(19) National Intellectual Property Office (NIPO)

(12) Patent applications for inventions

(10)Application Publication No. CN 115051845 A

(21) Application No. 202210639313 .6

(43) Application publication date 2022.09.13

(22) Application day 2022 .06 .08

(71) Applicant Beijing Qixingchen Information Security Technology Co.
No. 102, Qixingchen Building, Building 21, No. 8 Dongbeiwang West Road, Haidian District, Beijing, 100193, P.R.C.

Applicant Beijing NetGuard Nebula Information Technology Co.

(72) Inventors Chen Han

Liu Dunhui

(74) Patent Agency Beijing Jijia Intellectual Property Agency Co. Company 11227

Patent Attorney Gao Yong

(51) Int.Cl .

h04l 9/40 (2022 .01)

Claims 2 pages Instructions 8 pages

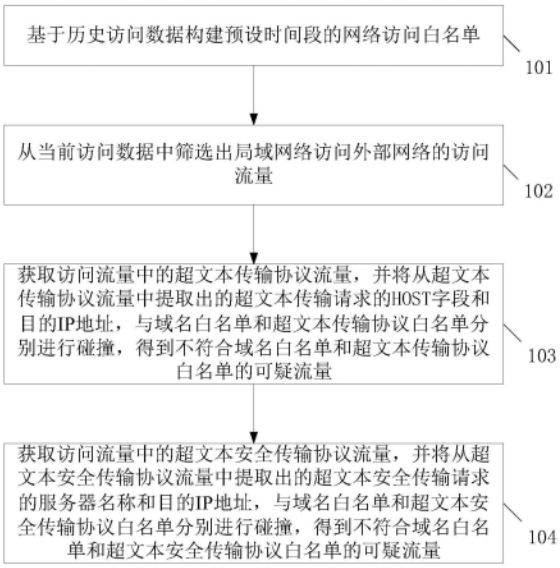
Drawings 4 pages

(54) Name of the invention

A method, apparatus, device, and storage medium for suspicious traffic identification

(57) Summary

The method for recognizing suspicious traffic provided by the present invention can construct a network access whitelist for a predetermined time period based on historical access data, and then filter the access traffic of a local area network accessing an external network from the current access data. The hypertext transfer protocol traffic among the access traffic is obtained, and the HOST field and the destination IP address of the hypertext transfer request extracted from the hypertext transfer protocol traffic are collided with the domain name whitelist and the hypertext transfer protocol whitelist, respectively, to obtain the suspicious traffic that does not conform to the domain name whitelist and the hypertext transfer protocol whitelist. And the server name and



CN 115051845 A

destination IP address of the hypertext secure transmission request extracted from the hypertext secure transmission protocol traffic are collided with the domain name whitelist and said hypertext secure transmission protocol whitelist, respectively, to obtain suspicious traffic that does not conform to the domain name whitelist and the hypertext secure transmission protocol whitelist. The traffic filtering method improves the accuracy of filtering malicious traffic.

1 . A method for identifying suspicious traffic, characterized in that it comprises:

constructing a network access whitelist for a predetermined time period based on historical access data, said network access whitelist comprising at least: a domain name whitelist, a hypertext transfer protocol whitelist, and a hypertext secure transfer protocol whitelist, said domain name whitelist storing domain names of domain name access requests satisfying a first rarity threshold, said hypertext transfer protocol whitelist storing HOST fields and destination IP addresses of hypertext transfer requests satisfying a second rarity threshold; said hypertext secure transfer protocol whitelist storing server names and destination IP addresses of hypertext secure transfer requests satisfying a third rarity threshold; and HOST field and destination IP address of a request, said hypertext secure transfer protocol whitelist storing a server name and destination IP address of a hypertext secure transfer request satisfying a third rarity threshold;

Filter the access traffic of the local area network to the external network from the current access data;

Obtain hypertext transfer protocol traffic in said access traffic and collide the HOST field and destination IP address of the hypertext transfer request extracted from said hypertext transfer protocol traffic with said domain name whitelist and said hypertext transfer protocol whitelist, respectively, to obtain suspicious traffic that does not conform to said domain name whitelist and said hypertext transfer protocol whitelist;

Obtaining Hypertext Secure Transfer Protocol traffic in said access traffic and colliding the server name and destination IP address of the Hypertext Secure Transfer request extracted from said Hypertext Secure Transfer Protocol traffic with said domain name whitelist and said Hypertext Secure Transfer Protocol whitelist, respectively, to obtain suspicious traffic that does not conform to said domain name whitelist and said Hypertext Secure Transfer Protocol whitelist.

2. The method according to claim 1, characterized in that said constructing a network access whitelist for a predetermined time period based on historical access data comprises:

Obtaining domain name access request traffic in said historical access data;

De-duplicating the domain name access requests containing the same domain name within said predetermined time period to obtain the number of source IP addresses in the de-duplicated domain name access requests;

comparing the number of source IP addresses in said de-duplicated domain name access request with said first rarity threshold to obtain the domain names of the domain name access requests that satisfy said first rarity threshold, and storing them in said domain name whitelist.

3. The method according to claim 1, characterized in that said constructing a network access whitelist for a predetermined time period based on historical access data comprises:

Get hypertext transfer request traffic in said historical access data;

De-weighting hypertext transfer requests containing the same HOST field and

destination IP address within said preset time period to obtain the number of source IP addresses in the de-weighted hypertext transfer requests;

The number of source IP addresses in said de-emphasized hypertext transfer request is compared with said second rarity threshold to obtain the HOST field and destination IP address of a hypertext transfer request that satisfies said second rarity threshold and is deposited in said hypertext transfer protocol whitelist.

4. The method according to claim 1, characterized in that said constructing a network access whitelist for a predetermined time period based on historical access data, comprising:

Obtaining hypertext secure transmission request traffic in said historical access data;

De-emphasize hypertext secure transmission requests containing the same server name and destination IP address within said preset time period to obtain the number of source IP addresses in the de-emphasized hypertext secure transmission requests;

Comparing the number of source IP addresses in said de-emphasized hypertext secure transmission request with said third rarity threshold, the server name and destination IP address of the hypertext secure transmission request that satisfies said third rarity threshold are obtained and stored in said hypertext secure transmission protocol whitelist.

5. The method according to claim 1, characterized in that said filtering out access traffic of a local area network accessing an external network from current access data comprises:

Obtaining a list of addresses within a local area network, said list of addresses having IP addresses of devices within the local area network stored therein;

Obtaining a source IP address and a destination IP address in said current access data, if said source IP address in said current access data exists in said address list and said destination IP address does not exist in said address list, then it is determined to be an access traffic of a local area network accessing an external network.

6. A method according to claim 1, characterized in that said first rarity threshold, said second rarity threshold and said third rarity threshold are the same value.

7. A method according to any one of claims 1 to 6, characterized in that it further comprises:

After a refresh of said network access whitelist, the network access whitelist in use is replaced in its entirety based on the updated network access whitelist.

8. A suspicious traffic identification device, characterized in that it comprises:

whitelist module for constructing a network access whitelist for a predetermined time period based on historical access data, said network access whitelist comprising at least: a domain name whitelist, a hypertext transfer protocol whitelist, and a hypertext secure transfer protocol whitelist, said domain name whitelist storing a domain name of a domain name access request that satisfies a first rarity threshold, said hypertext transfer protocol whitelist storing a HOST field and a destination IP address of a hypertext transfer request that satisfies a second rarity threshold, said hypertext transfer protocol whitelist stores a HOST field and a destination IP address for a hypertext transfer request that satisfies a third rarity threshold, and said hypertext secure transfer protocol whitelist stores a server name and a destination IP address for a hypertext secure transfer request that satisfies a third rarity threshold;

An access traffic module for filtering out access traffic of a local area network accessing an external network from current access data;

A first identification module for obtaining hypertext transfer protocol traffic in said access traffic and colliding the HOST field and the destination IP address of the hypertext transfer request extracted from said hypertext transfer protocol traffic with said domain name whitelist and said hypertext transfer protocol whitelist, respectively, to obtain suspicious traffic that does not conform to said domain name whitelist and said hypertext transfer protocol whitelist;and

A second identification module for obtaining hypertext secure transfer protocol traffic in said access traffic and colliding the server name and destination IP address of the hypertext secure transfer request extracted from said hypertext secure transfer protocol traffic with said domain name whitelist and said hypertext secure transfer protocol whitelist, respectively, to obtain suspicious traffic that does not conform to said domain name whitelist and said hypertext secure transfer protocol

whitelist of suspicious traffic.

9. An apparatus characterized in that it comprises: a memory and a processor; said memory, for storing a program;

Said processor for executing said program to implement the various steps of the method for identifying suspicious traffic as claimed in any one of claims 1 to 7.

10. A storage medium having stored thereon a computer program characterized in that said computer program, when executed by a processor, implements the various steps of a method for identifying suspicious traffic as claimed in any one of claims 1 to 7.

A method, apparatus, device, and storage medium for suspicious traffic identification

technical field

[0001] The present invention relates to the field of information security technology, and specifically relates to a method, device, apparatus and storage medium for recognizing suspicious traffic.

background technology

[0002] In the field of signaling security of networks, the initiators of network attacks, in order to hide themselves better, often use some features of HTTP (Hypertext Transfer Protocol) or HTTPS (Hypertext Transfer Protocol Secure) to disguise themselves as large site traffic to avoid detection. Due to the high similarity between malicious traffic and normal network access traffic, it is very easy to bypass the traffic pre-processing mechanism of conventional detection devices, resulting in a large number of missed reports. How to realize efficient and accurate identification of suspicious traffic has become an urgent technical problem.

Content of the invention

[0003] In order to solve the problem of serious underreporting of suspicious traffic that exists in the prior art, the present invention provides a method, device, apparatus, and storage medium for recognizing suspicious traffic, which has features such as more accurate recognition of suspicious traffic.

[0004] A method of recognizing suspicious traffic provided in accordance with specific embodiments of the present invention, comprising:

[0005] constructing a network access whitelist for a predetermined time period based on historical access data, said network access whitelist comprising at least: a domain name whitelist, a hypertext transfer protocol whitelist, and a hypertext secure transfer protocol whitelist, said domain name whitelist storing a domain name for a domain name access request that satisfies a first rarity threshold, said hypertext transfer protocol whitelist storing a HOST field and destination IP address for a hypertext transfer request that satisfies a second rarity threshold of a hypertext transfer request, said HOST field and destination IP address of a hypertext transfer request, and said hypertext secure transfer protocol whitelist stores a server name and destination IP address of a hypertext secure transfer request that satisfies a third rarity threshold;

[0006] Filter the access traffic from the local network to the external network from the current access data;

[0007] obtaining hypertext transfer protocol traffic in said access traffic and colliding the HOST field and destination IP address of the hypertext transfer request extracted from said hypertext transfer protocol traffic with said domain name

whitelist and said hypertext transfer protocol whitelist, respectively, to obtain suspicious traffic that does not conform to said domain name whitelist and said hypertext transfer protocol whitelist;

[0008] obtaining Hypertext Secure Transfer Protocol traffic in said access traffic and colliding the server name and destination IP address of the Hypertext Secure Transfer request extracted from said Hypertext Secure Transfer Protocol traffic with said domain name whitelist and said Hypertext Secure Transfer Protocol whitelist, respectively, to obtain domain name whitelist and said Hypertext Secure Transfer Protocol whitelist that do not match said suspicious traffic.

[0009] Further, said constructing a network access whitelist for a predetermined time period based on historical access data, comprising:

[0010] Obtaining domain name access request traffic in said historical access data;

[0011] de-duplicating domain name access requests containing the same domain name within said predetermined time period to obtain the number of source IP addresses in the de-duplicated domain name access requests;

[0012] comparing the number of source IP addresses in said de-emphasized domain name access request with said first rarity threshold to obtain the domain names of the domain name access requests that satisfy said first rarity threshold, and storing them in said domain name whitelist.

[0013] Further, said constructing a network access whitelist for a predetermined time period based on historical access data, comprising:

[0014] Obtaining hypertext transfer request traffic in said historical access data;

[0015] de-emphasizing hypertext transfer requests containing the same HOST field and destination IP address within said predetermined time period to obtain the number of source IP addresses in the de-emphasized hypertext transfer requests;

[0016] comparing the number of source IP addresses in said de-emphasized hypertext transfer request with said second rarity threshold to obtain a HOST field and a destination IP address of a hypertext transfer request that satisfies said second rarity threshold, and storing it in said hypertext transfer protocol whitelist.

[0017] Further, said constructing a network access whitelist for a predetermined time period based on historical access data, comprising:

[0018] Obtaining hypertext secure transmission request traffic in said historical access data;

[0019] de-duplicating hypertext secure transmission requests containing the same server name and destination IP address within said predetermined time period to obtain the number of source IP addresses in the de-duplicated hypertext secure transmission requests;

[0020] comparing the number of source IP addresses in said de-emphasized hypertext secure transmission request with said third rarity threshold to obtain the server name and destination IP address of a hypertext secure transmission request that satisfies said third rarity threshold, and storing it in said hypertext secure transmission protocol whitelist.

[0021] Further, said filtering out access traffic of a local area network accessing an external network from current access data comprises:

[0022] Obtaining a list of addresses within a local area network, said list of addresses storing IP addresses of devices within the local area network;

[0023] obtaining a source IP address and a destination IP address in said current access data, and if said source IP address in said current access data exists in said address list and said destination IP address does not exist in said address list, determining that it is access traffic from a local area network accessing an external network.

[0024] Further, said first rarity threshold, said second rarity threshold, and said third rarity threshold are the same value.

[0025] Further, said method of identifying suspicious traffic further comprises:

[0026] After a refresh of said network access whitelist, the network access whitelist in use is replaced in its entirety based on the updated network access whitelist.

[0027] A suspicious traffic identification device provided in accordance with specific embodiments of the present invention, comprising:

[0028] A whitelist module for constructing a network access whitelist for a predetermined time period based on historical access data, said network access whitelist comprising at least: a domain name whitelist, a hypertext transfer protocol whitelist, and a

hypertext secure transfer protocol whitelist, said domain name whitelist storing a domain name for a domain name access request that satisfies a first rarity threshold, said hypertext transfer protocol whitelist storing a HOST field and destination IP address for a hypertext transfer request that satisfies a second rarity threshold, said hypertext transfer protocol whitelist stores a HOST field and a destination IP address for a hypertext transfer request that satisfies a third rarity threshold, and said hypertext secure transfer protocol whitelist stores a server name and a destination IP address for a hypertext secure transfer request that satisfies a third rarity threshold;

[0029] Access traffic module for filtering out the access traffic of a local area network accessing an external network from the current access data;

[0030] a first identification module for obtaining hypertext transfer protocol traffic in said access traffic and colliding the HOST field and destination IP address of the hypertext transfer request extracted from said hypertext transfer protocol traffic with said domain name whitelist and said hypertext transfer protocol whitelist, respectively, to obtain domain name whitelist and said hypertext transfer protocol whitelist that do not meet said suspicious traffic; and

[0031] A second identification module for obtaining Hypertext Secure Transport Protocol traffic in said access traffic and transferring the data from the said

The server name and destination IP address of the hypertext secure transmission request extracted from said hypertext secure transmission protocol traffic are collided with said domain name whitelist and said hypertext secure transmission protocol whitelist, respectively, to obtain suspicious traffic that does not conform to said domain name whitelist and said hypertext secure transmission protocol whitelist.

[0032] An apparatus provided according to a specific embodiment of the present invention, comprising: a memory and a processor;

[0033] Said memory for storing a program;

[0034] Said processor for executing said program to implement the various steps of the method for identifying suspicious traffic as described above.

[0035] A storage medium provided according to specific embodiments of the present invention, having a computer program stored thereon, characterized by In that said computer program, when executed by the processor, implements the various steps of the method for identifying suspicious traffic as described above.

[0036] The suspicious traffic identification method provided by the present invention may construct a network access whitelist for a predetermined time period based on historical access data, wherein the network access whitelist comprises at least a domain name whitelist, a hypertext transfer protocol whitelist, and a hypertext secure transfer protocol whitelist, wherein the domain name whitelist stores domain names of domain name access requests that satisfy a first rarity threshold, the hypertext transfer protocol whitelist stores HOST field and destination IP address of a hypertext transfer request that satisfies a second rarity threshold, and the hypertext secure transfer protocol whitelist stores a server name and destination IP address of a hypertext secure transfer request that satisfies a third rarity threshold. The access traffic of the local area network accessing the external network is then filtered from the current access data. The hypertext transfer protocol traffic in the access traffic is obtained, and the HOST field and the destination IP address of the hypertext transfer request extracted from the hypertext transfer protocol traffic are collided with the domain name whitelist and the hypertext transfer protocol whitelist, respectively, to obtain the suspicious traffic that does not meet the domain name whitelist and the hypertext transfer protocol whitelist. Obtaining hypertext secure transfer protocol traffic in the access traffic and colliding the server name and destination IP address of the hypertext secure transfer request extracted from the hypertext secure transfer protocol traffic with the domain name whitelist and said hypertext secure transfer protocol whitelist, respectively, to obtain suspicious traffic that does not conform to the domain name whitelist and the hypertext secure transfer protocol whitelist. The traffic filtering method is filtered by separate whitelists for the hypertext transfer protocol and the hypertext secure transfer protocol, and based on the whitelists using the domain name with the corresponding destination IP address as the filtering object, which improves the accuracy of filtering the malicious traffic and reduces the underreporting.

illustrate

[0037] In order to more clearly illustrate the technical solutions in the embodiments or prior art of the present invention, the accompanying drawings that need to be used in the description of the embodiments or prior art will be briefly introduced below, and it will be obvious that the accompanying drawings in the following description are only embodiments of the present invention, and that other drawings can be obtained according to the accompanying drawings provided for a person of ordinary skill in the field, without creative labor. The accompanying drawings.

[0038] FIG. 1 is a flowchart of a method for recognizing suspicious traffic according to an exemplary embodiment;

[0039] FIG. 2 is a flowchart of the construction of a domain name whitelist provided according to an exemplary embodiment;

[0040] FIG. 3 is a flowchart of the construction of a hypertext transfer protocol whitelist provided according to an exemplary embodiment;

[0041] FIG. 4 is a flowchart of the construction of a hypertext secure transmission protocol whitelist provided according to an exemplary embodiment;

[0042] FIG. 5 is a flowchart of the acquisition of access traffic provided according to an exemplary embodiment;

[0043] FIG. 6 is a structural diagram of a suspicious traffic identification device according to an exemplary embodiment;

[0044] FIG. 7 is a structural diagram of an apparatus provided according to an exemplary embodiment.

practical way of doing sth.

[0045] The technical solutions in embodiments of the present invention will be clearly and completely described below in conjunction with the accompanying drawings in embodiments of the present invention

Described in its entirety, it is clear that the described embodiments are only a part of the embodiments of the present invention and not all of the embodiments. Based on the embodiments of the present invention, all other embodiments obtained by a person of ordinary skill in the art without making creative labor are within the scope of protection of the present invention.

[0046] Referring to FIG. 1, embodiments of the present invention provide a method of recognizing suspicious traffic, which may include the following steps:

[0047] 101, constructing a network access whitelist for a predetermined time period based on historical access data, the network access whitelist comprising at least: a domain name whitelist, a hypertext transfer protocol whitelist, and a hypertext secure transfer protocol whitelist, the domain name whitelist storing a domain name of a domain name access request that satisfies a first rarity threshold, the hypertext transfer protocol whitelist storing a second rarity threshold of a HOST field (request header field) and destination IP address of a hypertext transfer request, and the hypertext secure transfer protocol whitelist stores the server name and destination IP address of a hypertext secure transfer request that satisfies a third rarity threshold.

[0048] A whitelist is the opposite of a "blacklist". For example, in a computer system, there is a lot of software that applies black and white list rules, operating systems, firewalls, antivirus software, email systems, application software, etc., and almost all of them apply black and white list rules when it comes to control.

[0049] When blacklisting is enabled, users (or IP addresses, IP packets, emails, viruses, etc.) that are on the blacklist cannot pass through. If a whitelist is set up, users (or IP addresses, IP packets, emails, etc.) that are in the whitelist are prioritized to pass through and are not rejected as spam, making it much safer and faster. It is based on this characteristic of whitelist. Based on historical access data on the network to establish a preset time period of network access whitelist, such as within one hour of the network access whitelist. The network access whitelist may include a domain name whitelist, a hypertext transfer protocol whitelist, and a hypertext secure transfer protocol whitelist. Wherein the domain name whitelist is a statistic for each DNS request, i.e., a domain name of a domain name access request, in the DNS protocol traffic in the network, and its statistic is conditioned on the domain name of the domain name access request having a rarity threshold in the one-hour statistic time that cannot be lower than a first rarity threshold, wherein the setting of the rarity threshold is set on the basis of the principle of the principle that the more domain names accessed in the DNS requests in a finite network are more secure The rarity threshold is set here based on the principle that the more domain names accessed in a DNS request within a limited network, the safer it is. Based on the same principle for the hypertext transfer protocol whitelist there is stored the HOST field and the destination IP address of the hypertext transfer request that satisfies the second rarity threshold, and the hypertext secure transfer protocol whitelist there is stored the server name and the destination IP address of the hypertext secure transfer request that satisfies the third rarity threshold. Wherein the HOST field of the hypertext transfer request is the domain name or the destination IP address to be accessed.

[0050] 102. filter the access traffic of the local area network to the external network from the current access data.

[0051] Since the security of the network is mainly focused on the security of the intranet, i.e., the local area network's internal access to the external interconnected network, the internal to internal access to each other can be left unfiltered. Therefore there is a need to filter the access traffic of the local area network accessing the external network from the current access data. The scrubbing of access traffic can be done based on the built-in intranet address database, for example for example IPV4 as: 10 .0 .0 . /8 ,172 .16 .0 .0/12 ,192 .168 .0 .0/16,IPV6 is local with IPV6 unicast addresses (including link local Unicast address and site-local unicast address) and manually configured intranet address information. Based on the destination IP address and source IP address of the access traffic, you can get the access traffic from the intranet to the extranet.

[0052] 103, obtaining the hypertext transfer protocol traffic in the access traffic, and colliding the HOST field and destination IP address of the hypertext transfer request extracted from the hypertext transfer protocol traffic with the domain name whitelist and the hypertext transfer protocol whitelist, respectively, to obtain suspicious traffic that does not conform to the domain name whitelist and the hypertext transfer protocol whitelist.

[0053] De-collide the domain whitelist with the HOST+destination IP combination in Hypertext Transfer Protocol, and traffic hitting this whitelist is dropped directly.

[0054] De-collision Hypertext Transfer Protocol whitelist with HOST+destination IP, traffic hitting this whitelist is dropped directly.

[0055] 104, obtaining the HTTP traffic in the access traffic and colliding the server name and destination IP address of the HTTP request extracted from the HTTP traffic with the domain name whitelist and the HTTP whitelist, respectively, to obtain suspicious traffic that does not conform to the domain name whitelist and the HTTP whitelist.

[0056] De-collision domain whitelisting with SNI (Server Name) + Destination IP in Hypertext Secure Transport Protocol access traffic, traffic hitting this whitelist is dropped directly.

[0057] De-collide the Hypertext Secure Transport Protocol whitelist with SNI+destination IP, and traffic hitting this whitelist is dropped directly.

[0058] After obtaining the access traffic, the access traffic is collided in a combination of destination IP address and domain name, and the HTTP traffic that does not conform to the above domain name whitelist and the hypertext transfer protocol whitelist then enters into the subsequent suspicious traffic monitoring module, as well as the suspicious traffic that does not conform to the domain name whitelist and the hypertext secure transfer protocol whitelist enters into the subsequent suspicious traffic monitoring module The HTTP protocol and HTTPS white lists are then processed for further isolation and confirmation. So that the HTTP protocol and HTTPS protocol have separate whitelists and do not interfere with each other. As most websites have switched to HTTPS encryption protocol, there is already a big difference between normal traffic under HTTP and HTTPS protocols, so the whitelist of HTTP and HTTPS protocols is differentiated and the filtering is more refined. Moreover, in the filtering mechanism of whitelist, instead of pure IP or domain name filtering, the domain name as well as the destination IP address are used as objects for filtering, which effectively avoids bypassing the traffic pre-filtering system by disguising the HTTP HOST or Domain Borrowing and other types of malicious traffic, and improves the accuracy of the identification of suspicious traffic.

[0059] As a feasible implementation of the above embodiment, with reference to FIG. 2, the process of constructing a domain name whitelist may include

The following steps:

[0060] 201. Get domain access request traffic in historical access data.

[0061] 202, de-duplication of domain name access requests containing the same domain name within a preset time period, and obtaining the number of source IP addresses in the de-duplicated domain name access requests.

[0062] 203, the number of source IP addresses in the de-duplicated domain name access request is compared with the first rarity threshold to obtain the domain name of the domain name access request that satisfies the first rarity threshold and is deposited in the domain name whitelist.

[0063] Specifically, a DNS request domain name, a destination IP address, and a source IP that initiated the DNS request are recorded for DNS protocol traffic in the network, forming a DNS domain name + destination IP + source IP record and entering it into a database. The number of de-emphasized source IPs of each DNS domain name within an

hour is counted, and a certain rarity threshold (which may generally be greater than 90) is set for the number of source IPs, and DNS domain names that meet the condition of not being rare are calculated and added to a domain name whitelist.

[0064] Referring to FIG. 3, the process of constructing a hypertext transfer protocol whitelist may include the following steps:

[0065] 301, Getting hypertext transfer request traffic in historical access data .

[0066] 302, de-emphasizing hypertext transfer requests containing the same HOST field and destination IP address within a predetermined time period, obtaining a number of source IP addresses in the de-emphasized hypertext transfer requests.

[0067] 303, the number of source IP addresses in the de-duplicated hypertext transfer request is compared to the second rarity threshold to obtain the HOST fields and destination IP addresses of the hypertext transfer requests that satisfy the second rarity threshold and are stored in the hypertext transfer protocol whitelist.

[0068] Specifically, HOST values, corresponding destination IP addresses, and source IP addresses of HTTP protocol traffic of GET and POST methods in a network are recorded to form a HOST+destination IP+source IP record and put it into a database. Counting the number of de-duplicated source IPs based on HOST+destination IP pairs in an hour, setting a second rarity threshold (greater than 90) in terms of the number of source IP addresses, calculating the HOST+destination IPs that meet the condition of being not rare, and adding them to the hypertext transfer protocol whitelist (which is based on the number of accesses to the

The reason that the number of IPs as objects is calculated to be not rare is based on the fact that the more domains that are accessed in an HTTP request within a finite network the safer it is, and the reason for the restriction to HOST+destination IPs is to prevent certain cases of forged HOSTs).

[0069] Referring to FIG. 4, the process of constructing a hypertext secure transport protocol whitelist may include the following steps:

[0070] 401, Get hypertext secure transfer request traffic in historical access data.

[0071] 402, the hypertext secure transmission requests containing the same server name and destination IP address within the predetermined time period are de-duplicated to obtain the number of source IP addresses in the de-duplicated hypertext secure transmission requests.

[0072] 403, the number of source IP addresses in the de-duplicated hypertext secure transmission request is compared to the third rarity threshold to obtain the server name and destination IP address of the hypertext secure transmission request that satisfies the third rarity threshold and is deposited in the hypertext secure transmission protocol whitelist.

[0073] By recording the SNI (server name) of the HTTPS protocol in the network, as well as the corresponding destination IP address and the source IP address that initiated the HTTPS request, an SNI+destination IP address+source IP address record is formed and put into the database. The number of source IP addresses corresponding to each SNI in an hour is counted, and a certain rarity threshold (greater than 90) is set for the number of source IP addresses, and SNIs + destination IP addresses that meet the condition of being non-rare are counted and added to the HTTPS protocol whitelist.

[0074] It will be appreciated that the first rarity threshold, the second rarity threshold, and the third rarity threshold described above are the same value or may be different values, and the present invention is not limited herein.

[0075] In some specific embodiments of the present invention, filtering the access traffic of a local area network accessing an external network from the current access data as shown in FIG. 5 may include the following steps:

[0076] 501. obtaining a list of addresses within the local area network, the list of addresses storing IP addresses of devices within the local area network.

[0077] 502, the source IP address and the destination IP address in the current access data are obtained, and if the source IP address in the current access data exists in the address list and the destination IP address does not exist in the address list, it is determined to be the access traffic of the local area network accessing the external network.

[0078] Specifically, the source IP address and destination IP address in the traffic data are extracted, and it is judged that if the source IP address is in the intranet address list and the destination IP address is an address in the non-intranet address list, it is judged to be the traffic of an intranet host accessing the public network. Then according to all the received traffic, if it does not meet the above

rules of the intranet access to the public network is determined that it is not the data that needs to be detected by this detection model, it is directly discarded. And the traffic that satisfies the above intranet access to the public network is retained.

[0079] Where after the refresh of the network access whitelist, the network access whitelist in use is replaced in its entirety based on the updated network access whitelist. That is, the whitelist refresh mechanism uses a full replacement scheme, i.e. the new whitelist replaces the old whitelist list that is being used for traffic filtering in its entirety, to address the situation where after a change in the network link, the address of the same domain name after the same domain name has been resolved has changed, i.e. there is no longer a traffic match.

[0080] Based on the same design idea, referring to FIG. 6, embodiments of the present invention also provide a suspicious traffic identification device, which may perform various steps of the suspicious traffic identification method described in the above embodiments, which may include:

[0081] A whitelist module 601 for constructing a network access whitelist for a predetermined time period based on historical access data, the network access whitelist comprising at least a domain name whitelist, a hypertext transfer protocol whitelist, and a hypertext secure transfer protocol whitelist, the domain name whitelist storing domain names of domain name access requests that satisfy a first rarity threshold, the hypertext transfer protocol whitelist storing a second rarity threshold of HOST field and destination IP address for hypertext transfer requests, the hypertext secure transfer protocol whitelist stores the server name and destination IP address for hypertext secure transfer requests that satisfy a third rarity threshold.

[0082] An access traffic module 602 for filtering out access traffic of a local area network accessing an external network from current access data.

[0083] A first identification module 603 is used to obtain hypertext transfer protocol traffic in the accessed traffic and collide the HOST field and the destination IP address of the hypertext transfer request extracted from the hypertext transfer protocol traffic with a domain name whitelist and a hypertext transfer protocol whitelist, respectively, to obtain suspicious traffic that does not conform to the domain name whitelist and the hypertext transfer protocol whitelist. As well as

[0084] A second identification module 604 is used to obtain hypertext secure transfer protocol traffic in the access traffic and collide the server name and destination IP address of the hypertext secure transfer request extracted from the hypertext secure transfer protocol traffic with the domain name whitelist and the hypertext secure transfer protocol whitelist, respectively, to obtain the domain name whitelist and the hypertext secure transfer protocol whitelist that do not meet the Suspicious traffic.

[0085] The device has the same beneficial effects as the method of recognizing suspicious traffic described above, and the present invention will not be repeated herein.

[0086] Referring to FIG. 7, embodiments of the present invention also provide an apparatus comprising: a memory 701 and a processor 702;

[0087] Memory 701 for storing the program;

[0088] A processor 702 for executing a program to implement various steps of a method for identifying suspicious traffic as described above.

[0089] Embodiments of the present invention also provide a storage medium having stored thereon a computer program that, when executed by a processor, implements the steps of the method of identifying suspicious traffic as described above.

[0090] The method, apparatus, device, and storage medium for recognizing suspicious traffic provided in the above embodiments of the present invention, in the whitelist filtering mechanism, instead of pure IP or domain name filtering, the domain name as well as the returned (corresponding) IP address are used as objects to be filtered, which effectively avoids the bypassing of malicious traffic of the type of masquerading as HTTP HOST or Domain Borrowing, etc., by the Traffic pre-filtering system.

[0091] For each of the foregoing method embodiments, for the sake of simplicity of description, they are all expressed as a series of combinations of actions, but a person skilled in the art should be aware that the present invention is not limited by the order of the described actions, as certain steps may be carried out in an alternative order or at the same time, in accordance with the present invention. Secondly, the person skilled in the art should also be aware that the embodiments described in the specification are all preferred embodiments, and the actions and modules involved are not necessarily necessary for the present invention.

[0092] It should be noted that each embodiment in this specification is described in an incremental manner, and each embodiment focuses on the differences with other embodiments, and the same and similar portions of each embodiment can be seen in each other. For the device embodiments, since they are basically similar to the method embodiments, the descriptions are relatively simple, and it is sufficient to refer to part of the

description of the method embodiments for the relevant points.

[0093] The steps in the method of each embodiment of the present invention may be sequentially adjusted, combined and deleted according to actual needs, and the technical features recorded in each embodiment may be replaced or combined.

[0094] The modules and sub-modules in the various embodiments of the present invention species devices and terminals may be combined, divided and deleted according to actual needs.

[0095] In the several embodiments provided by the present invention, it should be understood that the disclosed terminals, devices, and methods, may be realized in other ways. For example, the terminal embodiments described above are merely schematic, e.g., the division of modules or sub-modules is merely a logical functional division, and the actual implementation may be divided in other ways, e.g., a plurality of sub-modules or modules may be combined or may be integrated into another module, or some features may be ignored, or not implemented. Another point is that the mutual coupling or direct coupling or communication connection shown or discussed may be an indirect coupling or communication connection through some interface, device or module, which may be electrical, mechanical or other forms.

[0096] The modules or sub-modules illustrated as separate components may or may not be physically separate, and the components that are modules or sub-modules may or may not be physical modules or sub-modules, i.e., they may be located in one place, or

It may also be distributed to a plurality of network modules or sub-modules. Some or all of these modules or sub-modules may be selected to fulfill the purpose of this embodiment scheme according to actual needs.

[0097] Alternatively, the various functional modules or sub-modules in various embodiments of the present invention may be integrated in a single processing module, or each module or sub-module may physically exist separately, or two or more modules or sub-modules may be integrated in a single module. The above integrated modules or sub-modules may be implemented either in the form of hardware or in the form of software functional modules or sub-modules.

[0098] The professional may further realize that the units and algorithmic steps of the various examples described in conjunction with the embodiments disclosed herein are capable of being implemented in electronic hardware, computer software, or a combination of the two, and that the composition and steps of the various examples have been described in the above description in general terms according to function for the sake of clarity as to the interchangeability of hardware and software. Whether these functions are performed in hardware or software depends on the particular application and design constraints of the technical solution. The skilled artisan may use different methods to implement the described functions for each particular application, but such implementations should not be considered outside the scope of the present invention.

[0099] The steps of the methods or algorithms described in conjunction with the embodiments disclosed herein may be implemented directly with hardware, a software unit executed by a processor, or a combination of both. The software units may be placed in random memory (RAM), memory, read-only memory (ROM), electrically programmable ROM, electrically erasable programmable ROM, registers, hard disks, removable disks, CD-ROMs, or any other form of storage medium known in the art.

[0100] Finally, it should also be noted that, in this document, relational terms such as first and second are used only to distinguish one entity or operation from another, and do not necessarily require or imply the existence of any such actual relationship or order between those entities or operations. Furthermore, the terms "including", "comprising", or any other variant thereof, are intended to cover non-exclusive inclusion, such that a process, method, article, or apparatus comprising a set of elements includes not only those elements, but also other elements not expressly listed, or other elements that are not expressly listed for the purpose of such a process, method, article or apparatus, or other elements that are not expressly listed for the purpose of such a process, method, article or equipment. elements, or also includes elements that are inherent to such process, method, article or apparatus. Without further limitation, the fact that an element is defined by the phrase "includes a ....." does not preclude the existence of additional identical elements in the process, method, article or apparatus that includes said element.

[0101] The foregoing description of the disclosed embodiments enables those skilled in the art to realize or use the present invention. Various modifications to these embodiments will be apparent to those skilled in the art, and the general principles defined herein may be realized in other embodiments without departing from the spirit or scope of the present invention. Accordingly, the present invention will not be limited to

these embodiments shown herein, but will be subject to the broadest scope consistent with the principles and novel features disclosed herein.

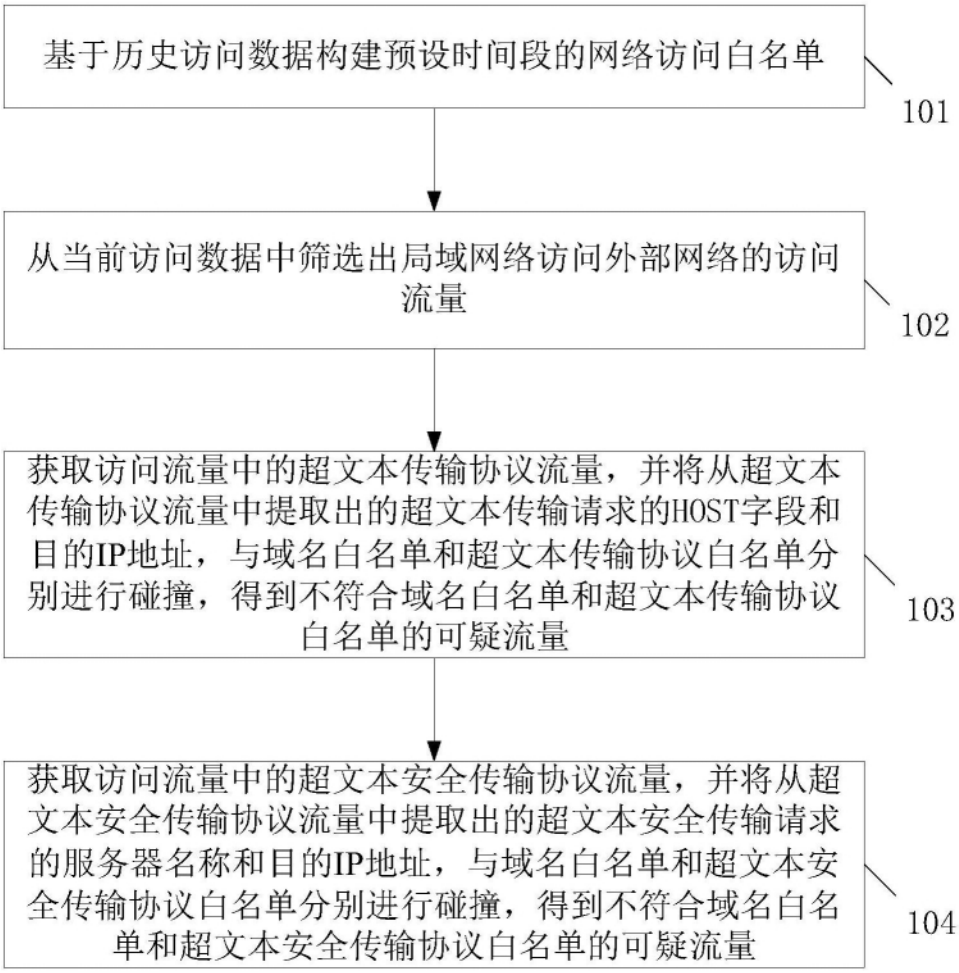
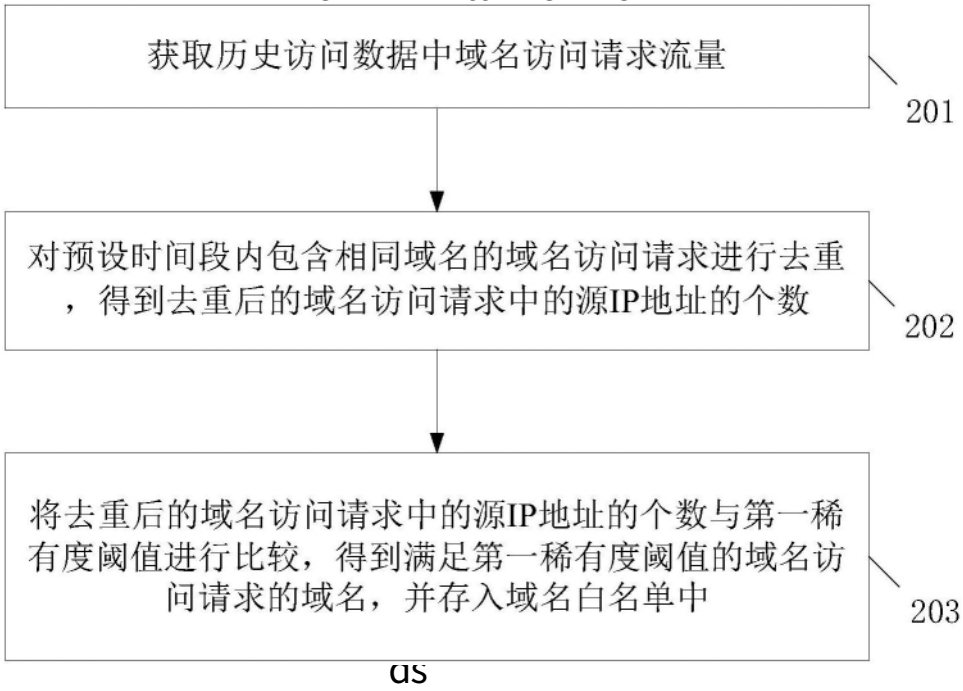

基于历史访问数据构建预设时间段的网络访问白名单
101
从当前访问数据中筛选出局域网络访问外部网络的访问流量
102
获取访问流量中的超文本传输协议流量，并将从超文本传输协议流量中提取出的超文本传输请求的HOST字段和目的IP地址，与域名白名单和超文本传输协议白名单分别进行碰撞，得到不符合域名白名单和超文本传输协议白名单的可疑流量
103
获取访问流量中的超文本安全传输协议流量，并将从超文本安全传输协议流量中提取出的超文本安全传输请求的服务器名称和目的IP地址，与域名白名单和超文本安全传输协议白名单分别进行碰撞，得到不符合域名白名单和超文本安全传输协议白名单的可疑流量
104

Figure 1



ds

Figure 2

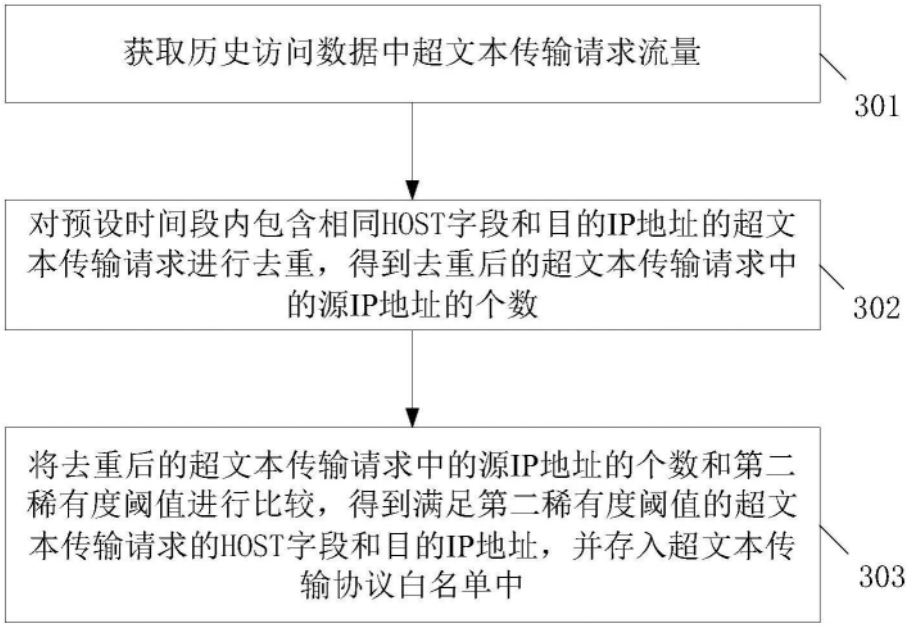


Figure 3

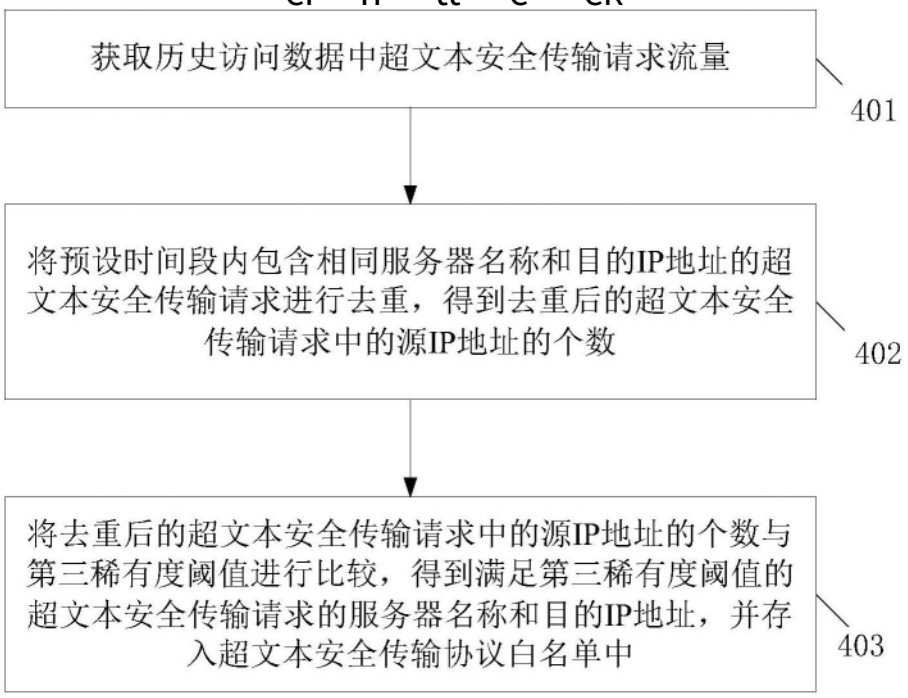


Figure 4

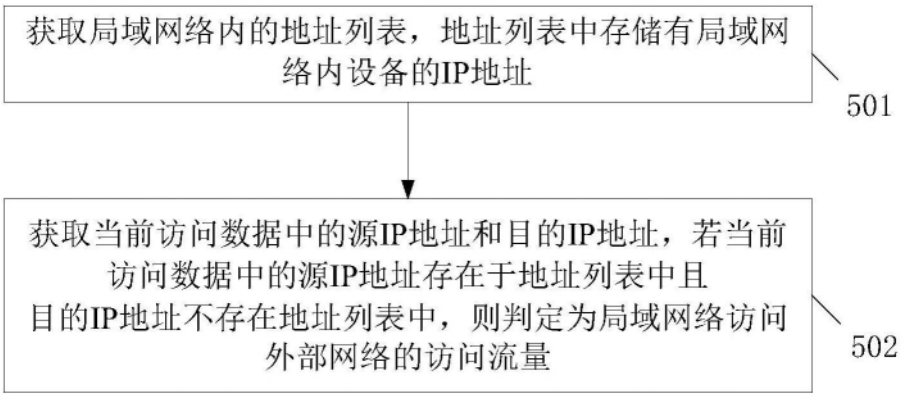


Figure 5

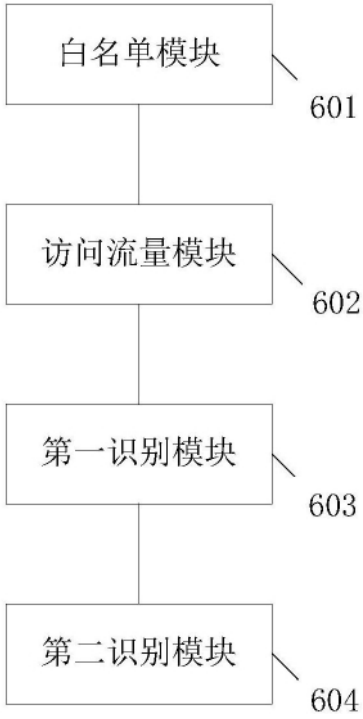


Figure 6

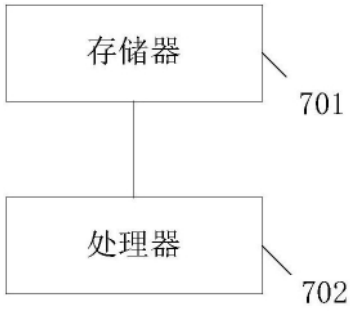


Figure 7