Extracting Internet Background Radiation from Raw Traffic Using Greynet

Lihua MIAO, Wei DING, Haiting ZHU
School of Computer Science and Engineering
Southeast University
Jiangsu, China
{lhmiao, wding, htzhu}@njnet.edu.cn

***Abstract*—Analysis based on Internet Background Radiation (IBR) has been shown to be effective for detecting Internet threats such as worms and DDOS attacks. In contrast with traditional methods using darknets, this paper proposes a scheme of extracting IBR from raw traffic gathered at a point of presence (PoP) by its ISP. This method is proceeding from a different angle based on redefined greynet and IBR's own characteristics. The method's basic principle is introduced first and then it is qualitatively analyzed using "precision" and "recall". On this basis, the method is implemented facing raw traffic in a particular format and applied to measured data with certain scale. Based on the successfully extracted IBR, subsequent analysis reveals that this scheme is effective and feasible.**

Keywords-Internet background radiation; greynet; grey space; Internet threats; backscatter

I. INTRODUCTION

Dark space is a region of globally routeable yet unused IP addresses [1]. Since there are no legitimate hosts or devices in it, any observed traffic must be anomalous (but not necessarily malicious). Traffic targeting dark space is called *Internet Background Radiation* (***IBR***) [2], which usually is the result of worm propagation, network scanning, backscatter from activities using spoofed source addresses (e.g. DDOS attacks) and misconfiguration [1]. Analyzing IBR has been shown to be an effective method for detecting and tracking Internet threats [3]. Similar systems have a variety of names, such as darknet [4], blackhole [3], network telescope [5], Internet motion sensor [6] and network sinks [7].

For enterprise network administrators, detecting and tracking potential threats are very necessary. Nevertheless, the deployment of darknet is not always feasible due to typically requiring large, contiguous blocks of unused IP addresses. To this end, ref. [8] and [9] introduce the concept of "greynet" which is a mix of lit (used) and dark (unused) IP addresses. The authors show that greynets can also be used to monitor IBR and relatively sparse greynets can achieve useful levels of network scan detection.

As Internet service providers, ISPs should use all resources and technologies to detect and track potential Internet threats in order to guarantee all downstream access networks' normal and stable operation. As a matter of fact, IBR always exists no matter whether darknets or greynets are deployed or not. As long as it is obtained through reasonable and effective methods, IBR could be used for network security management. In contrast with enterprise network operators, the upstream ISPs are not aware of which internal IP addresses are "lit up" (used). However, ISPs have the advantage of obtaining all access networks' traffic and related features at a higher level. Proceeding from this angle, this paper introduces how to extract IBR according to its own features from raw traffic passively captured at a point of presence (PoP) by its ISP and then conducts further analysis on this basis.

II. RELATED WORK

IBR can be used to capture Internet threats [1, 6, 15], detect censorship or Internet outage [20], etc. Much related work has been done using dark space and is all based on measured data. There are several well-known systems, including CAIDA's network telescope [5], the Internet Motion Sensor project in University of Michigan [6], Team Cymru Darknet Project [10], etc. These systems all obtain IBR based on monitoring dark address blocks and some of them can elicit additional information via active responses, such as responding to TCP SYN packets with TCP SYN+ACK packets. One shortcoming of such systems is that monitored dark space is typically fixed which makes them potentially avoidable for malicious attacks. Another is that their deployments are rather difficult for common enterprise networks owing to the requirement of contiguous blocks of unused IP addresses. Therefore, ref. [8] and [9] introduce greynet to conduct similar work. In addition, ref. [11] and [12] treat addresses which are inactive throughout a given time period as unused addresses. Unlike dark space monitors, systems using greynets are conducted based on productive address blocks, i.e. unused addresses interspersed among valid hosts, and thus typically do not use active responders which makes them powerless to elicit additional data. But the advantage is that greynets' own features (time-dependent and based on productive address blocks) make them harder to be avoided. There are several similar studies which are also conducted in productive networks. Ref. [13] expands the definition of IBR from "all data sent to unused IP addresses" to "all data sent to unused IP-port combinations" and examines the differences between the unused, previously used and used IP addresses in terms of received IBR. Ref. [14] obtains IBR by means of ***one-way flows*** (i.e. there is only traffic in one direction between communication endpoints)

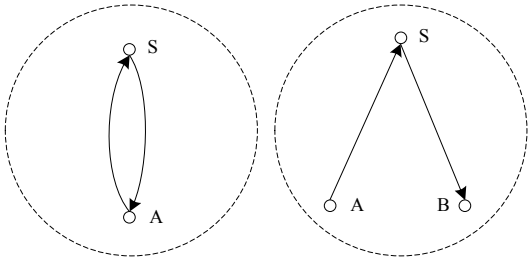
based on Netflow records captured from the academic and research network of Switzerland (SWITCH) but the false positive rate is relatively high.

To the best of our knowledge, there is little work studying how to filter IBR out of raw traffic from the perspective of ISPs. Thus, the paper is devoted to solve this problem. First, we redefine ***greynet***. On this basis, a scheme is proposed to extract IBR according to its own features. We apply this method to measured data with certain scale and further analysis shows that this method is effective and feasible.

III. EXTRACTING IBR VIA GREYNET

A. The Scheme of Extracting IBR

In the current Internet architecture, routers do not verify authenticity of the forwarded packets' source addresses. However, transportation and application layer protocols (e.g. TCP, DNS) implemented on end hosts always assume that the source addresses of received packets are authentic. This design flaw makes activities using spoofed source addresses (e.g. DDOS attacks) hard to be detected and blocked. It also is one of the main reasons of IBR's existence. Traffic generated by activities with spoofed source addresses is quite different from normal traffic in that it is asymmetric, showed in Fig. 1. Proceeding from this angle, this paper obtains IBR partly based on this asymmetry. Due to not using dark space, our system belongs to those based on greynets.



(a) Normal (b) Spoofed source address

Figure 1. Asymmetry caused by using spoofed source addresses

From the perspective of enterprise network operators, ref. [8] and [9] define greynet as a region of IP address space that is sparsely populated with dark addresses interspersed with lit IP addresses. In contrast, the upstream ISPs are not aware of the configuration parameters in downstream enterprise networks and can only acquire interested information by passively observing backbone links, especially the access links from the corresponding PoPs (Fig. 2). Therefore, we redefine "***greynet***" from the perspective of ISPs.

Let $EN = \{\boldsymbol{EN_i},\ i = 1 \sim n\}$ denote all enterprise networks which access the backbone network through PoP p, where $n > 1$ and EN_i stands for the IP address space of network i which is globally routeable, showed in Fig. 2.

Let ***Traffic****(T, p, srcIP, destIP)* denote traffic originated from *srcIP*, passing p and targeting *destIP* during time period T, where *srcIP* and *destIP* are two sets of IP addresses.

Let ***Type_of_packet****(T, p, ip_0)* denote a packet type collection of packets originated from a single address ip_0 and passing p during T. Packet types here are combinations of the protocol field and TCP flags (if any) in packet headers, e.g. TCP SYN, TCP SYN+ACK, UDP, etc.

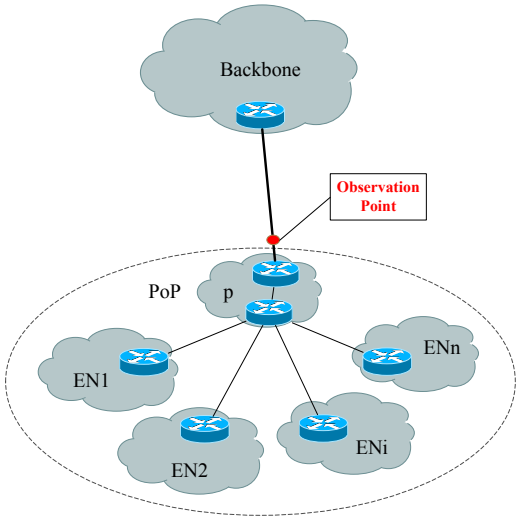


Figure 2. Schematic diagram of EN and the observation point

For $\forall\, g \in EN_i$, if $Traffic(T,p,\{g\},*) = \phi$, then term g a grey IP address. Let $greyEN_i$ denote the collection of grey IP addresses in EN_i.

For $\forall\, g \in EN_i$, if $Type\_of\_packet(T,\ p,\ g) = \{TCP\ RST,\ TCP\ RST+ACK\}$, then term g a pseudo-lit (pseudo-active) IP address. Let $\boldsymbol{pseudolitEN_i}$ denote the collection of pseudo-lit IP addresses in EN_i.

(D1) Greynet. Let $\boldsymbol{greyspaceEN_i} = greyEN_i \cup pseudolitEN_i$ denote the ***grey space*** in EN_i, $litEN_i = EN_i - greyspaceEN_i$ denote the lit space in EN_i, where $\forall\, g \in litEN_i$ is a lit (used) IP address. If $greyspaceEN_i \neq \phi$ and $litEN_i \neq \phi$, we say EN_i is a ***greynet***. If $litEN_i = \phi$, it means that EN_i is unproductive during T which would not occur under normal circumstances.

(D1) shows that the grey space in an enterprise network is composed of inactive and pseudo-active IP addresses during time period T. Under certain circumstances, enterprise networks' perimeter firewalls would respond to TCP SYN packets with TCP RST/RST+ACK packets for internal IP addresses and make these IP addresses pseudo-active. Because these TCP RST/RST+ACK packets are not actually responses from hosts using the corresponding addresses, we classify these IP addresses into the grey space.

For $\forall\, j \notin EN$, if $Traffic(T,\ p,\ \{j\},\ greyspaceEN_i) \neq \phi$, then we term j a suspicious IP address. Let $\boldsymbol{suspiciousEN_i}$ denote the collection of suspicious IP addresses against EN_i. Ref. [2] and [4] show that a suspicious address tends to perform similar behavior between different destination IP addresses. On the basis, we propose a scheme to extract IBR as below.

(D2) Extracting IBR. Let $\boldsymbol{radiationEN_i} = A - \boldsymbol{match}(A, B)$ denote IBR against EN_i during T where $A = Traffic(T, p, suspiciousEN_i, EN_i)$, $B = Traffic(T, p, litEN_i, suspiciousEN_i)$ and $match(A, B)$ stands for the flows in A which can be paired with flows in B. In Fig. 3, the red arrows indicate $radiationEN_i$.

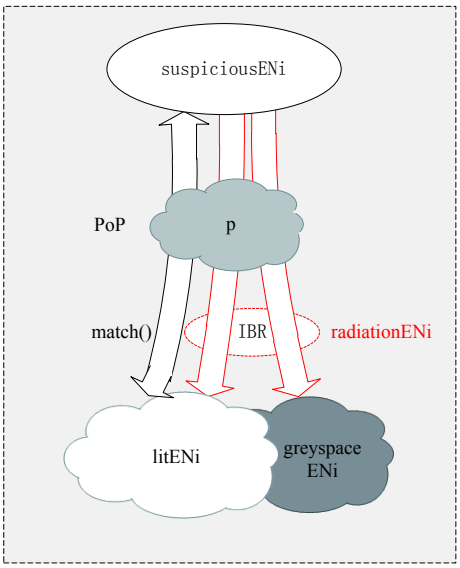


Figure 3. Schematic diagram of $radiationEN_i$

B. Qualitative Analysis of the Scheme

(D2) classifies the raw traffic into two categories: $radiationEN_i$ and the remainder. Let $\boldsymbol{IBRgreyspaceEN_i}$ denote IBR targeting $greyspaceEN_i$. Our purpose is to identify traffic targeting EN_i which have exactly the same features and purposes as $IBRgreyspaceEN_i$. We use $\boldsymbol{IBREN_i}$ to denote the accurate set of IBR against EN_i and the $radiationEN_i$ obtained in (D2) is its estimation set. Because we cannot obtain $IBREN_i$ accurately, we only analyze our method qualitatively using *precision* and *recall*, showed in (1) and (2). Here, $\boldsymbol{TP} = |IBREN_i \cap radiationEN_i|$, $\boldsymbol{FP} = |radiationEN_i| - TP$, $\boldsymbol{FN} = |IBREN_i| - TP$.

$$precision = \frac{TP}{TP+FP} = \frac{\left|IBREN_i \cap radiationEN_i\right|}{\left|radiationEN_i\right|} \quad (1)$$

$$recall = \frac{TP}{TP+FN} = \frac{\left|IBREN_i \cap radiationEN_i\right|}{\left|IBREN_i\right|} \quad (2)$$

In order to maximize *recall* in (2), we classify pseudo-lit IP addresses into the grey space of EN_i and estimate $IBREN_i$ using $radiationEN_i$ instead of $IBRgreyspaceEN_i$. We assume that any productive traffic sent from a host should be answered by its communication partner and moreover the observation point locates at a PoP, so we treat grey IP addresses in enterprise networks as inactive ones. This assumption might increase *FP* to some extent. As a matter of fact, almost all public point-to-point applications in Internet build up ***two-way flows*** (i.e. there is bidirectional traffic between communication endpoints) [14]. According to our observation, applications using multicast and broadcast (such as NTP) are typically used inside enterprise networks and thus cannot be seen at a PoP; P2P applications which are used in certain scale also work in a request/reply mode [19]. Thus, there is only one case which might cause misjudgment: private applications which do not require interaction in both directions. This kind of applications is not ubiquitous, so we assume that they have little impact on the overall. Therefore, the *precision* metric in (1) is close to 1. The qualitatively analysis above reveals that the obtained $radiationEN_i$ is valuable from the analytic perspective.

In addition, according to (D1), the grey space of an enterprise network is time-dependent. In contrast with dark address space, the grey space is dynamic and more concealed thus making it harder to be avoided.

There are two key problems which need to be solved in order to extract IBR from raw traffic using (D2). One is identifying $suspiciousEN_i$. This is solvable when $greyspaceEN_i$ is known. The other is realizing the function of *match()* reasonably. This function is meant to build up two-way flows when there is bidirectional traffic between communication endpoints or one-way flows when there is traffic only in one direction. This is hard to achieve in practice. We will discuss this topic in the following section based on measured data.

In the following sections, we use a group of measured traffic from a PoP and extract the corresponding IBR using (D2). The extracted IBR is further analyzed to prove that it's valuable from the analytic perspective.

At last, we propose a metric which could be used to evaluate the radiation intensity against enterprise networks.

(D3) Radiation Intensity. Let $INEN_i$ denote inbound traffic during T, the radiation intensity against EN_i is $\boldsymbol{RIEN_i} = |radiationEN_i|/|INEN_i|$. Then $\boldsymbol{|radiationEN_i|/|EN_i|}$ means the volume of IBR received by a single IP address on average.

IV. Extracting IBR from Measured Data

The implementation of (D2) differs as encountering different formats of raw traffic. In this section, we select a group of measured packet traces to verify that our method is effective and feasible.

A. Dataset and the Implementation of (D2)

The observed PoP is at the border of a province network in China Education and Research Network (CERNET). This PoP is now serving more than 100 campus networks and the whole IP space consists of up to 5000 /24 sized blocks. The administrative ISP captures packets (only the first 60 bytes) on a regular basis. Both directions (inbound and outbound) are measured using 1/4 flow sampling, i.e. only a quarter of the inner IP space is monitored. The collection of captured packets are saved as *Traces* (separated into two directions and ascending by their timestamps) and published after anonymization [18]. We select seven Traces captured from five enterprise networks (campus networks) in different scale as the

experiment data. The dataset used in this paper is showed in Tab. Ⅰ and the observation period T is [00:00, 24:00).

TABLE I. DATASET

Trace ID	IP count	EN_i	Date	Packets/sec (Inbound)	Packets/sec (Outbound)
A	4096	EN_1	2009-11-14	4.57K	4.41K
B		EN_1	2010-11-14	4.60K	4.19K
C1		EN_1	2011-11-17	3.65K	3.08K
C2	11648	EN_2	2011-11-17	16.86K	16.14K
C3	4096	EN_3	2011-11-17	1.40K	1.21K
C4	3616	EN_4	2011-11-17	3.52K	2.77K
C5	1024	EN_5	2011-11-17	0.39K	0.07K

The *IP count* in Tab. Ⅰ is the number of IP addresses after 1/4 flow sampling. The five selected networks own 24 different blocks ranging in size from /25 to /19. Note that A, B and C1 are captured from the same network on three different days from 2009 to 2011, and C1~C5 are captured from five different networks on the same day.

As for Traces, the implementation of (D2) can be divided into three steps as below. Let $TraceEN_i[in, out]$ denote all observed packets of EN_i during T where *in* and *out* indicate packets' directions.

(D4) ***IBR_Extraction(TraceEN$_i$[in, out], EN$_i$, T):***

Step 1: Obtain $greyspaceEN_i$ using $TraceEN_i[out]$ and $TraceEN_i[in]$;

Step 2: Obtain $suspiciousEN_i$ using $TraceEN_i[in]$ and the obtained $greyspaceEN_i$ in step 1;

Step 3: Filter all packets originated from $suspiciousEN_i$ out of $TraceEN_i[in]$. Filter all packets originating from $litEN_i$ and targeting $suspiciousEN_i$ out of $TraceEN_i[out]$. The filtered packets can still be called a Trace, say $\boldsymbol{FX_i}$. Group the packets in FX_i into flows in each direction and pair inbound flows with outbound flows. We can then obtain the $radiationEN_i$ in (D2).

B. Grouping Packets into Flows

(D4) is based on flow records which truly reflect end-to-end connections. We cannot simply adopt any common used flow specifications even those with large degree of acceptance. For example, Cisco's Netflow, even unsampled, is not suitable for our method. One reason is that a flow is terminated encountering a TCP FIN enabled packet. According to RFC793, a TCP FIN enabled packet could be followed by a TCP ACK packet. In this situation, the TCP ACK packet would be misjudged as a one-way flow. The other is that Netflow's default inactive timeout is 15 seconds. When the flow rate in one direction is quite different with the other between two communication endpoints, a short inactive timeout could cut one direction's flow into two parts while the flow in the other direction remains complete. Then the flow's tail which is cut off by timeout would be judged as a one-way flow. Even using 64 seconds as the inactive timeout [17], the situation mentioned above might still happen. Because (D2) regards one-way flows as IBR (Fig. 3), *FP* will thus be increased using Netflow records.

In order to furthest assure flows' integrity, we generate 5-tuple flows [16] with terminating conditions detailed below. (1) Terminate a flow when encountering a TCP RST enabled packet; (2) A TCP SYN/SYN+ACK packet is always the first packet of a new flow; (3) If the first packet of a flow is a TCP FIN/RST enabled packet, terminate this flow; (4) Inactive timeout t0. It changes with different Traces. We discuss this in detail in the following subsection.

When flows in both directions are ready, we pair inbound flows with outbound flows when their durations have intersection; or if they are both single packet flows, a time gap smaller than a threshold is allowed (we use 4 seconds as the threshold in practice).

C. Choosing Inactive Timeout t_0

As mentioned above, unreasonable t_0 might increase *FP*. In order to reduce *FP* furthest while ensuring the operability, we choose t_0 for a certain Trace, say $\boldsymbol{X}$ as below.

Let $\delta(t)$ denote the flow count of X using timeout t. As t increases, the approximate changing trend of $\delta(t)$ is showed in Fig. 4. Assume that the shortest acceptable timeout in practice is 15 seconds and the longest is 600 seconds, and let $\Delta = \delta(15s) - \delta(600s)$ denote the maximum error. Let $t_0 =$ {30s, 45s, 60s, 75s, 90s, 105s, 120s, 180s, 240s, 300s, 450s} and calculate the corresponding $\varepsilon = \left(\delta(t_0) - \delta(600s)\right) / \Delta$. Choose t_0 as the timeout of X when its $\varepsilon \approx 0.1$. We use this method against FX_i in (D4). The chosen timeouts are: {A: 240s, B: 105s, C1: 90s, C2: 120s, C3: 240s, C4:300s, C5:180s}.

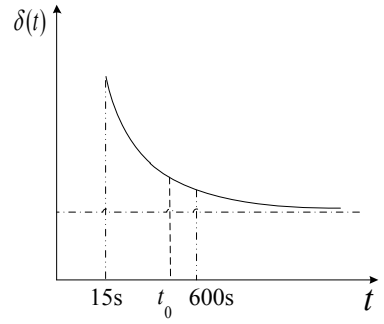


Figure 4. Schematic diagram of $\delta(t)$'s changing trend

D. The Extracted IBR

Using the methodology introduced above, basic statistics of extracted IBR are showed in Tab. Ⅱ.

Tab. Ⅱ can be concluded by four key points as below.

1) According to Tab. Ⅰ, A, B and C1 are captured from the same network in three different years from 2009 to 2011. Tab. Ⅱ shows that the volume of IBR shows a modest increase from 2009 to 2010 and has a big

increase in 2011. This might indicates that the volume of IBR increases with the passage of time which is also illustrated in [1].

2) C1~C5 are captured from five different networks on the same day, so the results are comparable. Tab. Ⅱ reveals that radiation intensity against EN_5 is the highest although it is the smallest network.
3) For every single Trace, the number of IBR packets received by a single address on average is close to that of the received IBR flows. This indicates that most IBR flows are single packet ones.
4) For C1~C5, the $RIEN_i$ (#flows) column shows that even the smallest $RIEN_2$ is up to 50%. This means that even for EN_2 whose radiation intensity is the lowest, almost 50% inbound flows are IBR flows. Large amount of short lived flows caused by IBR might have great impact on flow-oriented routing management policies.

TABLE II. BASIC STATISTICS OF EXTRACTED IBR

Trace	$\|radiationEN_i\|/\|EN_i\|$		$RIEN_i$ (%)	
	#packets	***#flows***	***# packets***	***#flows***
A	1504	1247	1.56	48.90
B	1596	1384	1.65	47.10
C1	5046	4633	6.55	76.44
C2	5255	5073	4.20	46.76
C3	7325	7254	24.80	92.23
C4	11801	11750	14.04	73.36
C5	32097	31764	97.58	99.81

V. ACTIVITIES IN INTERNET BACKGROUND RADIATION

In this section, we further analyze IBR extracted from **C1~C5** in last section. According to the activities in it, the extracted traffic is proved to be IBR. The classification method used below is based on *Type_of_packet()* defined before (D1).

For every Trace's IBR, classify IBR into several categories, such as TCP SYN single packet flows, TCP SYN+ACK single packet flows, TCP RST+ACK single packet flows, TCP FIN+RST multi-packet flows (cumulative OR of all packets' TCP flags), UDP flows, ICMP flows, etc. For C1~C5, the top three types of flows are the same: ***F1*** = {TCP SYN+ACK single packet flows}, ***F2*** = {TCP SYN single packet flows}, ***F3*** = {UDP flows}, showed in Tab. Ⅲ (unit: #flows). Tab. Ⅲ reveals that their sum account for more than 99% of the total extracted flows.

TABLE III. TOP THREE IBR FLOW TYPES

Trace	C1	C2	C3	C4	C5
|F1| (%)	87.74	84.46	89.58	92.63	95.56
|F2| (%)	8.01	10.96	6.64	3.89	2.15
|F3| (%)	3.84	3.90	3.46	3.18	2.11
SUM	99.59	99.32	99.68	99.70	99.82

To confirm that most of the extracted traffic belongs to IBR, we further analyze the activities in F1~F3 separately as below. Apparently, TCP SYN+ACK/SYN single packet flows are up to no good.

A. TCP SYN+ACK Single Packet Flows

Let ***radiation_saEN$_i$*** denote the collection of TCP SYN+ACK single packet flows in $radiationEN_i$. For C1~C5, the top two source ports in $radiation\_saEN_i$ are 80 and 7000. Let ***radiation_sa80EN$_i$*** denote flows in $radiation\_saEN_i$ whose source ports are 80 and ***radiation_sa7000EN$_i$*** denote flows with source port 7000. $|radiation\_sa80EN_i|/|radiation\_saEN_i|$ and $|radiation\_sa7000EN_i|/|radiation\_saEN_i|$ are showed in Tab. Ⅳ. Tab. Ⅳ reveals that flows sent by 80 and 7000 account for more than 93% of the total TCP SYN+ACK single packet flows.

TABLE IV. TOP SOURCE PORTS OF TCP SYN+ACK FLOWS

Port	C1	C2	C3	C4	C5
80 (%)	85.73	85.74	86.69	87.67	86.60
7000 (%)	8.64	8.61	8.07	7.47	6.90

Further analyze flows sent by 80 ($radiation\_sa80EN_i$, i=1~5). The top one IP address is x.x.108.20. Flows sent by it account for more than half of $radiation\_sa80EN_i$: {C1: 67.44%, C2: 67.60%, C3: 62.97%, C4:58.49%, C5: 57.83%}. We further check this address and it turns out to be a website of an unauthorized game server. We guess there is a DDOS attack against it during our observation and our scheme successfully filters out its backscatter which typically belongs to IBR.

Further analyze $radiation\_sa7000EN_i$, i=1~5. The top one address is x.x.107.234. Flows sent by it occupy the majority of $radiation\_sa7000EN_i$: {C1: 88.42%, C2: 88.46%, C3: 88.38%, C4: 88.33%, C5: 88.84%}. 7000 is the login port of a popular online game. Therefore, similar to the previous case, it also belongs to IBR successfully filtered out by our scheme.

According to our inspection, most of the TCP SYN+ACK single packet flows must be the backscatter from activities using spoofed source IP addresses such as DDoS attacks.

B. TCP SYN Single Packet Flows

Let $radiation\_sEN_i$ denote the collection of TCP SYN single packet flows in $radiationEN_i$. The top three destination ports of $radiation\_sEN_i$ are showed in Tab. Ⅴ. 1433 and 445 are attackers' common favorite ports while port 8909 just showed up lately. Further investigation reveals that 8909 might be opened by a video player with certain prevalence and the corresponding hosts thus are turned into HTTP anonymous proxies. This program bug is already patched in its latest version.

According to our analysis, most of the TCP SYN single packet flows are the result of network scanning and probes. Due to the lack of active responder, we cannot further analyze the purpose of these attempts.

TABLE V. TOP DESTINATION PORTS OF TCP SYN FLOWS

Port	C1	C2	C3	C4	C5
1433 (%)	60.90	56.83	53.46	56.43	50.43
445 (%)	7.40	12.37	25.36	18.72	29.07
8909 (%)	8.79	6.68	5.46	5.91	6.14

C. UDP Flows

Let $radiation\_udpEN_i$ denote the collection of UDP flows in $radiationEN_i$. For C1~C5, the top one source port of $radiation\_udpEN_i$ is 53. We use $radiation\_udp53EN_i$ to denote flows sent by this port in $radiation\_udpEN_i$, then its percentage of the total UDP IBR flows are: {C1: 78.62%, C2: 82.32%, C3: 92.53%, C4:97.55%, C5: 97.22%}. The corresponding packet percentages are {C1: 25.03%, C2: 56.33%, C3: 80.94%, C4:94.92%, C5: 95.45%}. We further examine $radiation\_udp53EN_i$, the top two source addresses are x.x.114.45 and x.x.58.274 who account for 54% and 45.5% approximately. They turn out to be two name servers of a company who provides web hosting service. We further analyze the corresponding packets and find out they are all "standard query response" to queries asking for the same domain name. We first thought that they were backscatters of DNS poisoning attempts, but then we found out that the queried domain name belongs to a small electrical company. Thus we guess that the observed traffic is the backscatter from DDoS attacks against these name servers.

VI. CONCLUSION

Capturing and analyzing Internet Background Radiation can help network administrators to detect anomalous events in their networks. Traditional methods monitor IBR using darknets and the original enterprise greynets. They can achieve very good results, but they have their own flaws. Proceeding from a new angle, this paper proposes a new scheme to obtain IBR based on its asymmetry characteristics. We use our method against a measured dataset with certain scale and examine the obtained traffic. Although there is room for improvement in the implementation algorithm's performance and parameter choice and moreover the process of inspecting the extracted traffic can be further standardized, the findings reveal that our scheme is feasible and the extracted IBR is valuable from the analytical perspective. In contrast with systems based on darknets, the grey space in our paper is dynamic and harder to be avoided. Compared with systems using the original greynets, our method based on redefined greynet can be carried out in a higher level and more suitable for ISPs.

ACKNOWLEDGMENT

The research was supported by the National Basic Research Program of China under Grant No. 2009CB320505 and the National Key Technology Research and Develop Program of China under Grant No.2008BAH37B04. The authors thank the reviewers for the comments that helped to improve the paper.

REFERENCES

[1] E.Wustrow, M. Karir, M. Bailey, F. Jahanian, and G. Huston. Internet Background Radiation Revisited. In Proceedings of the 10th annual Conference on Internet Measurement(IMC'10), ACM, 2010.

[2] R. Pang, V. Yegneswaran, P. Barford, V. Paxson, and L. Peterson. Characteristics of Internet Background Radiation. In Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement(IMC'04), Oct 2004.

[3] E. Cooke, M. Bailey, Z.M.Mao, D.Watson and F. Jahanian. Toward Understanding Distributed Blackhole Placement. In Proceedings of ACM CCS Workshop on Rapid Malcode, pp. 54-64. ACM Press, October 2004.

[4] M. Bailey, E. Cooke, D. Watson, F. Jahanian, and N. Provos. Practical Darknet Measurement. In Proceedings of the 40th Annual Conference on Information Sciences and Systems (CISS), Mar 2006.

[5] D. Moore, C. Shannon, G.M. Voelker, and S. Savage. Network Telescopes. Cooperative Association for Internet Data Analysis. Technical Report, 2004.

[6] Michael Bailey, Evan Cooke, Farnam Jahanian, Jose Nazario and David Watson. The Internet Motion Sensor: A Distributed Blackhole Monitoring System. In Proceedings of Network and Distributed System Security Symposium (NDSS '05), San Diego, CA, February 2005.

[7] V. Yegneswaran, P. Barford, and D. Plonka. On the Design and Use of Internet Sinks for Network Abuse Monitoring. In Proceedings of the Symposium on Recent Advances in Intrusion Detection, Sep. 2004.

[8] Warren Harrop and Grenville Armitage. Greynets: A Definition and Evaluation of Sparsely Populated Darknets. In Proceedings of the ACM SIGCOMM MineNet Workshop, Philadelphia, PA, August 2005.

[9] Warren Harrop and Grenville Armitage. Defining and Evaluating Greynets (Sparse Darknets). In Proceedings of the IEEE Conference on Local Computer Networks 30th Anniversary (LCN'05), 2005.

[10] Team Cymru Darknet Project. http://www.team-cymru.org/Services/darknets.html

[11] Y. Jin, G. Simon, K.Xu, Z.-L. Zhang and V. Kumar. Grey's Anatomy: Dissecting Scanning Activities Using IP Grey Space Analysis. In Proceedings of the Second Workshop on Tackling Computer Systems Problems with Machine Learning Techniques (SysML'07), 2007.

[12] Y. Jin, Z.-L. Zhang, K. Xu, F. Cao, and S. Sahu. Identifying and Tracking Suspicious Activities through IP Grey Space Analysis. In Proceedings of the 3rd annual ACM workshop on Mining network data, 2007.

[13] Kees Hoekzema. Background Radiation on Today's Internet with a Focus on Used IP Addresses. 8th Twente Student Conference on IT, Enschede, January, 2008.

[14] Konstantinos Karampogias. Evaluating and Improving the Detection of Internet Background Radiation. Semester thesis at ETH Zurich, Feb. 2011.

[15] S. Soltani, S. A. Kyaham, H. Radha. Detecting Malware Outbreaks Using a Statistical Model of Blackhole Traffic. In Proceedings IEEE International Conference on Communications (ICC'08), pp. 1593-1597, Beijing, 2008.

[16] DongJin Lee and Nevil Brownlee. Passive Measurement of One-way and Two-way Flow Lifetimes. ACM SIGCOMM Computer Communication Review, 37(3):17-28, 2007.

[17] K.C. Claffy, H.W. Braun, and G.C. Polyzos. A Parameterizable Methodology for Internet Traffic Flow Profiling. Selected Areas in Communications, IEEE Journal on, vol. 13, pp. 1481-1494, 1995.

[18] IP Trace Distribution System. http://iptas.edu.cn/src/system.php

[19] Wang Gang, Ding Wei and Dong Shi. Classification of UDP Traffic from P2P Applications. In Proceeding of 17th IEEE International Conference on Networks (ICON), 2011.

[20] A. Dainotti, C. Squarcella, E. Aben, K. C. Claffy, M. Chiesa, M. Russo, A. Pescap. Analysis of Country-wide Internet Outages Caused by Censorship. ACM SIGCOMM/SIGMETRICS Internet Measurement Conference IMC 2011, November 2011.