# Leurre.com:
# on the Advantages of Deploying a Large Scale Distributed Honeypot Platform

F. Pouget, M. Dacier, V.H. Pham
Institut Eurecom
2229, route des Crêtes, BP 193
06904, Sophia-Antipolis, France
{pouget, dacier, pham}@eurecom.fr

**Abstract**

There are several well known techniques to observe criminal activities on the Internet by monitoring its traffic. One option consists in using global telescopes or dark nets which offer interesting views of global trends. Another solution consists in centralizing firewall logs and intrusion detection system alerts to extract some information. In this paper, we advocate the usefulness of a third approach that focuses on the need of local views to get more precise information on some attacks. With this idea in mind, we have developed and deployed for the last six months a distributed honeypot environment in several distinct countries. We show in this paper that 1) local sensors present strong similarities to a certain degree, and 2) they also highlight very clear local patterns. As a conclusion, we demonstrate the usefulness of distributed honeypots and we hope to encourage more partners from all over the world to join our honeypot, named the Leurre.com

**Keywords**

Honeypots, distributed architecture, internet threats, monitoring, local observation

# 1 Introduction

There are a few approaches to observe malicious traffic on the Internet. Some solutions consist in monitoring blocks of unused address spaces. Several names have been used to describe this technique such as network telescopes [Caida, Moo01], blackholes [Song01, Morr04] and darknets [Cym04]. This technique has been used both by host-based honeypot tools [Spit04] and by wide address space monitors [Moo01, Song01, Moo02]. Some other solutions consist in passive measurement of live networks by centralizing and analyzing firewall logs or IDS alerts [Sans04, Yegn04]. Coarse-grained interface counters and more fine-grained flow analysis tools such as NetFlow [Netfl] offer another readily available source of information.

So far, nobody has investigated the possibility of using a large number of simple, cheap and similar sensors deployed all over the Internet. As a consequence, we have deployed for more than six months such platforms, thanks to motivated partners as part of the

LEURRE.COM project. In this paper, we first discuss statistical results. We show that some platforms present strong similarities but that they also exhibit very clear (and surprising) local patterns. This helps us making our point that local sensors are needed to acquire a good understanding of Internet threats. A global knowledge of the threats must be completed by a good understanding of local malicious activities. Finally, we hope this paper will incite new partners to join this project, and, by doing so, enrich the local views.

The paper is divided into four major Sections: Section 1 describes the set up we have used to collect, store and analyze malicious data. Section 2 highlights the added value of local sensors compared to other existing solutions. Section 3 presents the common features observed on several platforms. Section 4 illustrates the platforms differences thanks to a few relevant examples.  Section 5 concludes this paper.

# 2 Distributed Honeypot Setup

## *2.1 Platform Architecture*

We have presented in previous publications [DaPD04, DPDe04] some experiments based on so called "high interaction honeypots". These experiments have shown that most of the attacks are caused by a limited number of attack tools and that there are very stable processes occuring in the wild. As a follow up, we found out that low interaction honeypots, despite their ability to be easily fingerprinted, represent a suitable solution to gather statistics regarding most automated tools. Furthermore, they are a better solution from a practical deployment point of view. Indeed, we do not want our platforms to be corrupted as most of the honeypots applications intend to [Honey, Spit04]. We only want to observe the first attack waves in order to get a better understanding of current malicious activities.
As a result, we have deployed a platform similar to the one presented before, but with emulated operating systems and services. This decision is all the more justified that we have shown in [DaPD04] that most of the attacks are *blind* and do not even try to fingerprint the victim in a first step. We have developed our own platform based on some open source utilities[1]: it emulates three different Operating Systems; Windows 98, Windows NT Server and Red Hat 7.3 respectively. The platform only needs a single host station, which is carefully secured by means of access controls and integrity checks. Every day, we connect to each machine to retrieve traffic logs and check security logs. Finally, data is stored into a centralized database. The dump files are also analyzed by means of other utilities and this additional information is collected as well.
A detailed description of the database architecture lies outside the scope of this paper. Instead, we refer the interested reader to our previous publications on this topic. However to make this paper self contained, we list here after the kind of information that are derived from the dump files:

- IP geographical location
- Domain name resolution

[1] The platform implements a modified version of Honeyd at this time [Prov04].

- Passive OS fingerprinting
- TCP stream analysis
- Etc

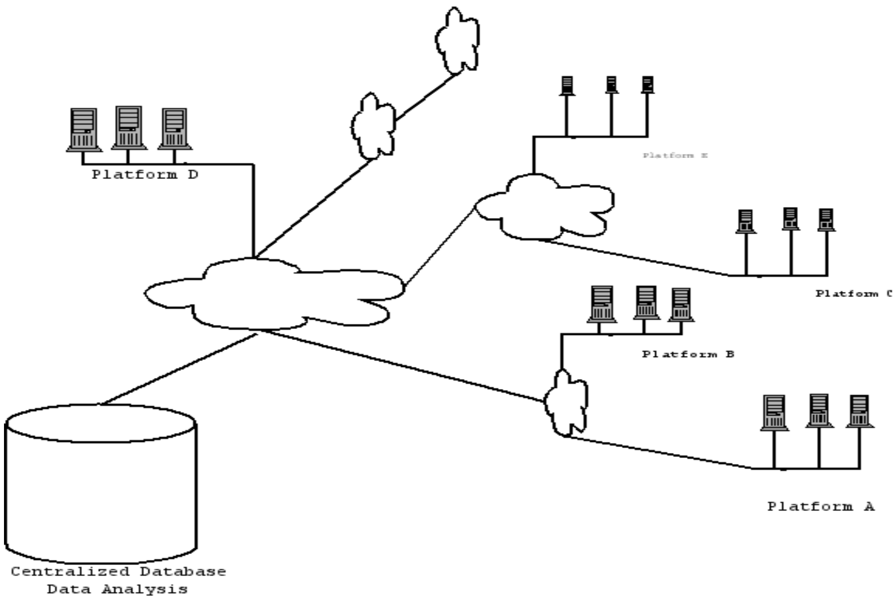

**Figure 1: Leurre.com, the distributed honeypot architecture**

In the following, we will make use of the expression *ports sequence* which we define as follows:
- *Ports Sequence*: an ordered list of ports targeted by an attack source on a virtual machine. For instance, if source A sends requests on port 80 (HTTP), and then on ports 8080 (HTTP Alternate) and 1080 (SOCKS), the associated ports sequence will be {80;8080;1080}.

## 2.2 LEURRE.COM: Platforms Positioning

We have introduced in Section 2.1 a low interaction honeypot platform emulating three virtual machines. LEURRE.COM project aims at disseminating such platforms in various places [Pou04]. Figure 1 represents a simple scheme of the distributed platform architecture. Partners are invited to join this project and to install one platform on their own premises. Eurecom takes care of the installation by providing the platform image and configuration files. The install process is fully automated thanks to a provided CD-ROM. In exchange, Eurecom gives the partners an access to the database and its enriched information[2]. A dedicated web interface has also been developed to make research faster and more efficient.
The project has triggered interest from many organizations (academic, industrial and governmental) disseminated all over the world. At this time writing, we have deployed

[2] A Non-Disclosure Agreement is signed to protect partners from each others

more than 20 platforms. Table 1 lists the countries hosting at least one platform. More are in the installation phase.

**Table 1: Leurre.com, sensors locations**

| **Countries** |
|---|
| Australia |
| Belgium |
| Colombia |
| France |
| Germany |
| Italy |
| Ivory Coast |
| Lithuania |
| Poland |
| Taiwan |
| USA |

In the next sections, we detail some of the findings highlighting the differences but also the similarities identified between platforms. For the sake of conciseness, a very limited number of examples are given.

# 3 Benefits of Local Views

## *3.1 Current Information Sources*

There are a few sources that provide information on the attacks. Generally speaking, we can classify them into four main categories:

- *Incident mailing lists:* anonymous admin members make an inventory of some incidents they observe on their own networks [Sans04,Secfo04]. However, there is a dilemna between providing too much information (at the risk of informing malicious people on their network structure) and not enough to get interesting expert feedbacks. Moreover, they often report bugs which are due to errors in the way they manage their networks.
- *Centralized reports:* some web sites such as MyNetWatchman or Dshield ask volunteers to send their firewall or IDS logs. They *analyze* such data and give output by means of monitoring consoles and graphs [Talis04]. Information is limited and not always accurate enough, as we will show in the next Section. However, this provides very good overview of current *trends*.
- *CERTs:* CERTs are reporting centers for Internet security problems [Auscert, Certcc]. They provide technical advices, they coordinate responses and they disseminate information to some given communities. They also analyze product vulnerabilities, publish technical documents and present training courses. However, in most cases, they limit the information diffusion to their *customers*.

Even if they present work of great value, they mainly participate to the propagation of 'known' important global information.

These three approaches present very interesting advantages. However, there is a lack of local and precise information of the attacks. We want to get a better understanding of local threats and thus to extract accurate information from this. Many analogies can be made here with numerous scientific fields which require permanent observations, like weathercast, migratory animals, or volcanic eruptions. The three previously mentioned approaches cannot provide this local information. In addition, Cooke et Al. have demonstrated in [Cook04] that observations made within a certain address space may not always be generalized to other address spaces. In other words, global observations of attacks can be very different from the attacks observed on one particular network. This motivates the development of local sensors. In the next subsections, we give some basic examples on how information can differ from the three previously listed information sources.

## 3.2 Related Work

We point out in this subsection that there are a few initiatives that have advocated local observations of attacks:

- *lucidic.net:* This project has been launched in some mailing lists in mid 2002 but, as far as we know, it has not led to any concrete realization at this time writing [Luci04]. The initial concept though was very close to the Leurre.com project.
- *Honeypot Farms:* This concept has been suggested by Lance Spitzner. However, it remains limited to a theoretical description in an article published in Security Focus in 2003 [Spit03] where it is explained that farming consists in deploying the honeypots in a single, consolidated location, instead of deploying large numbers of honeypots, or honeypots on every network. This single network of honeypots becomes the honeypot farm, a dedicated security resource. Attackers are then redirected to the farm, regardless of what network they are probing. However, no technical implementation is discussed.
- *The Distributed Honeypot System (DHS):* This project has been proposed by Bakos et Al in [Bak04]. The concept is very close to the Leurre.com project, as it is explained in the DHS web page that: "a Distributed Honeypot System (DHS) can be defined as a collection of honeynets or honeypots that are distributed throughout the Internet or other large network and that send their data to a central analysis point." However, there is no available data or precise information on the distributed system that has been deployed.
- *The Brazilian Honeypots Alliance*: This project focuses on the Brazilian IP address space [Braz04]. They propose to set up a network of low-interaction honeypots. First, data and results are available for Brazilian partners only. Second, they do not take benefits of their local platforms to get local information. The available information is limited to simple global analysis.

None of these initiatives have brought concrete results as far as we know. *A contrario*, our project is opened to any partner eager to access the database and look into the data. We show in the following the first results that confirm the interest of a distributed platform.

## 3.3 Few Illustrative Examples

We have listed in Section 3.1 some web sources that provide information on the attacks. However, this information often differs from local observations that can be made. There are numerous examples which validate this claim. One is presented in figures 2 and 3. In figure 2, we represent the evolution of attacks targeting ports sequence {445} on one of the French platforms. In figure 3, we present for the very same period the Internet Storm Center/Dshield report [Sans04]:

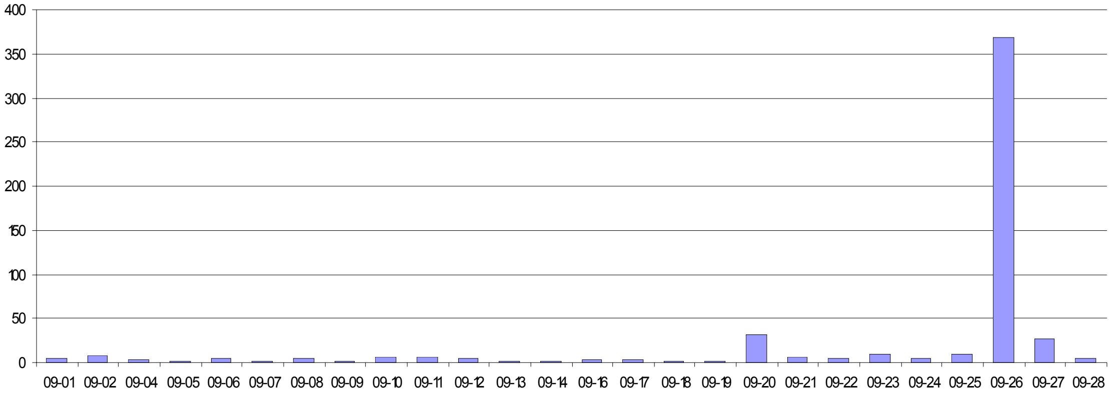

**Figure 2: Evolution of attacks targeting platform France3 on port {445}**

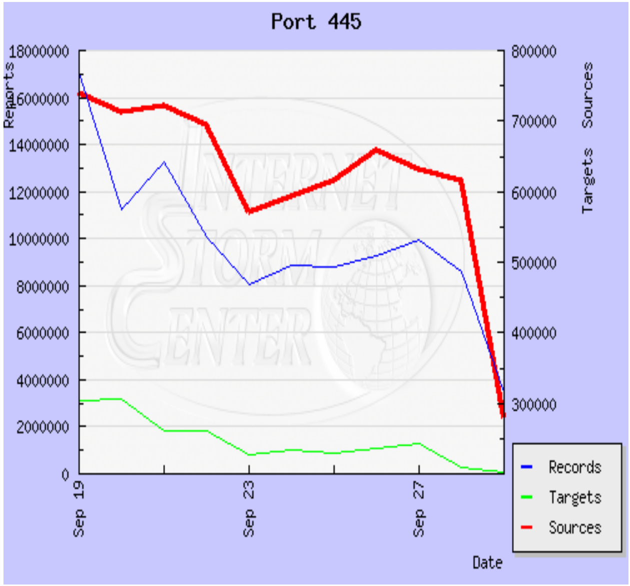

**Figure 3: Evolution of attacks reported by the Internet Storm Center on port {445}**

The peek observed on September 26th does not appear in any Dshield reports or Incident mailing list posts. The reason for this to have happened is still under investigation. Nevertheless, it clearly shows that local observations might differ from global trends. This claim is defended in [Cook04]. The authors demonstrate differences in traffic observed in class A IP ranges and smaller subnetworks along at least three dimensions: over all protocols and services, over a specific protocol and service and over a particular worm signature. Thus, there are good reasons to analyze local trends.

We show in the next two Sections that deployed sensors can present similar attack patterns, but also strong differences. These particular characteristics highlight the need of deploying more sensors in many various places.

# 4 Similarities Between Views

## 4.1 Common Attack Sources

Many similarities can be observed from the platforms we have set up. One example is the attacking IPs observed on many platforms. We find that 5% of the attack sources have been observed on at least two platforms. From these IPs, 55% are clearly identified as backscatters. Backscatters are residues of DoS attacks [Moo01]. Thus the addresses of many platforms have been spoofed and used during DoS attacks. These IPs come mainly from China. They are the victims of DoS attacks: a simple glance at the involved ports shows that they are web servers (ports 80 and 8080).

The other 45% common IP addresses are also very interesting. All of them are seen scanning the Internet. We can suppose the source machines have scanned a whole IP range. Indeed, these IP addresses are observed on the platforms belonging to the same class A. We also note that it is very rare to observe IPs on all platforms. This implies that scans are limited to some specific IP ranges. They avoid scanning the 2^32 addresses.

It seems that IPs which are observed in many platforms present very strong characteristics so they can be easily identified. We have introduced two examples, respectively backscatters and large scans. Some others need to be clarified and precised. But this is out of the topic of the paper. The objective is essentially to show that similarities exist between such platforms. In a future work, these results will be compared with big telescopes observations. We can also imagine blacklisting such IPs, once they are clearly identified.

## 4.2 Common Attack Characteristics

We list in this Section some other similarities that can be easily identified between platform attacks:

1. Most of the attacks come from Windows machines. More precisely, it is found that about 80% to 95% of the observed attack sources are Windows machines[3]

[3] Operating Systems have been determined by means of Passive Fingerprinting techniques [Disco,Etter,Pof]. Generally speaking, we avoid any active technique that would alert the attackers.

independently of the chosen platform. It is not surprising since many attacks target Windows ports (see next section). Thus, it simply means that most of the attacks propagate through Windows stations.

2. There are some similarities between domains. As illustrated in table 2, we note that the ratio of machines identified as personal computers is quite constant over all the platforms. This value is quite high, as 35% of the machine names we have observed correspond to personal computers. With more details, the values are obtained by simple pattern matching from the DNS reverse resolution tables we get. Expressions like '%dsl%', '%dialin%', '%cable%' tend to indicate personal computers. One drawback is the high number of undetermined names from the DNS reverse name resolution. In average, 40% of the addresses are undetermined. Another possible solution would be to associate observed IPs with known IP ranges allocated by ISPs to particular connections.

**Table 2: Domain Name Analysis**

| **Platforms** | **% attack sources clearly identified as personal computers** |
|---|---|
| France | 33 |
| Germany | 28 |
| Taiwan | 48 |
| USA1 | 43 |

# 5 Differences Between Views

## *5.1 Attack Frequences*

Table 3 presents the average number of distinct attack sources observed by day[4] for some of our platforms. The number of received packets is also given. As one can see, results can be very different:

- Between platforms in different countries: for instance, one platform in Germany is attacked in average twenty times more frequently than the one in Lithuania.
- Between platforms in a same country: in France, the two platforms named France1 and France2 correspond to two platforms located in two academic networks. This is why they have similar attack profiles. However, they belong to two distinct class A blocks. On the other hand, the so called "France3" platform presents a very different attack pattern. France1 is an academic network while France3 is an industrial one. This might explain why the first one is attacked 5 times more than the second one. This conjecture needs to be validated. It seems, however, to be confirmed by a closer look at the other platforms.

[4] Electrical shortcuts and network interventions have prevented us to collect data for some days on a few platforms. These inactive periods are not taken into account in the table values.

**Table 3: Attack Frequencies per platform**

| Countries | # attack sources/day | # received packets |
|---|---|---|
| Australia | 89 | 364 |
| France1 | 132 | 1060 |
| France2 | 160 | 919 |
| France3 | 609 | 4604 |
| Germany | 924 | 25799 |
| Lithuania | 49 | 632 |
| Taiwan | 287 | 22571 |

Taiwan and Germany face more attacks each day than the three French platforms all together. However the Taiwanese platform is attacked by fewer sources –four times less to be precise – than the German one. Surprisingly enough, the results are different in terms of packets. This simply means, as plaforms are totally similar, that, in the average, attacks against our Taiwanese platform are made of more packets than attacks against the German one. This is confirmed in the following when looking at the attack details.

## 5.2 Attack Tools

We have presented in the previous subsection the frequence of the attacks per platform. One first conclusion is that platforms are not equally targeted every day. This subsection intends to show that the attack types also differ from one platform to another. As for an illustration, we present in table 4 the top 10 ports sequences for each platform. They are listed for each column in decreasing order of importance. For instance, column 2 represents the 10 more observed ports sequences attacks on platform France1. We do have more than 3 (resp. 1) platforms in France (resp. Germany, Lithuania, Taiwan) but we focus on a limited number in the following for the sake of conciseness.

**Table 4: Top 10 targeted ports sequences per platform**

| Australia | France1 | France2 | France3 | Germany | Lithuania | Taiwan |
|---|---|---|---|---|---|---|
| {1026} | {445} | {445} | {445} | {445} | {1433} | {135} |
| {5554,9898} | {1026} | {139} | {135} | {135} | {445} | {445} |
| {9898} | {1027} | {1026} | {135,4444} | {80} | {5554,1023,9898} | {139} |
| {1027} | {1433} | {1027} | {1025} | {137} | {1023} | {80} |
| {1433} | {137} | {135} | {139} | {2745} | {5554} | {1025} |
| {445} | {135} | {5000} | {5554,9898} | {1025} | {80} | {2745,135,1025,3127,6129,139,80} |
| {1029} | {139} | {1433} | {139,445} | {2745,1025,3127,6129,80} | {5554,9898} | {135,1025} |
| {1028} | {5554,1023,9898} | {5554,1023,9898} | {9898} | {1433} | {2745} | {135,139} |
| {5554} | {4899} | {137} | {2745} | {135,4444} | {22} | {4899} |

| | | | | | | |
|---|---|---|---|---|---|---|
| {4899} | {5554} | {5554} | {5554} | {5000,135} | {4899} | {5554,1023, 9898} |

Windows ports 135, 445 and 139 are the most targeted ports. In addition, attacks are clearly different on each platform. Clearly enough, we find some similar ports but:

- Orders of importance are different.
- Some ports sequences are only observed on few platforms. This is the case of {5000,135} which has been observed many times on the German platform and never on the Lithuanian one.

We need to point out that one ports sequence might be associated to several attack tools. In this case, a deeper analysis of the packets payloads and attack features is required to distinguish between all of them. We have developed a technique to clearly identify such tools. We invite the interested reader to have a look at our previous publications for more information on that [DaPD04,PoDa04]. The idea is that the previous results can be refined in terms of *attack tools* instead of *ports sequences.*

It seems quite clear that the attacks differ from one platform to another. To confirm this, we show in figure 4 the repartition of attacks on each platform depending on the number of targeted machines. As a reminder, each platform emulates three different machines. Thus we look at attack sources having targeted only one of the three virtual machines, then those having targeted two out of the three virtual machines, and finally those having targeted all virtual machines.

**Table 5: Attacks on virtual machines**

| Platforms | Attacks on 1 virtual machine only | Attacks on 2 virtual machines only | Attacks on the 3 virtual machines |
|---|---|---|---|
| Australia | 58.9% | 8.6% | 32.5% |
| France1 | 61.7% | 4.5% | 33.8% |
| France2 | 55.1% | 5.1% | 39.8% |
| Germany | 78.8% | 5.3% | 15.9% |
| Taiwan | 48.1% | 26.5% | 25.4% |

First, we have not observed attacks which have targeted the three virtual machines in another order than the increasing or decreasing numerical order of their IP addresses. Thus, these attacks are essentially sequential scans. Moreover, except for Taiwan, the percentage of attacks which target exactly two machines is very low. A deeper analysis is required to check if they are due to truncated scans or more sophisticated propagation mechanisms.

We also notice that the percentage of attacks on one machine is very high on the German platform. This tends to show that this platform faces more precise attacks. A deeper analysis is also required at this stage, like the analysis technique we have presented in [PoDa04] for a better understanding of the involved attack tools.

## 5.3 Attack Origins

The origin of the attacks can be very different from one platform to another. We show in figure 4 the origin of the attacks for 4 platforms located in France, Germany and Taiwan. We only present the 5 most important attacking countries per platform. The other countries are grouped in the '*others*' category. On each platform, around 110 countries have been observed. Thus five countries represent more than half of the attacks for many platforms.

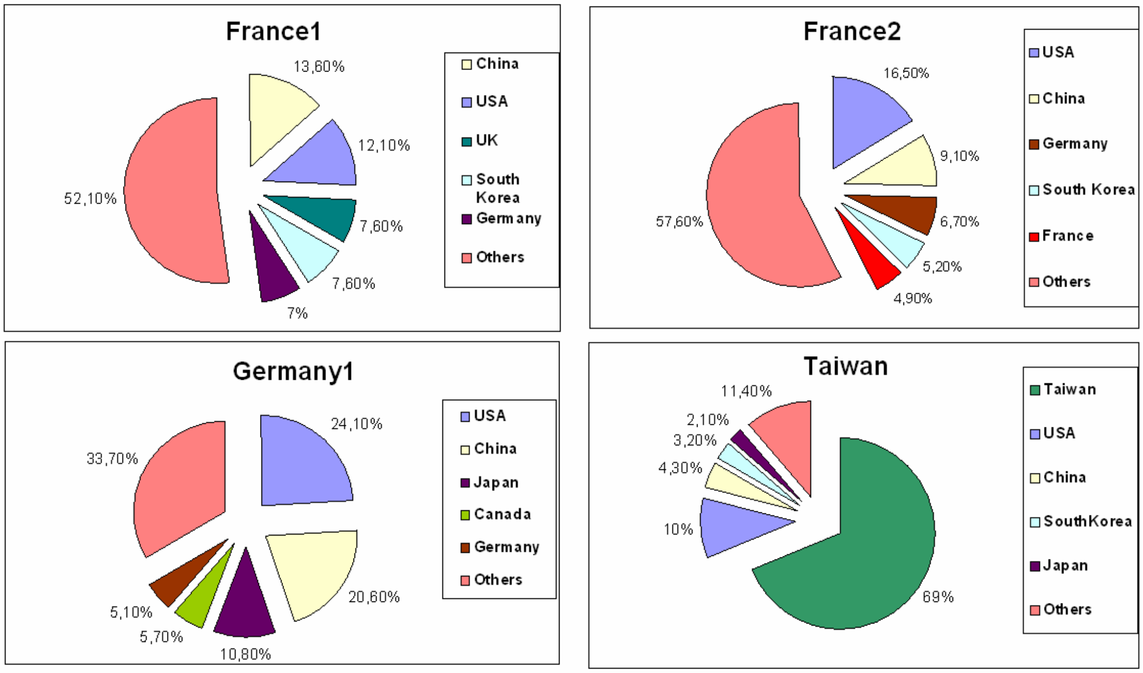

**Figure 4: Origin of the attacks for some platforms**

There are at least three points we can observe from these graphs:

- Taiwan presents very particular features. 69% of the observed attacks are coming from the same country than where the honeypot is residing. Such a phenomenon is not observed any other platforms.
- The USA and China are on the top three attacking countries, but they have different ratios on each platform.
- Some countries like Canada or UK appear on one platform only. They are less important on the other platforms.
- The two French platforms present some differences. France3 is an industrial network and we observe more French attacks than in France1, an academic network.

There is an important correlation between attack origins and platform positioning/types. Some attacking countries are observed in all platforms (with various ratios). Others are more virulent depending on the network type or the platform positioning. This confirms the main result we have noticed in the previous Sections: these platforms present different attack patterns. Platforms are targeted by different attacks coming from different countries.

# 6 Conclusion

We have presented very simple examples to illustrate the interest of local observations of malicious activities. This paper does not pretend to make a deep analysis of the phenomena we observe on the distributed platforms, but rather a clear demonstration of the usefulness of deploying multiple sensors in various places. As we have shown, they present very particular characteristics that enable learning more things than with current techniques. Similarities or differences they present confirm that it is worth analyzing local trends of the attacks. This information is of interest for the administrator who needs to clearly understand the threats against the network he is in charge of.

Most of the previous points must be carefully analyzed. This is part of our future work. Finally, we hope this paper will convince new partners to join our efforts in the collection of a diversified set of data. Also, it is our hope to see people bringing in their expertise to analyze rigorously our data sets thanks to techniques that we might not be familiar with.

# 7 References

[Auscert] AusCERT, Australian Computer Emergency Response Team home page: http://www.auscert.org.au.

[Caida] CAIDA, the Cooperative Association for Internet Data Analysis web site: http://www.caida.org

[Bak04] G. Bakos, R. Gray, "Analysis of the Data-Collection Capabilities of a Large-Scale, Distributed Honeypot System", 2004. Available on line at: http://www.ists.dartmouth.edu/projects/honeypots/.

[Braz04] The Brazilian Honeypots Alliance home page: http://www.honeypots-alliance.org.br/.

[Certcc] CERT Coordination Center home page: http://www.cert.org.

[Cook04] E. Cooke, M. Bailey, Z.M. Mao, D. Watson, F.Jahanian, D. McPherson, "Toward Understanding Distributed Blackhole Placement". *In Proc. of the 2004 ACM Workshop on rapid Malcode WORM'04*, Whashington DC, pp.54-64, 2004.

[Cym04] Team Cymru. "The Darknet Project". Available at http://www.cymru.com/Darknet/index.html, June 2004.

[DaPD04] M. Dacier, F. Pouget, H. Debar, "Attack Processes found on the Internet". *NATO Symposium IST-041/RSY-013*, Toulouse, France, April 2004.

[Disco] Disco Passive Fingerprinting Tool. home page: http://www.altmode.com/disco.

[DPDe04] M. Dacier, F. Pouget, H. Debar, "Honeypots, a Practical Mean to Validate Malicious FaultAssumptions". *In Proc. Of the 10th Pacific Ream Dependable Computing Conference (PRDC04)*, February 2004.

[Etter] Ettercap NG utility home page: http://ettercap.sourceforge.net.

[Honey] The Honeynet Alliance Project home page : http://www.honeynet.org.

[Luci04] The Lucidic Distributed Honeypot Project home page: http://www.lucidic.net.

[Moo01] D. Moore, G.M. Voelker, S. Savage. "Inferring Internet denial-of-service", *In Proc. of the 10th USENIX Security Symposium*, Whashington D.C., pp.9-22, Aug. 2001.

[Moo02] D. Moore. "Code-Red: a case study on the spread and victims of an internet worm". Available at: http://www.icir.org/vern/imw-2002/imw2002-papers/209.ps.gz, 2002.

[Morr04] C. Morrow, B. Gemberling. "How to Allow your Customers to blackhole their own traffic". 2004. Available at: http://www.secsup.org/CustomerBlackhole/

[Netfl] Cisco Systems. "Netflow Services and Applications", 1999.

[PoDa04] F. Pouget and M. Dacier. Honeypot-based forensics. *In Proc. of the AusCERT Asia Pacific Information Technology Security Conference 2004 (AusCERT2004)*, May 2004.

[P0f] p0f Passive Fingerprinting Tool, home page: http://lcamtuf.coredump.cx/p0f-beta.tgz.

[Pou04] F. Pouget. Leurre.com, the Eurecom Honeypot Project introduction. Available at: http://www.eurecom.fr/~pouget/leurrecom.htm.

[Prov04] N. Provos. Honeyd home page: http://www.honeyd.org.

[Sans04] SANS Institute. "Internet Storm Center ISC". Available at: http://isc.incidents.org

[Secfo04] Security Focus BugTraq Mailing List. Archives available at: http://www.securityfocus.com/archive/1, 2004.

[Song01] D. Song, R. Malan, R. Stone. "A snapshot of global Internet worm activity". Tech. Report, Arbor Networks, 2001.

[Spit03] L. Spitzner. "Honeypot Farms", Infocus, Aug. 2003. Available on line at: http://www.securityfocus.com/infocus/1720.

[Spit04] L. Spitzner. "The Honeynet Project". http://www.project.honeynet.org. 2004.

[Talis04] Talisker Security Wizardry Computer Security, "The Computer Network Defence Internet Operational Picture", available on line at http://securitywizardry.com/radar.htm, 2004.

[Yegn04] V. Yegneswaran, P. Barford, S. Jha. "Global intrusion detection in the DOMINO overlay system". *In Proc. of the Network and Distributed System Security Symposium NDSS'04*, San Diego, CA, Feb. 2004.