

Systems And Methods For Intercepting, Filtering and Blocking Content From Internet in Real Time

5 Field of Invention

Embodiments of the present invention generally relate to computer and digital processing systems for network monitoring, and more particularly, to systems and methods for intercepting, filtering and blocking content from internet in real
10 time.

Background

In today's world, it is not uncommon for legal regulatory authorities and
15 governments of various countries to monitor a great deal of internet activity with a view to determine illegal or terrorist activities which has escalated due to global terrorism and criminal activities on the Web in recent years.

Nowadays, the web provides a platform for a plethora of illegal activities ranging from credit card fraud, through identity phishing, to transferring money and
20 orders. Among multiple illegal activities, threat posed by illegal websites is highly detrimental and hence, blocking such illegal websites is the need of the hour. Examples of such illegal websites vary for different jurisdictions depending upon local laws and regulations. For example, gambling websites are illegal in Singapore, child pornography websites are illegal in UK, and so
25 on.

One possible way to block such illegal websites is to block them at the level of intermediaries and ISPs (internet service providers). Such blocking usually requires regulatory backing and consequently, the action varies for each jurisdiction. Moreover, the infrastructure of a large Internet Service Provider (ISP) or Network Service Provider (NSP) typically comprises a constantly growing network of heterogeneous routers interconnecting millions of customer-devices. This network enables the network customers to exchange data of various formats created and consumed by a number of applications.

In recent years it has be observed that customers obtain electronic threats (eThreats) mainly from the internet. For example, eThreats comprise a variety of attacks which can be classified into three main categories: worm-related, non-worm related (e.g., virus, Trojan), and probes (e.g., spyware, adware, identity theft, and phishing).

Furthermore, Network Service Provider (NSP) are vulnerable to eThreats propagated across networks. Generally, information pertaining to illegal websites may be collected and analyzed by retrieving multiple packets and determining if such packets contain such information pertaining to illegal websites.

Subsequently, information pertaining to such illegal sites may be categorized from which corresponding packets are transmitted if the analyzed packets include harmful information, and such information can be stored in a database for further action.

However, such technologies are mostly based on signatures and databases, and consequently, corresponding analysis requires lot of time, which may be hours, days or may be weeks, or even longer in some cases.

Therefore, such delays prevent real time monitoring and analysis of illegal websites. Bloom Filters is based on the signature-based approach and are used to build a system that scans Internet traffic. Packets enter the system and are processed by Internet Protocol (IP) wrappers. The data in the packet goes to the input buffer and then flows through the content pipeline. As the packet passes through the pipeline, multiple Bloom engines scan different window lengths for signatures of different lengths. Subsequently, data leaves the content pipeline, flows to the output buffer, streams through the wrappers, and then packets are re-injected into the network. If a Bloom engine detects a match, a hash table is queried to determine if an exact match occurred. If the queried signature is an exact match, the malicious content can be blocked and an alert message is generated within a User Datagram Protocol (UDP) packet, informing a network administrator, an end-user or an automated process that a matching signature has been detected. Even though, the bloom filters provide a solution focused on throughput performance criteria, but can only deal with certain kinds of known eThreats that can be identified by their hashing or regular expression signature.

As may be seen, the solutions described above provide a partial solution to a very small part of a particular problem. Moreover, they don't block the content online in real time.

Accordingly, there exists a need in the art for systems and methods for intercepting, filtering and blocking content from internet in real time.

Summary of the Invention

According to an aspect of the present invention, a system for intercepting, filtering and blocking content from internet in real time, includes, a filtering device mounted at a network service provider for retrieving and filtering web content in real-time to generate filtered content. In use, the filtering device is
5 capable of transmitting the filtered content for decoding and reconstruction. In addition, the system further includes a collection device mounted at a law enforcement agency for receiving the filtered content for further classification, decoding and reconstruction.

According to another aspect of the present invention, a method for intercepting,
10 filtering and blocking content from internet in real time includes the steps of, retrieving web content including multiple packets from a source, stripping headers and abstracting information such as MAC addresses, IP addresses and the like from the packets by a dissector; grouping packets belonging to same flow together by a flow manager; determining a protocol to be used for
15 each group of the packets by a protocol checker; abstracting application layer data by a protocol decoder and grouping the application layer data into multiple layer 7 applications; and sending the application layer data for further processing.

20 **Brief Description of Drawings**

In the drawings, like reference characters generally refer to the same part throughout the different views. The drawings are not necessary to scale, emphasis instead generally being placed upon illustrating the principles of the invention. In the following description, various embodiments are described with
25 reference to the following drawings, in which:

Fig. 1 illustrates system architecture of a system for intercepting, filtering and blocking content from internet in real time, according to an embodiment of the present invention;

Fig. 2 illustrates system architecture of a system for intercepting, filtering and blocking content from internet in real time, according to another embodiment of the present invention; and

Fig. 3 illustrates system architecture of a filter of the system for intercepting, filtering and blocking content from internet in real time, according to an embodiment of the present invention.

10

Detailed Description

Various embodiments of the present invention relate to systems and methods for intercepting, filtering and blocking content from internet in real time. More specifically, a network tap is placed in ISP (Exchange Router/Gateway Router/Border Router) before the Internet traffic leaves the National Gateway. Consequently, the network tap taps and sends traffic for decoding and reconstruction in real-time.

Fig. 1 illustrates system architecture of a system 120 for intercepting, filtering and blocking content from internet in real time. In one embodiment, the system 120 may be employed between law enforcement agency 115 and one or more internet service providers (ISPs) 125₁, 125₂, 125₃, and the like. As may be seen, the internet 105 is accessed via a national gateway 110, which makes possible the use of lawful interception solutions as provided by various embodiments of the present invention.

25

Fig. 2 illustrates system architecture of a system 200 for intercepting, filtering and blocking content from internet in real time. In accordance with an embodiment of the present invention, the system 200 for intercepting, filtering and blocking content from internet in real time, includes, a filtering device 205
5 mounted at a network service provider for retrieving and filtering web content in real-time to generate filtered content. In use, the filtering device is capable of transmitting the filtered content for decoding and reconstruction.

In accordance with an embodiment of the present invention, the system 200 further includes a collection device 210 mounted at a law enforcement agency
10 for receiving the filtered content for further classification, decoding and reconstruction.

In accordance with an embodiment of the present invention, the filtering device 205 includes a fabric switch 212 for retrieving and distributing the web content, and, multiple filters 215 for receiving the web content from the fabric switch 212,
15 filtering the web content based on multiple parameters to generate the filtered content and forwarding the filtered content to the collection device 210. In use, traffic of multiple 10Gbps/40Gbps/100Gbps can be tapped and spanned to the fabric switch 212, which can then distribute the traffic by 10Gbps ports to multiple filters 215. Generally, the multiple filters 215 are configured to handle
20 up to 160Gbps line rate traffic throughput. In addition, the multiple filters 215 are capable of layer 7 header deep packet inspection (DPI) and payload content filtering.

In accordance with an embodiment of the present invention, the multiple parameters to generate the filtered content include keywords, REGEX patterns,
25 IP addresses, URLs and the like. In use, such as, for example, multiple pre-

configured keywords for online betting such as online betting, football betting, horse racing betting, football odds, betting odds, bookmakers, Asian handicaps, and the like may be pre-configured or profiled into one or more filters where traffic is tapped at the ISP. Subsequently, when there is any keyword
5 whose condition is met, the traffic is filtered out and sent for real-time decoding and reconstruction.

Those of ordinary skill in the art will appreciate that multiple keywords, regex patterns and the like can be added from time to time to improve the multiple filters 215 capability. Furthermore, custom filtering signatures or keywords can
10 be added which in time will grow providing greater visibility and accuracy to 205. In addition, various third party filtering signatures can also be used together with the keywords, regex patterns and the like to enhance the multiple filters 215.

Furthermore, on those encoded data packets that the payload content cannot
15 be inspected by the multiple filters 215, the filtering device will pass the packets to the Collection Device 210 where the reconstruction unit 230 will make use of real time packet reconstruction (RTPR) to process the packets and reconstruct the content of the packet payloads, enabling the content to be decoded. Consequently, the visibility is wider and accurate. In use, this also covers future
20 methodologies on how the content are transported as the world moves to Web 2.0.

In accordance with an embodiment of the present invention, the collection device 210 includes, a classification unit 225 for decoding and classifying multiple packets contained in the filtered content in accordance with L7

protocols, and, a reconstruction unit 230 for reconstructing and retaining the packets by employing real time packet reconstruction (RTPR).

In accordance with an embodiment of the present invention, the collection device 210 further includes one or more application programming interfaces (API) 235 for using the packets obtained after RTPR for further action. In use, the API includes content filtering, content blocking 245, warrant management 240 and the like.

In accordance with an embodiment of the present invention, the system 200 further includes a mediation device 220 for passing tapped traffic from at least one telecom service provider (ISP) domain to the collection device 210 in LEA domain with certain standard format matched.

Fig. 3 illustrates system architecture of the filter 215 of the system 200 for intercepting, filtering and blocking content from internet in real time, according to an embodiment of the present invention. In accordance with an embodiment of the present invention, each filter 215 includes a High Throughput Network Processor Unit (NPU) for data processing. In use, multiple 10Gbps (up to 40Gbps) traffic may be spanned to the input ports of each filter. Subsequently, the traffic travels through the fabric switch, which is then sent to NPU for processing. Thereafter, the NPU performs the packet filtering according to pre-configured/defined keywords patterns, as explained above.

In accordance with an embodiment of the present invention, the system further includes a dissector for stripping headers and abstracting information such as MAC addresses, IP addresses and the like from the web content. In use, the whole list of sites with details like source-destination IP addresses, URL links and even content is made viewable to the LEA. Subsequently, the LEA can

then have the call to decide to block which URL or IP address through the API integration with ISP other network security equipment, such as, for example, the content filtering and blocking module.

In accordance with an embodiment of the present invention, the system further
5 includes a flow manager for grouping the web content into multiple packets belonging to same flow.

In accordance with an embodiment of the present invention, the system further includes a protocol checker for determining a protocol to be used for each group of the packets. In use, the protocol checker includes HTTP, POP3, SMTP, and
10 the like.

In accordance with an embodiment of the present invention, the system further includes a protocol decoder for abstracting application layer data and grouping the application layer data into multiple layer 7 applications.

In accordance with an embodiment of the present invention, the system further
15 includes at least one database module for storing the filtered content.

In accordance with an embodiment of the present invention, the system further includes a web management GUI for accessing the at least one database module.

In accordance with an embodiment of the present invention, a method for
20 intercepting, filtering and blocking content from internet in real time includes the steps of, retrieving web content including multiple packets from a source, stripping headers and abstracting information such as MAC addresses, IP addresses and the like from the packets by a dissector; grouping packets belonging to same flow together by a flow manager; determining a protocol to
25 be used for each group of the packets by a protocol checker; abstracting

application layer data by a protocol decoder and grouping the application layer data into multiple layer 7 applications; and sending the application layer data for further processing.

In accordance with an embodiment of the present invention, the method further
5 includes the step of determining if further decoding is needed.

In accordance with an embodiment of the present invention, the method further includes the step of sending data to application data recorder if further decoding is needed.

In accordance with an embodiment of the present invention, the method further
10 includes the step of saving data in at least one database module.

In accordance with an embodiment of the present invention, a method for intercepting, filtering and blocking content from internet in real time includes the steps of, spanning the tapped traffic by multiple 10Gbps port to the filter. If it is determined that the traffic is encrypted, the filter dumps the traffic. However, if
15 the traffic is not encrypted, the filter determined whether it is Web HTTP traffic or not. In case it is not Web traffic, it is dumped and only Web HTTP traffic is retained. In due course, the filter checks whether it is encoded or not. In case the traffic is not encoded, the filter processes the traffic using the L7 DPI Filtering capability to filter out the traffic of interest based on keywords patterns
20 and the like, as explained above. In case of encoded traffic, it is bypassed from the L7 DPI Filtering process and directly sent for decoding and reconstruction, which enables the LEA to view the traffic that contains the target keywords patterns using the Web Management GUI display.

Therefore, as may be seen, various embodiments of the present invention
25 provide significant advantages, such as, for example, but not limited to, real-

time decoding and reconstruction that can allow the illegitimate sites objects to be identified (displayed) and law enforcement agency (lea) can instantly decide to make use of the API integration with content filtering or blocking system to block such sites or content.

Claims

1. A system for intercepting, filtering and blocking content from internet in real time, said system comprising:

a filtering device mounted at a network service provider for retrieving and filtering web content in real-time to generate filtered content, said filtering device
5 being capable of transmitting said filtered content for decoding and reconstruction; and,

a collection device mounted at a law enforcement agency for receiving said filtered content for further classification, decoding and reconstruction.

10 2. The system as claimed in Claim 1, wherein said filtering device comprises:

a fabric switch for retrieving and distributing said web content; and,

a plurality of filters for receiving said web content from said fabric switch, filtering said web content based on a plurality of parameters to generate said
15 filtered content and forwarding said filtered content to said collection device.

3. The system as claimed in Claim 2, wherein said plurality of filters are capable of layer 7 header deep packet inspection (DPI) and payload content filtering.

20

4. The system as claimed in Claim 2, wherein said plurality of parameters comprise keywords, REGEX patterns, IP addresses, URLs and the like.

5. The system as claimed in Claim 1, wherein said collection device comprises:

a classification unit for decoding and classifying plurality of packets contained in said filtered content in accordance with L7 protocols; and

a reconstruction unit for reconstructing and retaining said plurality of packets by employing real time packet reconstruction (RTPR).

5

6. The system as claimed in Claim 5, wherein said collection device further comprises at least one application programming interface (API) for using said plurality of packets obtained after RTPR for further action.

10 7. The system as claimed in Claim 6, wherein at least one API comprises content filtering, content blocking, warrant management and the like.

8. The system as claimed in Claim 2, wherein each filter of said plurality of filters comprises a High Throughput Network Processor Unit (NPU) for data
15 processing.

9. The system as claimed in Claim 1, wherein said system further comprises a dissector for stripping headers and abstracting information such as MAC addresses, IP addresses and the like from said web content.

20

10. The system as claimed in Claim 1, wherein said system further comprises a flow manager for grouping said web content into a plurality of packets belonging to same flow.

11. The system as claimed in Claim 1, wherein said system further comprises a protocol checker for determining a protocol to be used for each group of said plurality of packets.

5 12. The system as claimed in Claim 11, wherein said protocol checker comprises HTTP, POP3, SMTP, and the like.

13. The system as claimed in Claim 1, wherein said system further comprises a protocol decoder for abstracting application layer data and
10 grouping said application layer data into a plurality of layer 7 applications.

14. The system as claimed in Claim 1, wherein said system further comprises at least one database module for storing said filtered content.

15 15. The system as claimed in Claim 1, wherein said system further comprises a web management GUI for accessing said at least one database module.

16. A method for intercepting, filtering and blocking content from internet in
20 real time, said method comprising the steps of:

retrieving web content from a source, said web content comprising a plurality of packets;

stripping headers and abstracting information such as MAC addresses, IP addresses and the like from said plurality of packets by a dissector;

grouping said plurality of packets belonging to same flow together by a flow manager;

determining a protocol to be used for each group of said plurality of packets by a protocol checker;

5 abstracting application layer data by a protocol decoder and grouping said application layer data into a plurality of layer 7 applications; and

 sending said application layer data for further processing.

17. The system as claimed in Claim 16, wherein said method further
10 comprises the step of determining if further decoding is needed.

18. The system as claimed in Claim 16, wherein said method further comprises the step of sending data to application data recorder if further decoding is needed.

15

19. The system as claimed in Claim 16, wherein said method further comprises the step of saving data in at least one database module.

20