Tracking global wide configuration errors

Radu State, Jerome Francois, Olivier Festor

Tracking global wide configuration errors

Jerome Francois, Radu State, Olivier Festor
email :(francois, state, festor)@loria.fr
Management of Dynamic Networks and Services
Laboratoire Lorrain d'Informatique et de ses Applications de Lorraine
Nancy, France

***Abstract*— Both honeypots and network telescopes are extremely powerful tools to analyze malicious network activities. In this paper we evaluate these two concepts by looking at data obtained from two such sources over a common time period. Our viewpoint is biased towards a network management perspective and we will present in this paper patterns and trends in misconfigured network devices. We will show that simple to avoid and well known errors are still present in real deployed networks and that patterns of faulty configuration can be put into evidence.**

I. Introduction

Over the past years, honeypots and network telescopes emerged as highly efficient toolkits to capture and analyse malicious network usage. The differences between these two concepts are due to the reactiveness and scale. While a honepot serves as a decoy and should mislead a hacker into unveiling patterns of malicious usage, a network telescope is usually a large unused network having as only purpose to capture traffic destinated towards it. Both concepts have shown their practical values to both the research and operational communities, although as of today no comparative study has been yet done. In this paper we analyse data obtained from a large network telescope [1] and a distributed honeypot platform [2] and we look at this data from a network management perspective. Although data related to traces from denial of services attacks is highly informative from a security point of view, we were challenged in our work to see if we can discover patterns of network misconfigurations. Our paper is structured as follows : we will start in section 2 with an introduction to honeypots and backscatter traffic. Section 3 presents our analysis and research on observed misconfigurations. Finally, section 4 presents related works and section 5 concludes the paper.

II. Backscatter packets and Honeypot

A. Network telescope

One way to detect network intrusions is to monitor network traffic and detect malicious flows. An alternative is to monitor only the traffic which has no reason to exist. A network telescope is a range of unused IP addresses which are monitored in order to observe malicious traffic because normally no regular communications should exist to these addresses. A telescope is characterized by the length of the prefix of the subnetwork which is observed and this is the major characteristic which determines the amount of data gathered by the monitoring process [1]. Traditional telescopes are passive and store information only about incoming traffic , but some more advanced telescopes do exist, capable to perform more active monitoring [3].

B. Backscatter packets

Denial-of-service attacks are traditional attacks aiming at crashing a specific service or to degrade its performance. This can be achieved by flooding the victim with TCP packets with the SYN flag set in order to force the targeted host to deal with the packets and send the corresponding responses. Because the attacker aims at securing his privacy, spoofed IP can be used and forged packets are sent to the victim. The packet is received and the response is generated and sent to this spoofed destination. The attacker is not overloaded by the response packets of its own attack and additional traffic will hit a third party (the spoofed source). These response packets are called backscatters packets [4] [5]. For an excellent introduction to these topics please see [6]. The figure 1 shows a simplified scenario, where a hacker uses 3 spoofed IP addresses : one is assigned to a real machine and the others are not assigned to real operational machines, but instead are monitored by a network telescope.

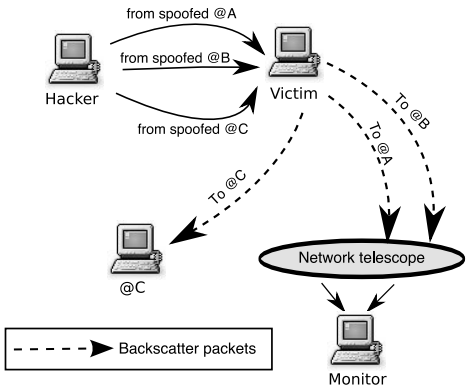

Fig. 1. Backscatter principle

C. Honeypots and HoneyNets

In [2] a honeypot is defined as "an environment where vulnerabilities have been deliberately introduced in order to observe attacks and intrusions". Therefore, honeypots are computers connected to the Internet with no apparent protection to intrusions. Moreover, most of time, the honeypots are not

used for normal activity and all traffic is abnormal. Contrary to the network telescope which monitors a large range of IP addresses and gathers a lot of data, the honeypots can be deployed in different locations and monitor only few addresses. The goal of a distributed honeynet is to obtain statistical information about intrusions by grouping together informations of the different honeypots in a single database.

III. Detecting misconfigurations

A. Sources

		Telescope	Honeypot
#monitored addresses		16 777 216	129
Number of incoming packets	05	334 866 721	475 519
	06	106 514 961	1 211 820
	07	0	1 495 525
	08	457 980 950	1 821 534
	09	170 139 630	1 371 280
	10	0	2 317 525
	11	566 326 329	2 292 083
	12	69 716 055	1 451 770
Number of unique IP addresses	05	184 282	18 392
	06	121 828	39 419
	07	0	34 011
	08	229 719	49 076
	09	168 631	60 666
	10	0	77 032
	11	281 084	84 485
	12	162 607	82 500
Size of data	05	23,7 GB	69 MB
	06	7,6 GB	176 MB
	07	0 GB	217 MB
	08	24,1 GB	264 MB
	09	12,3 GB	199 MB
	10	0 GB	337 MB
	11	40,4 GB	333 MB
	12	12 GB	211 MB
Number of days	05	6	all
	06	3	all
	07	0	all
	08	7	all
	09	3	all
	10	0	all
	11	7	all
	12	2	all

TABLE I

Global information about the data. The months are represented in number (05, 06, 07...)

To compare the honeynet and telescope data, we have chosen to analyse data from a common timeframe to both such platforms : May 2004 to December 2004. For the telescope we have used the data from a very large telescope (the CAIDA's project for which only backscatter data are available [7]). This is a /8 telescope and so monitors 2^{24} IP addresses [1]. The data are divided into several files : one by day and represents a total of about 1.7 billions captured packets. However, the packets were not captured during the entire period but only for some days. The data from the honeynet is not so abundant, but covers all days of the period. Because the daily quantity is not relevant with respect to the telescope, all the observations have been grouped on a monthly basis. We have used the data from the Leurre.com project which is composed of 43 honeypots, where each emulates under VMWare 3 different vulnerable operating systems. For more details about the (H2) platform please see [8]. For the period from May to December 2004, 12437056 incoming packets have been captured. The table I shows a summary about the data.

B. Abnormal source addresses

During a preliminary analysis of the data, we were amazed by the large quantity of observed IP addresses that should in theory never appear on the Internet. Several factors jointly produce them : misconfigured enterprise routers/firewalls, missing ISP level ingress/egress filtering and maybe defective devices. The table II gives a summary of such addresses as well as their target deployment usage.

Range	Description
10.0.0.0 → 10.255.255.255	Class A private addresses
172.16.0.0 → 172.31.255.255	Class B private addresses
192.168.0.0 → 192.168.255.255	Class C private addresses
224.0.0.0 → 239.255.255.255	Class D multicast addresses
240.0.0.0 → 255.255.255.255	Class E addresses reserved for experimental use
127.0.0.0 → 127.255.255.255	Loopback addresses
0.0.0.0 → 0.255.255.255	addresses of network 0 (class A)
169.254.0.0 → 169.254.255.255	addresses of DHCP client which can't obtain an address from the server
192.0.2.0 → 192.0.2.255	Loopback addresses

TABLE II

Abnormal source addresses on Internet

1) Backscatter analysis: The left barchart of figure 2 shows the proportion (per 100 000) of the different types of abnormal addresses in comparison with the total number of unique IP addresses for the observed days. This graph allows to observe both the main types of abnormal addresses and their corresponding global proportion.

There is a category which is about constant (colored in black). It is the proportion of network 0 addresses (class A). Normally 0.0.0.0 can be used only as source broadcast address on local segments but not on the global Internet. However the global proportion increases significantly from June to August with peaks in June, at the end of August and the beginning of September. Very strangely is also the apparition of multicast addresses as source addresses. Multicast addresses can be only used as a destination address and will never appear as source addresses. Moreover this increase in abnormal addresses is also due to private IP addresses used in outgoing reply packets. These packets are received by the telescope (and for these packets the source appears to be a private IP address). The most probably source of these packets are misconfigured routers/firewalls/NATs. These packets are generated by victims, such that we can safely assume that the majority is not malicious. A second explanation is that these victims or a subset of them detected the attack and replied with forged packets, possible for instance with tools like

IPPersonality [9]. This increase can be also caused by an ISP deploying some new policy based routing rules, which were misconfigured. The concerned computers are connected to Internet but don't receive the responses of their own requests. Another justification of the apparition of private addresses (the class C for instance, which are generally used by home users) are a definite evidence of misconfigured network devices. However, the main issue is that the ISP does not block these addresses.

We observed that 169.254.0.0 traffic is not visible in the telescope, although it appears in the honeynets. This traffic is due to hosts which have do not receive DHCP replies. Since the traffic in the telescope only have backscatter data, this traffic remains invisible to a telecope, because the concerned hosts do not become the target of attacks. However, these hosts might be infected by worms and as such can directly initiate outgoing traffic to the Internet.

2) Honeypot data: We performed a similar analysis with the data from the honeynet (at the right on the same figure 2) but in this case a bar represents a month period. The results show a different pattern than the backscatter analysis. First the graph shows two peaks but not at the same time. The first in May and the second in July. The usage of private class IP addresses is also significant and the explanation might be the same i.e. the misconfiguration of local network and providers that don't do ingress filtering. However the main type of abnormal IPs is the range of addresses automatically assigned by a computer when the DHCP server don't respond to its request for obtaining an address. The cause is probably due to local networks with a non valid configuration of the DHCP service.

For comparing the two traces, we had to compare data from backscatter traffic observed from the telescope with data (directly incoming and backscatter) from the honeynets. We could not rely entirely on only the backscatter traffic from the honeynets due to the lack of massive datasets.

C. MS Windows specific ports

The MS Windows operating systems uses a series of defaults ports for its own network protocols : ports 137, 138, 139 and 445. The Netbios service is designed for sharing resources on a local network and this port should not be available from the Internet. To prevent these attacks, these ports should be filtered by a firewall.

1) Backscatter analysis: The figure 3 shows the number of unique IP addresses with an open port per 100 000 unique addresses that suffered denial of service attacks. The ports 137, 138 and 445 seem to be protected even if there is a little peak for the port 445 in November. However it's clear that the port 139 is less filtered as we can see on the several peaks of the graphs. It seems that in 2004, professional networks and home computers were generally protected by firewalls contrary to some years before.

2) Honeypot analysis: The honeypot data contains only one IP address having the port 139 open, such that the use of honeypot is not a good way to detect this kind of misconfiguration. Only a telescope with a large IP range can efficiently detect it. However you can notice that the only visible port is also the one which is the most frequently observed as opened by the telescope.

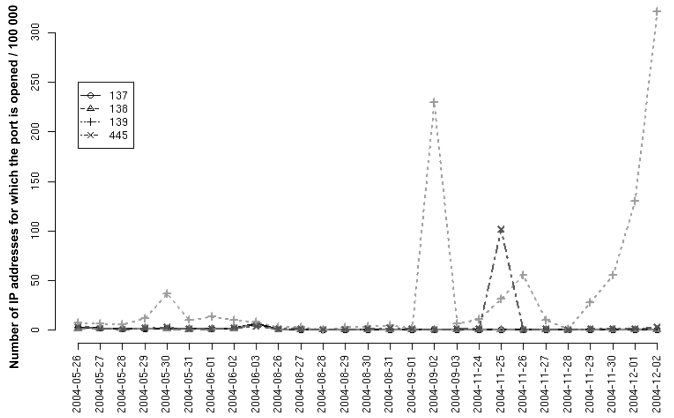

Fig. 3. Number of unique IPs with an open port per 100 000 unique IP addresses and according to each specific windows port. The chart represents the backscatter data.

D. Analysis of ICMP 'Destination unreachable' message

When a host connects to another host which is not available, an ICMP message is sent to the source with the type 3 equal to 'Destination unreachable message'. An additional code [10] is also used to provide additional information. We analyzed the following 8 codes in our work :

- 0 : net unreachable
- 1 : host unreachable
- 2 : protocol unreachable
- 3 : port unreachable
- 4 : fragmentation needed and don't fragment was set
- 9 : communication with destination network is administratively prohibited
- 10 : communication with destination host is administratively prohibited
- 13 : communication administratively prohibited

Some firewalls will typically answer with codes 9 or 10 to show that a device or service is filtered. Although such information can be very helpful when troubleshooting a network, it can leak information about existing devices/open ports to an attacker and could determine him to try more advanced reconnaissance techniques. Firewalls that well engineered , configured by more security conscious network managers might directly reply with a RST. We can observe a configuration pattern evolving over the year 2004. In the beginning of the year, firewall configuration was globally verbose and attacker friendly. Administrative prohibited (code 13) type of messages and host unreachable (code 1) are used abundantly. This type of configuration leaks important information to an

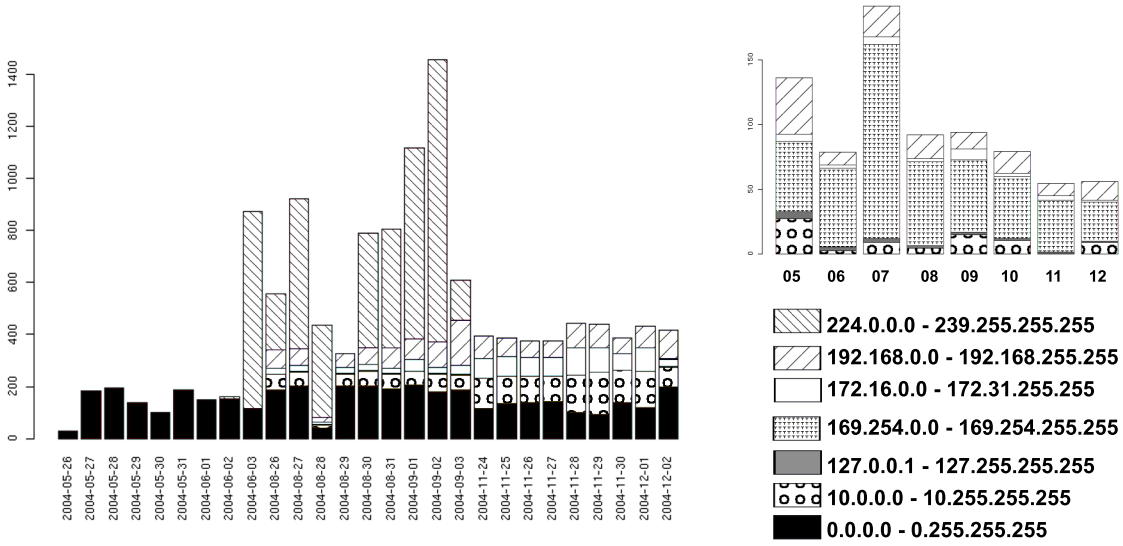

Fig. 2. Number of unique IP addresses of the different categories of abnormal addresses per 100 000 unique IP addresses. (Left : backscatter data, right : honeypot data by month)

attacker. Receiving an ICMP message with an administrative prohibited code implies that the service port is open and that maybe a simple access control based on the source of the IP packet disallows any communication. In this case, spoofing this information and blindly attacking the service can lead to a success (for the attacker). Similarly, receiving an ICMP with a code 1, is helpfull when scanning a network. This is called also inverse scanning, because learning which hosts are not on the network allows to infer the IP addresses that are allocated on the same network. We see that a global trend towards a less verbose configuration is visible and more and more firewalls reply with a port unreachable message. The same behavior can be observed on the honeynets.

The drastic change observed in 4 can be du to a novel configuration of a firewall. The essential fact however is the comparison between the telescope and honeynet, for which we can state only hypotheses. In the telescope data we have a drastic change due to 2 main reasons :

- high decrease of the code 13 : firewalls configuration changes to hide the firewalls
- switching between code 1 and code 3 : the attacks are targeted towards real host contrary to before

For the honeynet data, we observe the same changes but these are not so drastic (except for the code 1). Moreover the changes do not occur at the same time period.

The important fact is that the two methods give somehow identical results but for the telescope the change is much more important, then the smooth change observed on the honeynet. In fact, all IP addresses monitored by the telescope are from the same subnetwork. If an attacker spoofs using this range of addresses (for instance by bad random generator, use of a specific subnetwork...) and targets the same hosts, the telescope observes a drastic change when the firewalls of these hosts change their policy.

For the honeynet, the picture looks totally different. For this data we have contiguous data and the IP addresses are well distributed in the address space.

E. Attacked services

A natural question is related to which services are the most attacked services and whether these services should have been accessible to the Internet. We did this analysis on a day by day basis and some of our main findings follow :

- The most attacked port and consistently ranked number 1 over all this period is port 80 : it seems that web servers are the major target of denial of service attacks. This service should be accessible to the Internet such that we can not point out a misconfiguration.
- port 6667 shows up frequently in the attacks. This port is typically used for IRC talks (or IRC anonymizing proxies like psyBNC). We suppose that these attacks are targeted at specific servers and can be associated to Internet war games waged to take the control of a a IRC channel. Similarly to the previous case, this service/port should always be accessible, unless a compromised machine is used to serve as proxy to cover a malicious bot master.
- Name Servers (port 53) are also attacked (although to a lesser extend than IRC). For our study, we did not want to be invasive and to check the function of the attacked servers, but in general a well secured DNS server should never allow a zone transfer (if important information can be leaked out) such that these cases are very important examples of misconfigurations.
- Attacks against BGP routers (identified by received SYN/ACKs from a port 179) are also highly interesting and can be observed, since these attacks aim at either de-

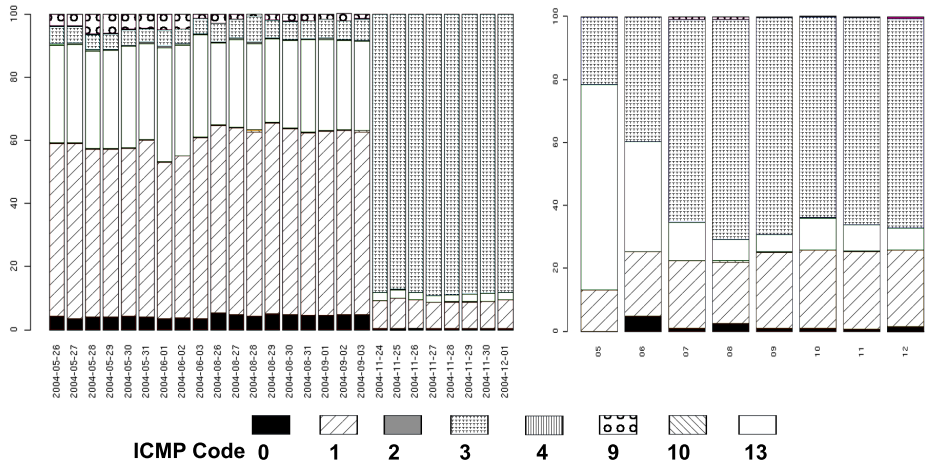

Fig. 4. Proportion of the different ICMP codes for the icmp type 3 (Destination unreachable). Backscatter data are represented at the left and honeypot data is at right

Port	Vulnerability
1011	Augudor
1025	Spybot
1433	Spybot
6000	Lovgate
7000	SubSeven
7001	Freak88
7300	NetMonitor
8000	Gaobot

TABLE III

SOME SERVICES WHICH ARE IN THE MOST ATTACKED SERVICES AND WHICH PRESENT KNOWN VULNERABILITIES

connecting a network domain, or can serve as preliminaries for a routing prefix hijack. Similarly to the previous case, BGP connections should be filtered and available only to well known peers having contractual bindings

The table IV compare the most attacked services between the telescope (3 days) and the honeypot for May. The overlap of the ports is small : only the port 80. However, if we consider table V for September, the views match because 7 common ports appear in both the honeynet and in the telescope data. To conclude, even if sometimes, the two methods allow to get the same results, it appears that the results can be different and therefore the methods can be considered as complementary.

Moreover, in these tables (IV and V) an interesting fact is to have the port 7000 which is known as a backdoor. In the table III the ports which are in the most attacked services with known vulnerabilities are listed. These ports are commonly used by backdoors or ports used for the spreading of a worm. While it's unclear for us, why someone would do a denial of service attack against them, having these ports open and accessible to the Internet are clear signs of misconfigured routers/firewalls. Even if internal machines were compromised by an internal/external user, an additional filtering properly secured firewall to deny access to them, would have limited the exposure and the impact of the intrusion.

IV. RELATED WORKS

In [4], the authors used a telescope to monitor denial-of-service attacks activity. Some limitations are clearly defined like that reflector attacks are possible, maybe the addresses are not randomly generated in all the address space and the ISP can do ingress filtering. In [5] another approach is chosen to analyse backscatter data by using spectral analysis and this method showed that the spoofed addresses are not generated with a good randomness property. The author in [3] proposes another type of telescope which can respond to the request of the attackers and get more information about them. This kind of architecture is more complex and demands to monitor less address space but in order to limit the required bandwidth, some additional sampling methods are needed. Finally, some statistics like the mean length of attacks of the mean number of targeted subnetworks can be computed and the authors have observed that only 5% of the traffic is backscatter traffic. Honeypot analysis is presented in [2] and [11] and some of the described results cover the discovery of new threats, the localization of the attacks, the operating systems of the attackers as well as their behavior and network scanning activities.

V. CONCLUSIONS

This paper presents a preliminary study of very large network data from both a network telescope and a distributed honeynet. We were challenged in our work by two questions. The first one concerned the differences and similarities in views from such different platforms and the identification of peculiar properties of each of these. Our second motivation was to analyse configuration errors and patterns of misconfiguration. We showed that backscatter data available from the network telescope is very useful for these purposes. The main difficulty encountered during our work is related to processing such large datasets : data counts to more than 120 GB and this task pushed our computational resources to their limits. Future

Honeypot May		Telescope 2004-05-26		2004-05-27		2004-05-28	
80	35 (63.64)	80	734 (7.61)	80	973 (10.03)	80	980 (16.27)
6667	5 (9.09)	21	15 (0.16)	21	17 (0.18)	139	14 (0.23)
3389	3 (5.45)	6667	15 (0.16)	4662	15 (0.15)	21	13 (0.22)
7000	3 (5.45)	139	13 (0.13)	139	13 (0.13)	22	11 (0.18)
1107	1 (1.82)	1002	12 (0.12)	25	11 (0.11)	113	10 (0.17)
1205	1 (1.82)	22	10 (0.10)	8080	11 (0.11)	25	10 (0.17)
1214	1 (1.82)	8080	10 (0.10)	110	10 (0.10)	8080	9 (0.15)
1235	1 (1.82)	110	9 (0.09)	113	10 (0.10)	443	8 (0.13)
1254	1 (1.82)	113	8 (0.08)	135	10 (0.10)	110	6 (0.10)
1271	1 (1.82)	111	6 (0.06)	22	8 (0.08)	178	6 (0.10)

TABLE IV

THE MOST ATTACKED SERVICES DURING MAY WHICH HAVE SENT SYN/ACK. THE FIRST NUMBER IS THE PORT AND THE SECOND THE NUMBER OF UNIQUE IP ADDRESSES WHICH ARE CONCERNED. THE NUMBER BETWEEN PARENTHESIS IS THE PERCENTAGE ACCORDING TO ALL UNIQUE COUPLE IP ADDRESS - OPEN PORT

Honeypot September		Telescope 2004-09-01		2004-09-02		2004-09-03	
80	116 (50.88)	80	956 (14.89)	80	1100 (19.66)	80	508 (17.69)
7000	49 (21.49)	7000	37 (0.58)	139	413 (7.38)	7000	24 (0.84)
7100	11 (4.82)	7200	13 (0.20)	7000	30 (0.54)	7100	21 (0.73)
22	9 (3.95)	7100	12 (0.19)	7100	22 (0.39)	7200	18 (0.63)
7200	7 (3.07)	21	10 (0.16)	7200	18 (0.32)	3389	12 (0.42)
7090	6 (2.63)	25	9 (0.14)	21	14 (0.25)	21	11 (0.38)
3389	4 (1.75)	22	8 (0.12)	3389	11 (0.20)	8080	8 (0.28)
21	3 (1.32)	443	8 (0.12)	22	10 (0.18)	139	5 (0.17)
113	2 (0.88)	8080	8 (0.12)	8080	8 (0.14)	6000	5 (0.17)
6667	2 (0.88)	3389	7 (0.11)	25	7 (0.13)	1524	2 (0.07)

TABLE V

THE MOST ATTACKED SERVICES DURING SEPTEMBER WHICH HAVE SENT SYN/ACK. THE FIRST NUMBER IS THE PORT AND THE SECOND THE NUMBER OF UNIQUE IP ADDRESSES WHICH ARE CONCERNED. THE NUMBER BETWEEN PARENTHESIS IS THE PERCENTAGE ACCORDING TO ALL UNIQUE COUPLE IP ADDRESS - OPEN PORT

work will address more advanced data mining and statistical analysis techniques.

Acknowledgment We thank Fabien Pouget and Marc Dacier from the Leurre.com project for their collaboration in the honeypot project. Moreover, we would like to thank Emile Aben and Colleen Shannon at CAIDA for granting us access to the telescope backscatter data. This paper was supported in part by the EC IST-EMANICS Network of Excellence (#26854).

REFERENCES

[1] D. Moore, C. Shannon, G. M. Voelker, and S. Savage, "Network telescopes," CAIDA, Tech. Rep., 2003.

[2] F. Pouget, M. Dacier, and H. Debar, "Attack processes found on the Internet," in *NATO Research and technology symposium IST-041/RSY-013 "Adaptive Defence in Unclassified Networks", 19 April 2004, Toulouse, France*, Apr 2004.

[3] V. Yegneswaran, P. Barford, and D. Plonka, "The design and use of internet sinks for network abuse monitoring," 2004.

[4] D. Moore, C. Shannon, D. Brown, and G. M. Voekler, "Inferring Internet denial-of-dervice activity," *IEEE/ACM Transactions on Computer System (TOCS)*, 2006.

[5] K. E. Giles, D. J. Marchette, and C. E. Priebe, "On the spectral analysis of backscatter data," in *Hawaii International Conference on Statistics and Related Fields*, 2004.

[6] J. Mirkovic, S. Dietrich, D. Dittrich, and P. Reiher, *Internet Denial of Service : Attack and Defense Mechanisms*, ser. Radia Perlman Computer Networking and Security. Prentice Hall PTR, december 2004.

[7] C. Shannon, D. Moore, and E. Aben, "The caida backscatter-2004-2005 dataset - may 2004 - november 2005, http ://www.caida.org/data/passive/backscatter_2004_2005 _dataset.xml."

[8] F. Pouget and T. Holz, "A pointillist approach for comparing honeypots," in *DIMVA 2005, Conference on Detection of Intrusions and Malware & Vulnerability Assessment, July 7-8, 2005, Vienna, Austria - Also published in LNCS Volume 3548*, Jul 2005.

[9] G. Roualland and J.-M. Saffroy, "http ://ippersonality.sourceforge.net."

[10] IANA, "http ://www.iana.org/assignments/icmp-parameters," 2005.

[11] F. Pouget and M. Dacier, "Honeypot-based forensics," in *AusCERT2004, AusCERT Asia Pacific Information technology Security Conference 2004, 23rd - 27th May 2004, Brisbane, Australia*, May 2004.