

2/5

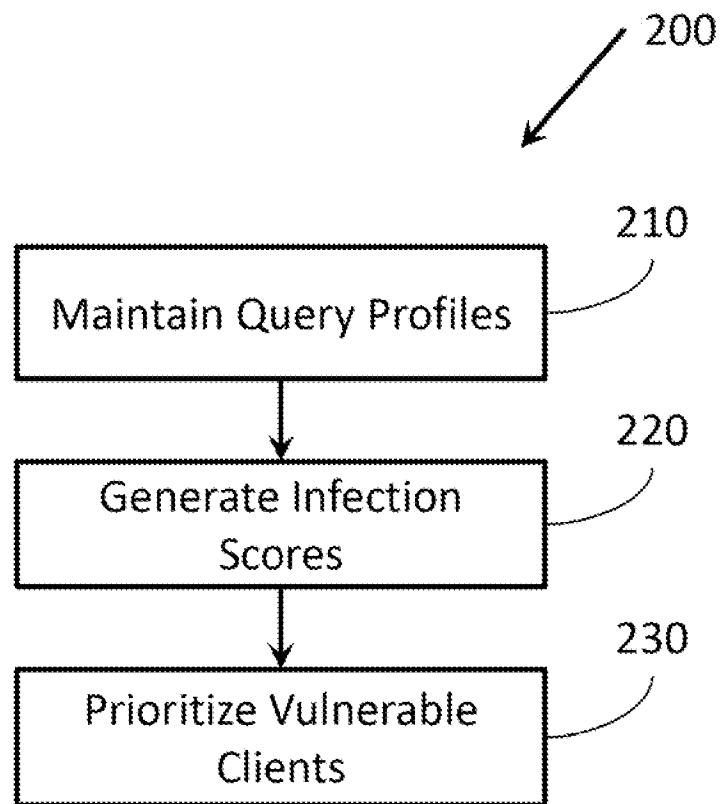
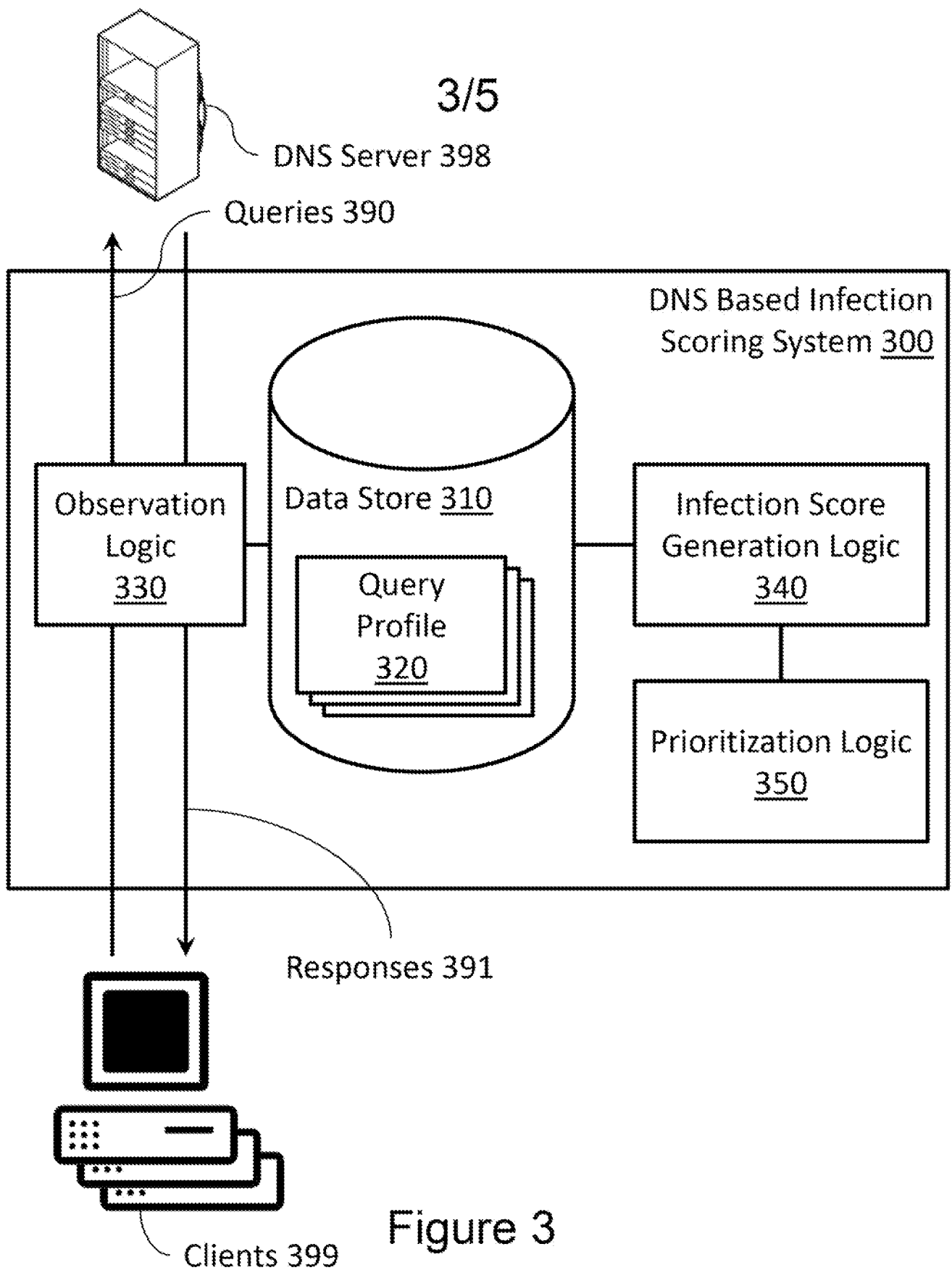


Figure 2



4/5

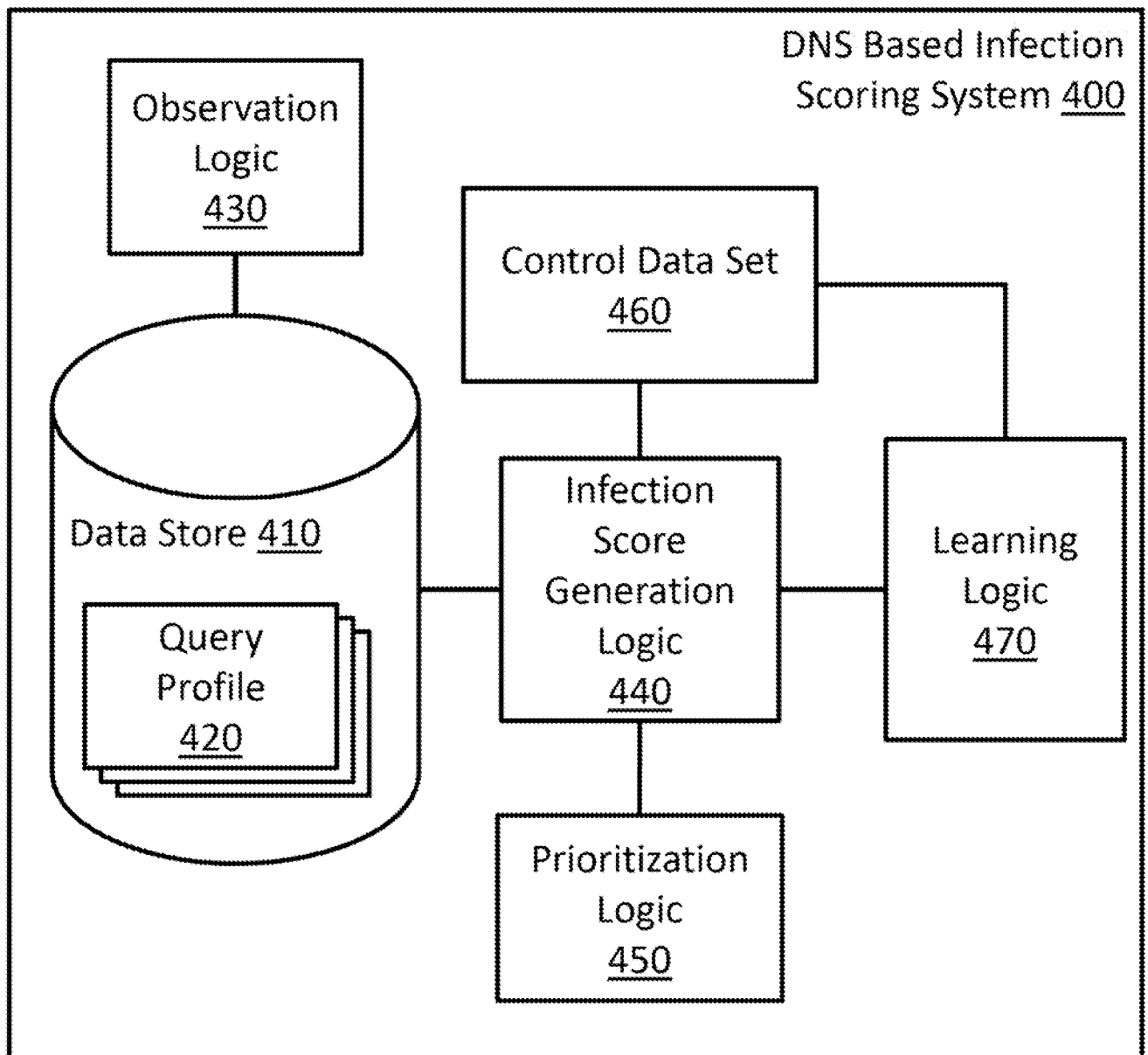


Figure 4

5/5

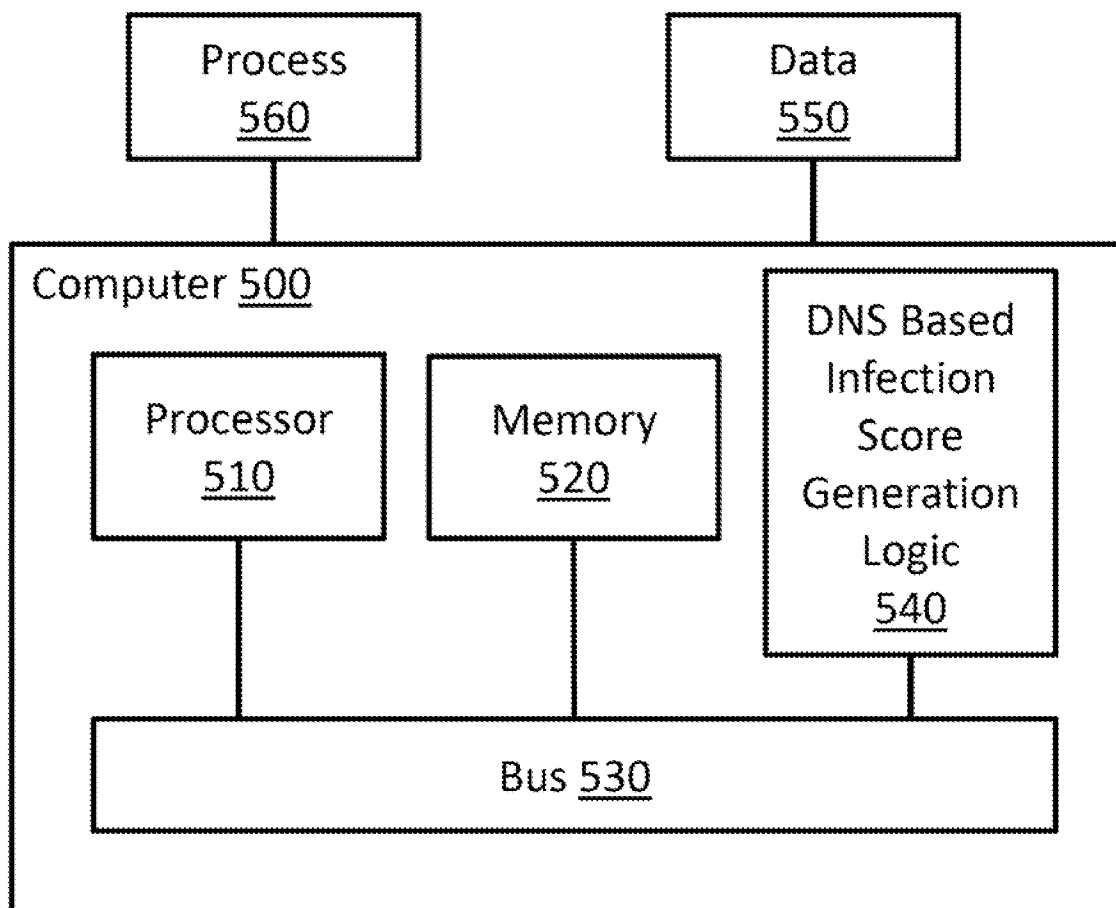


Figure 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2014/042711**A. CLASSIFICATION OF SUBJECT MATTER****G06F 21/00(2006.01)i**

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F 21/00; G06F 15/173; H04L 29/06; G06F 11/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean utility models and applications for utility models

Japanese utility models and applications for utility models

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS(KIPO internal) & Keywords: DNS, query, profile, infection, score, prioritize, remedial action, whitelist, blacklist, greylist, client

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2014-0101759 A1 (EMMANOUIL ANTONAKAKIS et al.) 10 April 2014 See paragraphs [0007]-[0009], [0017], [0047], [0057]-[0058]; and figures 1, 9.	1-15
Y	US 2010-0319069 A1 (EDOARD GRANSTEDT et al.) 16 December 2010 See paragraphs [0048], [0053]-[0054], [0068], [0071], [0075]; and figures 1b, 5.	1-15
A	US 2008-0028463 A1 (DAVID DAGON et al.) 31 January 2008 See paragraphs [0035]-[0036], [0047]-[0049]; and figures 1A, 14.	1-15
A	US 2012-0246315 A1 (MARTIN KAGAN) 27 September 2012 See paragraphs [0009]-[0015]; and figures 1-2.	1-15
A	US 2011-0078309 A1 (ERIC BLOCH et al.) 31 March 2011 See paragraphs [0066]-[0069]; and figure 3.	1-15



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

15 February 2015 (15.02.2015)

Date of mailing of the international search report

16 February 2015 (16.02.2015)

Name and mailing address of the ISA/KR

International Application Division
Korean Intellectual Property Office
189 Cheongsu-ro, Seo-gu, Daejeon Metropolitan City, 302-701,
Republic of Korea

Facsimile No. ++82 42 472 3473

Authorized officer

AHN, Jeong Hwan

Telephone No. +82-42-481-8440



INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2014/042711

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2014-0101759 A1	10/04/2014	US 2011-167495 A1 US 8578497 B2	07/07/2011 05/11/2013
US 2010-0319069 A1	16/12/2010	AU 2010-259950 A1 GB 201120260 D0 GB 2482273 A US 8407791 B2 WO 2010-144796 A2 WO 2010-144796 A3	01/12/2011 04/01/2012 25/01/2012 26/03/2013 16/12/2010 24/02/2011
US 2008-0028463 A1	31/01/2008	US 2014-245436 A1 US 8566928 B2 WO 2007-050244 A2 WO 2007-050244 A3	28/08/2014 22/10/2013 03/05/2007 23/04/2009
US 2012-0246315 A1	27/09/2012	CN 102656579 A CN 102667749 A EP 2488957 A1 EP 2497034 A1 EP 2497034 A4 US 2012-0246290 A1 WO 2011-056796 A1	05/09/2012 12/09/2012 22/08/2012 12/09/2012 31/07/2013 27/09/2012 12/05/2011
US 2011-0078309 A1	31/03/2011	US 7849502 B1 US 7849507 B1 US 8087082 B2	07/12/2010 07/12/2010 27/12/2011