An Analysis of Correlations of Intrusion Alerts in an NREN

Vaclav Bartos
Brno University of Technology
Brno, Czech Republic
ibartosv@fit.vutbr.cz

Martin Zadnik
CESNET
Prague, Czech Republic
zadnik@cesnet.cz

***Abstract*—An ever increasing impact and amount of network attacks have driven many organizations to deploy various network monitoring and analysis systems such as honeypots, intrusion detection systems, log analyzers and flow monitors. Besides improving these systems a logical next step is to collect and correlate alerts from multiple systems distributed across organizations. The idea is to leverage a joint effect of multiple monitoring systems to build a more robust and efficient system, ideally, lacking the shortcomings of the individual contributing systems. This paper presents an analysis of alert reports gathered from several such detectors deployed in national research and education network (NREN). The analysis focuses on the correlations of reported events in temporal domain as well as on the correlations of different event types.**

I. Introduction

Blacklists are often used to collect and distribute information on misbehaving entities. However, the blacklists include very coarse information (usually only the identifier). Also, the spatio-temporal incident characteristics that are prerequisite to blacklist usage have not been studied thoroughly. Moreover, the blacklists contain only the most visible misbehaving entities reported by a particular detection system. An attacker who attacks multiple targets but with lower intensity might evade being reported in the blacklist despite being detected due to a high risk of a false positive.

To this end, this paper contributes by an assessment of data (in Sec. III) gathered on suspicious events detected by various detection systems deployed in CESNET (Czech NREN) over a period of 6 months. In particular, the data analysis in Sec. IV aims to answer two questions. Whether it is usual that IP addresses appear in collected alerts repeatedly and whether there are correlations between sources of different kinds of malicious traffic.

II. Related work

It is a common expectation that incidents exhibit some spatio-temporal correlations but the amount of literature to document these correlations is surprisingly low. The largest portion of the literature discusses spam characteristics, *e.g.* [9]. We are aware of only few works that do not focus solely on spam and include also other types of malicious traffic. The author of [7] focuses primarily on spatial correlations of spam and SSH bruteforce attacks. His work shows that there are "Internet bad neighbourhoods" which are application specific and may be observed on various scales such as prefixes, ISPs and countries. Subsequently, he elaborated the concept of bad neighbourhoods on spam in [1], [2], [8]. A thorough study [4] focusing exclusively on SSH bruteforce attacks shows that some attacks are distributed as well as stealthy.

A work studying botnet spatio-temporal behavior [6] introduces a network quality termed uncleanliness to estimate future botnet addresses. They analyzed several internal and external alert sources (*e.g.* blacklists). Their results demonstrate the evidence for both spatial and temporal correlation of uncleanliness and relationship between botnet membership and spamming and scanning activities. On the other hand, in [3] the authors show that there is an overlap of blacklists of the same type whereas only little overlap of different types within a period of one week.

Most of the related works utilize already preprocessed data, i.e. blacklists, as the major source of input data. The blacklists are usually constructed using data from many sources located at various places all over the world. In contrast to this, our work is based on alerts of various detection systems deployed at a single NREN network. We thus investigate whether the correlations in malicious traffic discovered in the global data are observable locally as well. Moreover, our analysis focuses on different types of malicious events than those studied before. We are not aware of any work mutually analyzing spatial and temporal correlations of alerts about port scanning, bruteforce password guessing, unauthorized web accesses and TCP SYN flood attacks, which are the types of events studied in this work.

III. Data sets

For the analysis presented in this paper we gathered alert data from a diverse range of detection systems – honeypots as well as flow-based traffic analysis systems[1]. These detectors are based on diverse software and deployed in various campus networks connected to CESNET or in the CESNET network itself (with one exception). We do not have direct control over most of the detectors since they are operated by administrators of the campus networks and we only get results through CESNET's alert sharing system called Warden. The dataset is thus very heterogenous, with data from sources with different characteristics and configurations. While this might not be ideal for analysis, we argue that such a non-uniform deployment is common in many real networks and therefore valid for analysis. Also, we do not aim at characterizing

[1]The dataset is available in an anonymized form at: http://www.fit.vutbr.cz/~ibartosv/alert_dataset/

TABLE I: Datasets used for the analysis.

dataset	attack type	detector	detector type	observed IP range	# alerts	# unique source IPs	date range
scan1	scan	LaBrea	honeypot	512	843911	416035	2013-08-01 – 2014-01-31
scan2 [1]	scan	honeyd	honeypot	2048	309190	27524	2013-08-01 – 2014-01-31
scan3	scan	HostStats	flow-based	∼1 M	201848	75565	2013-08-01 – 2014-01-31
scan4 [2]	scan	honeyscan	hybrid	256	126169	49456	2013-08-01 – 2014-01-31
scan5	scan	Dionaea	honeypot	7	16082	8454	2013-08-01 – 2014-01-31
scan6	scan	Dionaea	honeypot	1	7007	4252	2013-08-01 – 2014-01-31
scan7 [3]	scan	LaBrea	honeypot	256	1274	511	2013-08-01 – 2014-01-31
scan8	scan (SSH)	SSHCure	flow-based	∼1 M	23959	4901	2013-10-01 – 2014-01-28
scan-ext	scan (SSH)	SSHCure	flow-based	65536	7116	2290	2013-10-01 – 2014-01-19
sshbf1	BF (SSH)	Kippo	honeypot	1	2832	1525	2013-08-01 – 2014-01-31
sshbf2	BF (SSH)	Kippo	honeypot	7	2218	1035	2013-08-01 – 2014-01-31
sshbf3	BF (SSH)	SSHCure	flow-based	∼1 M	24281	6573	2013-10-01 – 2014-01-28
sshbf-ext	BF (SSH)	SSHCure	flow-based	65536	3699	2167	2013-10-01 – 2014-01-19
webbf1	BF (web login)	HIHAT	honeypot	1	11684	4640	2013-08-01 – 2014-01-31
web1	web access	Dionaea	honeypot	7	1799	1093	2013-08-01 – 2014-01-31
web2	web access	Dionaea	honeypot	1	1473	1182	2013-08-01 – 2014-01-31
synflood1	SYN flood	HostStats	flow-based	∼1 M	207	155	2013-08-01 – 2014-01-31

[1] Some TCP ports are blocked by university's main firewall so scans of these ports do not reach the honeypot.
[2] Only scans targeting both the honeypot segment and some other address in the same /16 network are reported.
[3] Only scans from Czech networks are reported.

malicious traffic itself but rather at analysing alerts generated by a typical set of detection systems. Which we believe is exactly what is needed for developing good alert aggregation and reputation modelling techniques.

We split the alerts by their source and attack type into individual datasets summarized in Tab. I. There are several attack types reported. *Scan* refers to port scanning activity, a large number of connection attempts from a single address to distinct addresses. *BF* stands for *bruteforce* and denotes attempts to log in via SSH or a Web form by automated password guessing. *Web accesses* are trials to access some of the predefined URLs on a honeypot. A *SYN flood* attack is reported when a single address is sending a huge amount of TCP SYN packets to another address.

For each dataset, we also indicate the size of IP range the corresponding detector observes. The flow-based systems observe either traffic at an organization gateway (/16 IP prefix) or traffic on all border links of the CESNET backbone network. Sum of all address ranges which CESNET connects to the Internet is approximately one million. In case of honeypots, the indicated range is the number of IP addresses the honeypot is listening on.

The only detector outside the CESNET network (and also outside the Czech republic and quite far in the IP space) is SSHCure [5] deployed at the University of Twente, the Netherlands. Datasets from this detector are labeled by *-ext* suffix.

The columns *# alerts* and *# unique source IPs* indicate number of alerts and number of unique source IP addresses in each dataset. The column *date range* shows the range from which we have the data. Most of the datasets span 6 months.

We are aware that the datasets may contain false alerts. However, honeypots report accesses to IP address ranges where no legitimate traffic should appear and it is a common practice to consider all such traffic malicious. In this work, all flow-based systems utilize very strict detection policy, e.g. a scanning is reported only if the scanning IP address generates more than 200 connection attempts to different targets. Therefore we assume that the number of false alerts is negligible.

Besides information on the alert type and the reporting detector, each alert includes source IP address and timestamps of the beginning and the end of an attack. However, not all detectors report start and end times of attacks. Therefore we merge all alerts on the same source IP address reported by the same detector within 2 hours. That is, if the time interval between two consecutive alerts reporting the same address is longer than 2 hours, the alerts are considered to represent two distinct attacks. Otherwise the alerts are merged together. Tab. I depicts the numbers of alerts after this aggregation process.

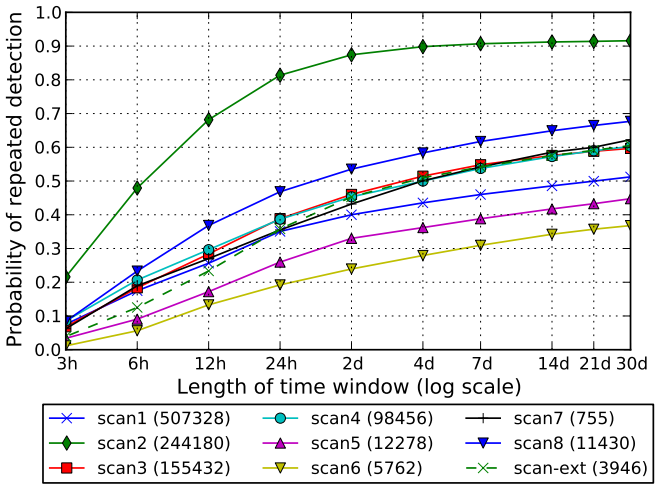
IV. Analysis results

We analyze the datasets from two perspectives – time correlations of alerts and correlations between individual types of alerts.

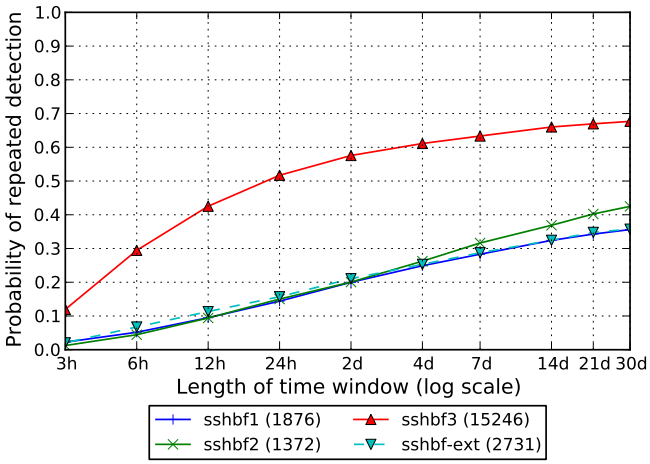
A. Temporal correlations

In the analysis of temporal correlations we ask whether is it usual that the same IP address is detected and reported as malicious repeatedly and how long does it take for such address to be reported again. More precisely, for an attack detected at time t we estimate the probability that there will be another attack from the same address in a time window $(t, t + l]$, where l is the length of the time window. We label that probability $P(l)$.

Fig. 1 shows the probability $P(l)$ for l ranging from 3 hours to 30 days. Each line in the plot corresponds to a dataset containing one type of alerts, i.e. datasets labeled with the same prefix (such as *scan*) merged together. We do not plot a line



(a) Scanning



(b) Bruteforce attacks

Fig. 2: Probability of a repeated detection of the same address for individual detectors.

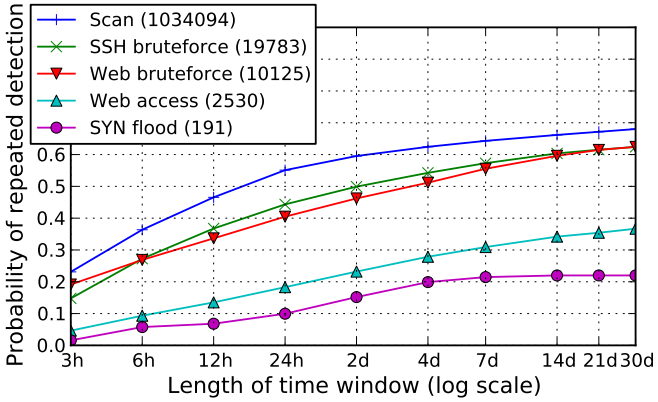


Fig. 1: Probability of a repeated detection of the same address for individual types of attacks.

for all datasets combined since it is almost the same as the one for scans.

As can be seen, around 55% of addresses performing scans are detected as scanning again within the next 24 hours by some of the detection systems. And if we store reports about scanning for a month, almost 70% of detected scans will originate from addresses already known. Just a little lower number of repeated detections can be seen in the case of bruteforce attacks. 27% of machines trying to log in by guessing passwords try this activity again within the next 6 hours and half of them try it again in two days.

On the other hand, malicious web accesses and SYN floods are not detected repeatedly from the same address often. Only 37% of alerts reporting the web accesses and 22% of SYN floods are followed by another such alert from the same address in a month. In the case of web attacks, this can be due to small number of detectors which monitor just a few addresses and thus see just a small fraction of all real attacks. The same explanation can not be used for SYN flood attacks since they are reported by a detector observing all communication going to and from the CESNET network. However, SYN flood attacks are not very frequent, so the low repeatability of SYN flood alerts is better explained simply by small number of such alerts in the dataset. Also, source addresses may be easily spoofed in the case of SYN flood attacks.

No line in the plot shows a steep increase between two lengths of time windows[2]. This means that the delays between consecutive alerts reporting the same address are distributed quite evenly and that there is no common interval in which the attackers would usually repeat their attacks against the same network.

When we merge all datasets into one, 68% of alerts are followed by an alert with the same source address within 30 days. However, these 68% of alerts accounts for 30% of all addresses in the dataset only. As much as 70% of addresses appear in the data only once. It may partially be caused by spreading attacking activities throughout the whole Internet so there is low chance that an attack hit some of the addresses monitored by our detection systems more than once. Another reason is the utilization of IP address as an identifier. The IP address of a host may be assigned dynamically, moreover, mobile devices connect to the Internet from many places and thus having different IP addresses each time. It may therefore happen that a single physical machine performing scans or other malicious activities (*e.g.* due to infection by some malware) appears as several sources as its IP address changes. It would therefore be useful to have possibility to automatically detect which address ranges are used for dynamic assignment since it is not useful to store alerts about such addresses for a long time. It would be even better if there would be a possibility to reliably identify individual machines and track them as they move in the IP address space. We encourage more research on this topic.

Fig. 2 shows the same information as Fig. 1 but for individual datasets which were previously merged. Fig. 2(a) shows data on scans, Fig. 2(b) shows data on SSH bruteforce

[2]Only the SYN flood line has a little different shape from the others. We beleive this is just a statistical anomaly caused by the small number of alerts and their specific distribution in our particular dataset.

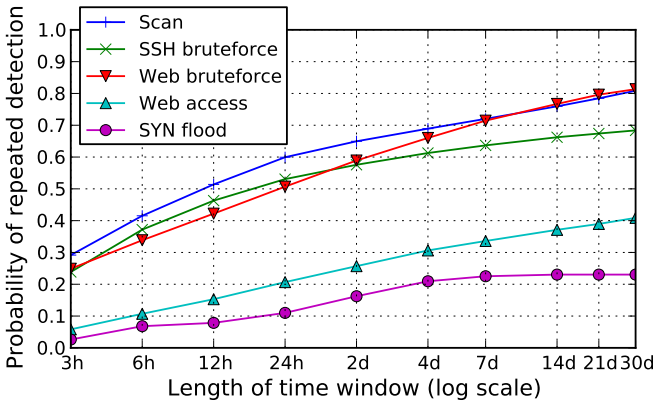


Fig. 3: Probability of a repeated detection of the same /24 prefix for individual types of attacks.

attacks. These plots show, that the extent to which addresses are detected repeatedly by a single detector varies greatly among different detectors. It might depend on the detection capabilities of the detectors, but since most of our data sources are honeypots which rarely miss a connection attempt, the most important characteristics are the amount of traffic and the size of the IP address space a detector observes.

This explanation fits well on *scan2* which exhibits higher chance to report the same address in short time window than the other detectors. *Scan2* observes much larger number of IP addresses than any other honeypot in our datasets. The reason why it does not also report the highest number of alerts is due to a firewall protecting the organizational network including the honeypots (blocking some of the most often scanned TCP ports).

The measured repeatability of alerts is however not strictly proportional to the size of observed IP space. Detectors *scan3* and *scan8* observes much larger IP space than the others but they do not report much more repeated alerts. This is because of discrepancy in the scale of an event necessary for detection by a flow-based system and by a honeypot. A honeypot can report even a single connection attempt on some of its addresses, a flow-based detection system must however see a lot of unsuccessful connection attempts before it can be sure enough that it is not a legitimate traffic.

The situation is different for SSH bruteforce attacks. These must consist of a lot of connections by definition, so all of them can be captured by a flow-based detection system as well as a honeypot. And since *sshbf3* monitors much larger IP space than the other detectors, it is much more likely that it detect the same address attacking more than once within a short time.

In summary, regardless of the detector, $P(l)$ always grows as l increases with growth rate slowly decreasing. This is a common trend. The absolute percentage of repeated detections is however heavily dependent on a particular detection system and the environment it is deployed in.

So far we have analyzed the probability that an alert will be followed by another alert with exactly the same address. We can relax this condition and look for alerts with addresses with the same prefix. Fig. 3 shows $P(l)$ distribution after such modification for prefix length of 24 bits. It thus shows the probability that an alert will be followed by another alert with an address from the same /24 subnet. Comparison with Fig. 1 reveals that repeatability of SYN flood and web access alerts does not increase significantly when /24 prefixes are used instead of full addresses. This is because there are quite low number of sources in these datasets and they rarely have a common /24 prefix.

However, the use of prefixes have very significant impact on other types of alerts. The increase of $P(l)$ ranges from 5% to almost 20%. In the case of scans and web bruteforce attacks the increase is significantly higher for longer time windows that for the short ones, resulting in less curved lines than those in Fig. 1. This may be explained by misbehaving hosts changing their addresses in the order of days, but only in small IP range, so they stay in the same /24 prefix. Such behavior is common in environments with dynamic assignment of IP addresses.

Taken all datasets together, only 40% of /24 prefixes appears only once (compare to 70% of alerts not repeated when considering individual addresses). These results suggest that it might be useful to model reputation not only for addresses but also for prefixes. It however needs more in-depth analysis which we leave for future work.

In summary, this subsection shows that a significant fraction of detected malicious events are done by addresses reported previously and that this behavior is well observable even when only a small part of the Internet (a single NREN) is monitored. The results of the analysis may be useful for development and tuning of alert correlation methods and reputation modelling systems as it gives clues on how long is it useful to store alerts in order to detect the addresses attacking repeatedly.

B. Correlations between individual datasets

In this section we look at correlations between sources of different types of malicious traffic, i.e. how many addresses from one dataset group can be found in another group as well, where datasets are grouped by their type of malicious traffic. Datasets with SSH scans only (*scan8* and *scan-ext*) are put into its own group here in order to study correlations of these scans with SSH bruteforce attacks.

The correlations are shown in Tab. II. A number at row r and column c tells a percentage of addresses in dataset r that are also present in dataset c. For example, 44% of the addresses performing bruteforce attacks on SSH (*sshbf*) were also reported as scanners (*scan*). Numbers in parentheses represent the count of unique addresses in each dataset.

The highest correlation is between scans reported by systems observing SSH traffic only and other scan reports. This is not surprising since it is expected that scans of SSH are captured by generic systems, too. In fact, it tells us that 17% of SSH scanners would be undetected without the sensors specialized to SSH traffic.

Very high correlation is also between web attacks and scans. More than 80% of addresses attacking web services were also reported as scanners. Most of them probably scanned port 80 only, but 5% of them were also reported by SSH-only detectors as scanners and 4% as originators of SSH bruteforce

TABLE II: Percentage of addresses common to pairs of dataset groups.

	scan	scan (SSH)	sshbf	webbf	web	synflood
scan (513240)	—	0.95	0.78	0.00	0.33	0.01
scan (SSH) (5869)	83.1	—	46.1	0.00	1.79	0.41
sshbf (9030)	44.4	30.0	—	0.00	0.96	0.04
webbf (4640)	0.39	0.00	0.00	—	0.00	0.00
web (2080)	82.1	5.05	4.18	0.00	—	0.67
synflood (155)	45.8	15.5	2.58	0.00	9.03	—

attacks. This means that there is a small part of attacking IP addresses which performs very different kinds of attacks. Similar results can be observed for SYN flood attacks where 45% of attackers were also reported as scanners.

Another case is correlation of scanning and bruteforce attacks on SSH. It is known that SSH bruteforce attacks are usually preceded by scanning of TCP port 22 from the same source [5]. Our results however show that it is not so common – only 30% of attacking addresses were also detected as scanners. This might suggest that some attackers get their lists of addresses to attack by another way than by scanning the network themselves.

An interesting case of low correlation is the *webbf* dataset. Although bruteforce attempts to log in via SSH and via Web forms exhibit similar characteristics with respect to repetition of alerts (as shown in Fig. 1), source addresses of these types of attacks are completely different. In fact, the *webbf* dataset have almost no addresses common with any other dataset. This means that addresses trying to login to web services using brute force are rarely involved in other types of malicious activities and vice versa.

V. Conclusion

We have shown that known characteristics of malicious traffic found previously in data from blacklists and from globally deployed arrays of sensors are valid when observing traffic in a local network as well. Also, we show it on a different kinds of attacks than those studied before. Knowledge of these characteristics is important when designing IP reputation systems

The analysis of temporal characteristics of alerts shows that some IP addresses can be detected as malicious repeatedly if the information about previously reported attacks are stored long enough. However, around 70% of addresses were reported only once throughout all our datasets. We believe that that portion will be lower if we observe larger part of the Internet. This may be achieved by sharing incident reports among network operators, for example. Also, the chance to see more than one report from the same address is probably significantly affected by dynamic address assignment, which causes that a single malicious host may have many different IP addresses (usually in small range) over time. This theory is supported by the fact, that if we consider /24 prefixes only instead of full addresses, the number of non-repeated alerts lowers to 40%. Ability to identify individual hosts regardless of their current IP address would therefore be very useful (although it may pose privacy issues). Or at least an ability to detect prefixes with dynamic address assignment might be of use – such addresses should be kept on blacklists for shorter time than those assigned statically, for example.

Many characteristics shown by our analysis might be expected. However, we also found some surprising ones. For example, while it is common to expect that bruteforce attacks on SSH are usually preceded by a network scan from the same source, our data show that it is true in 30% of cases only. 70% of password-guessing addresses were never detected as scanners.

While the analysis is based on data mostly from the CESNET NREN, we believe the results are valid for other networks of similar size as well.

As our future work we plan to compare the data from our local detectors and honeypots with data from publicly available global blacklists. We also plan to analyze distribution of sources of malicious traffic by their geographic location and by their autonomous systems.

Acknowledgments

This work was supported by the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070), funded by the European Regional Development Fund and the national budget of the Czech Republic via the Research and Development for Innovations Operational Programme, as well as Czech Ministry of Education, Youth and Sports via the project Large Research, Development and Innovations Infrastructures (LM2011033). Also, this research has been partially supported by the CESNET Large Infrastructure project no. LM2010005 funded by the Ministry of Education, Youth and Sports of the Czech Republic.

We also want to thank to operators of the detection systems in the CESNET network and to developers of the Warden system. Special thanks goes to Rick Hofstede from University of Twente for sharing results of their deployment of SSHCure.

References

[1] G. C. M. Moura *et. al.* Internet bad neighborhoods: The spam case. In *Proceedings of CNSM*, 2011.

[2] G. C. M. Moura *et. al.* Internet bad neighborhoods aggregation. In *Proceedings of NOMS 2012*, pages 343–350, USA, 2012.

[3] J. Zhang *et. al.* Characterization of blacklists and tainted network traffic. In *PAM*, volume 7799 of *LNCS*, pages 218–228. 2013.

[4] M. Javed and V. Paxson. Detecting stealthy, distributed ssh brute-forcing. In *Proceedings of SIGSAC CCS*, pages 85–96, USA, 2013.

[5] L. Hellemons *et. al.* SSHCure: A Flow-Based SSH Intrusion Detection System. In *Proceedings of the 6th AIMS, Luxembourg*, volume 7279 of *LNCS*, pages 86–97, Berlin, 2012.

[6] M. P. Collins *et. al.* Using Uncleanliness to Predict Future Botnet Addresses. In *Proceedings of IMC*, IMC '07, pages 93–104, USA, 2007.

[7] G. C. M. Moura. Internet bad neighborhoods. In *Ph.D. dissertation*. University of Twente, 2013.

[8] G. C. M. Moura, R. Sadre, and A. Pras. Internet bad neighborhoods temporal behavior. In *Proceedings of NOMS*, 2014.

[9] A. Ramachandran and N. Feamster. Understanding the network-level behavior of spammers. *SIGCOMM Comput. Commun. Rev.*, 36(4):291–302, Aug. 2006.