Behavioral Analysis of Zombie Armies

Olivier THONNARD [a,1] , Wim MEES [a] and Marc DACIER [b]
[a] *Royal Military Academy, Polytechnic Faculty, Brussels*
[b] *Symantec Research Labs, Sophia Antipolis, France*

Abstract. *Zombie armies* - or botnets, i.e., large groups of compromised machines controlled remotely by a same entity - pose today a significant threat to national security. Recent cyber-conficts have indeed demonstrated that botnets can be easily turned into digital weapons, which can be used by cybercriminals to attack the network resources of a country by performing simple Distributed Denial-of Service (DDoS) attacks against critical web services. A deep understanding of the long-term behavior of botnet armies, and their strategic evolution, is thus a vital requirement to combat effectively those latent threats. In this paper, we show how to enable such a long-term, strategic analysis, and how to study the dynamic behaviors and the global characteristics of these complex, large-scale phenomena by applying different techniques from the area of knowledge discovery on attack traces collected on the Internet. We illustrate our method with some experimental results obtained from a set of worldwide distributed server honeypots, which have monitored attack activity in 18 different IP subnets for more than 640 days. Our preliminary results highlight several interesting findings, such as *i)* the strong resilience of zombie armies on the Internet, with survival times going up to several months; *ii)* the high degree of coordination among zombies; *iii)* the highly uneven spatial distribution of bots in a limited number of "unclean networks", and *iv)* the large proportion of home users' machines with high-speed Internet connexions among the bot population.

Keywords. Intelligence monitoring, Threat analysis, Zombie armies.

Introduction

In the recent years, many security experts have drawn attention to the increasingly important security problem related to *zombie armies* - also called botnets, which are groups of malware-infected machines that are remotely controlled and coordinated by a same entity. Still today, zombie armies and botnets constitute, admittedly, one of the main threats on the Internet, as they are used for different kinds of illegal activities (e.g., bulk spam sending, online fraud, denial of service attack, etc) [2,19]. More importantly, the analysis of recent "cyber conflicts", such as the presumed cases related to Estonia and Georgia [17,6,7], have lead experts to the conclusion that botnets can be easily turned into digital weapons, which can be used by cybercriminals (or dissidents) to attack the network resources of a country by performing very simple Distributed Denial-of Service (DDoS) attacks against critical web services (e.g., DNS servers, network routers, gov-

[1]Corresponding Author: Olivier Thonnard, Royal Military Academy, Avenue de la Renaissance 30, 1000 Brussels, Belgium; E-mail: olivier.thonnard@rma.ac.be.

ernment or financial websites, etc), which can lead to substantial economical or financial loss. Although no clear evidence of the implication of any governmental organization in those attacks could be underlined, one important lesson learned from these events is that botnets are primarily used by dissidents or activists to perform this type of attacks in periods of political disturbances. A deep understanding of the long-term behavior of botnet armies, and their evolution, is thus a vital requirement to be able to combat effectively those latent threats.

While most previous studies related to botnets have focused on understanding their inner working [24,5,1], or on techniques for detecting individual bots at the network-level [8,9], in this work we are more interested in studying the global behaviors of those armies from a strategic viewpoint. That is, we are not interested in studying a particular botnet from the inside, or in the analysis of the various protocols used by bots to communicate with their C&C server. But instead, we want to perform a **long-term**, **strategic analysis** of those armies from a behavioral point of view, i.e.: how long do they stay alive on the Internet, what is their average size and their spatial distribution, and more importantly, how do they evolve over time with respect to different criteria such as their origins, or the type of activities (or scanning) they perform.

The first contribution of this paper consists in introducing a systematic method that enables us to perform such a strategic analysis of zombie armies, based on the botnet scanning traffic observed in a global honeynet. Our approach is based on an appropriate combination of different knowledge discovery and data mining techniques, which consists of the following components:

1. detection and characterization of coordinated attack events;
2. unsupervised clique-based clustering, so as to discover correlations among attack events;
3. dimensionality reduction techniques, which allow us to visualize and to assess the cliques correlations;
4. a fuzzy, multi-criteria decision-making process that leverages the results obtained in the previous steps, in order to identify sequences of attack events that are very likely attributed to the same zombie army.

As second contribution, we present some preliminary results obtained from a proof-of-concept framework in which we implemented the techniques mentioned here above. The experiments have been performed on attack traces collected with a worldwide distributed honeynet, which has observed global attack activity in over 18 different IP subnets from Sep 2006 until July 2008 (i.e., about 640 days). Our experimental results highlight several interesting facets of the botnet phenomenon:

- with a mean lifetime of about 98 days, zombie armies seem to be quite resilient. In some extreme cases, we observed certain armies surviving for more than 18 months, which indicates that *taking down botnets still constitutes a real challenge*. On average, zombie armies had at least 8,500 distinct, observable sources during their lifetime.
- regarding the origins, malicious sources involved in zombie armies seem to be highly unevenly distributed in the IPv4 address space; they clearly form a relatively small number of tight clusters within a number of "unclean networks", which are thus responsible for a large deal of malicious activities related to server-side attacks (e.g., network scanning, bot propagation).

- over all zombie armies observed so far, at least 43% of the botnet population is made of home users' machines with high-speed Internet connexions (cable, DSL). Windows 2000 and WinXP Pro were the primarily operating systems among zombie machines (i.e., more than 90% of the bots).
- similarly to real-world armies, certain groups of zombie machines seem to be able to coordinate their efforts, e.g., by coordinating different tasks such as network reconnaissance and subsequent targeted attacks.
- finally, most of the identified zombie armies had a significant attack capability, not only in terms of the available bandwidth that can possibly be offered by all zombies together, but also the number of ports they are able to probe or to exploit.

The rest of the paper is structured as follows: in Section 1, we give a brief overview of the honeynet used in our experiments, and we define the notion of coordinated *attack events* as observed by the honeypots. In Section 2, we describe the components of our knowledge discovery framework that we use to identify global attack phenomena, whereof most are related to some activities of zombie armies. In Section 3, we present our experimental results and the kind of findings we can obtain by applying this method to a set of attack events collected on the Internet. Finally, we conclude in Section 4.

Note that this research builds on prior work in malicious traffic analysis. More particularly, we have presented in [28] a more formal and complete discussion of our framework, especially regarding the aspect fuzzy, multi-criteria decision-making. To make this paper as self-contained as possible, we have summarized as much as possible our previous contributions in Section 2. This paper will mostly focus on the practical results obtained in each step of our analysis framework, rather than the formal aspects of the different techniques.

1. Collecting Attack Traces with a Global Honeynet

1.1. Leurre.com Honeynet - Dataset Overview

Our data set is made of network attack traces collected with a distributed set of sensors (called *server honeypots*), which are deployed in the context of the *Leurre.com Project* [14,22]. Because honeypots are systems deployed for the sole purpose of being probed or compromised, any network connection that they establish with a remote IP can be considered as malicious, or at least suspicious.
Launched in 2003 by Eurecom, a research Institute based in Sophia Antipolis (France), this project maintains a worldwide distributed system of honeypots running in more than 30 different countries covering the five continents. The main objective of the project is to get a realistic picture of certain classes of global attack phenomena happening on the Internet, by collecting unbiased quantitative data in a long-term perspective. In the first phase of the project, the data collection infrastructure relied solely on low-interaction sensors based on *Honeyd* [23] to collect unsolicited traffic (also sometimes termed "Internet background radiation" [18]). In early 2008, a second phase of the project was started with the deployment of medium-interaction honeypots based on the *ScriptGen* [15] technology, in order to enrich the network conversations with the attackers. Scriptgen sensors are able to automatically learn about new protocol interactions, such that they can handle *0-day* exploits, and eventually capture shellcode samples and malware

Table 1. Overview of some prevalent types of activities observed in the honeynet, grouped by port sequence. The network traffic has been collected from Sep'06 until June'08.

Observed Port Sequence	Targeted Service	Volume of Sources (%)	Main Origins (countries)
|I	ICMP (Echo request/reply)	755,227 (28%)	US(20%),KR(11%),CN(10%),BR(6%), others(53%)
|1026U|1027U|1028U	Windows Messenger	373,361 (14%)	CA(100%)
|1026U	Windows Messenger	216,040 (8%)	US(50%),null(17%),CA(6%), others(27%)
|445T	Microsoft-DS	208,060 (8%)	CS(32%),RS(19%),US(6%), others(43%)
|I|139T, |I|139T|445T	ICMP (Allaple), MS-Netbios-ssn, Microsoft-DS	130,392 (5%)	KR(20%), others(80%)
|135T	Microsoft DCE/RPC	112,764 (4%)	JP(16%),US(13%),CS(7%),RS(7%), PL(6%),DE(6%), others(45%)
|5900T	VNC	104,238 (4%)	US(17%),CN(6%),FR(6%),KR(6%), others(51%)
|2967T	Symantec AntiVirus (ssc-agent)	101,062 (4%)	US(23%),CN(8%),JP(6%), DE(5%),PK(5%), others(53%)
|1433T	MS-SQL	87,332 (3%)	CN(32%),US(15%),others(53%)
|139T	MS-Netbios-ssn	50,781 (2%)	US(17%),CA(8%),TW(5%),FR(5%), others(65%)
|I|80T	ICMP, Web	48,649 (2%)	US(54%),KR(11%),CN(8%), CA(7%), others(20%)
|1434U	MS-SQL-Monitor (Slammer)	36,627 (1%)	CN(44%),US(14%),JP(6%), others(36%)
|22T	SSH	36,094 (1%)	CN(24%),US(13%),KR(8%), TW(5%), others(50%)
|80T	Web	28,005 (1%)	US(27%),CN(7%),FR(7%), DE(7%),null(5%), others(47%)
|137U	MS-Netbios-ns	25,630 (<1%)	US(16%),BR(9%),AR(6%), FR(5%),ES(5%), others(59%)
|I|445T	ICMP, Microsoft-DS	18,273 (<1%)	US(14%),CN(13%),TW(8%),FR(7%), JP(7%),null(6%),DE(5%), others(41%)
|4899T	Remote Admin	15,935 (<1%)	CN(15%),US(15%),KR(10%), RU(5%), others(54%)

binaries when they are targeted by code injection attacks. All network traces captured on the platforms are automatically uploaded into a centralized database. The collected traffic is also enriched with a diverse set of contextual information, such as: the geographical location and the ISP's of malicious sources (via Maxmind), reverse DNS lookups, VirusTotal[2] and Anubis[3] reports for each sample of downloaded malware, passive OS fingerprinting (with P0f), Snort IDS alerts, and more recently, we also added the correlation of the observed IP sources with different IP blacklisting services (e.g., Spamhaus[4], Emergingthreats[5] blocking lists, and a fast-flux bot tracker[6]).

For the purpose of this study, we have used a 640-day attack trace collected by 36 platforms located in 20 different countries and belonging to 18 different class A-subnets. Note that, in the scope of this paper, we only considered the traffic collected by low-interaction sensors; but we are actively looking into extending our analysis techniques to integrate the attack traffic gathered by the medium-interaction (ScriptGen) platforms. Table 1 gives an overview of the most prevalent types of activities grouped by targeted port sequences, and their origins, as observed in the honeynet.

From this traffic, we have then selected only the most prevalent types of activities observed on the sensors, i.e., about 130 distinct attack profiles for which an activity

[2]http://www.virustotal.com
[3]http://anubis.iseclab.org
[4]http://www.spamhaus.org
[5]http://www.emergingthreats.net
[6]http://dnsbl.abuse.ch

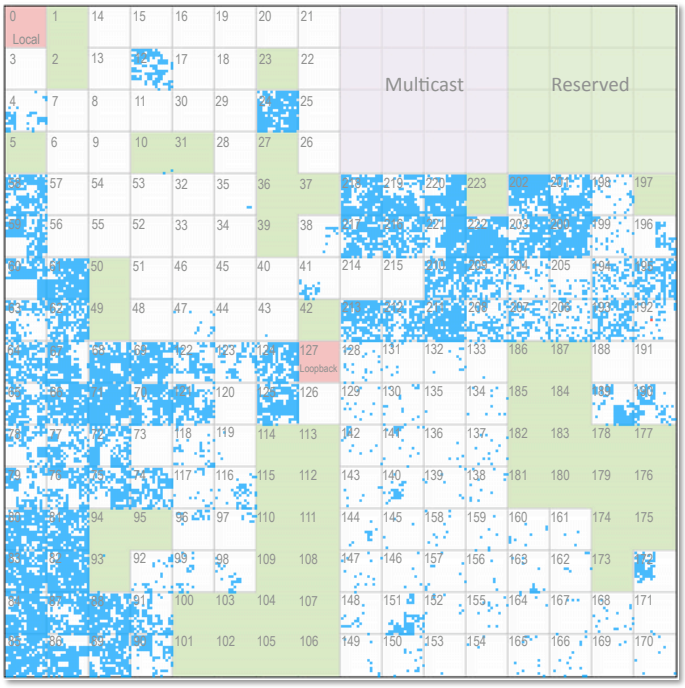

Figure 1. Distribution of malicious sources in the IPv4 address space using a fractal mapping (Hilbert curve).

involving a sufficient number of IP sources had been observed at least once on a given day. This data set comprises totally 1,195,254 distinct sources, which have sent about 3,423,577 packets to the sensors. Fig. 1 illustrates the distribution of malicious sources for these activities using a fractal mapping (e.g., a Hilbert curve). Note that spoofed IP addresses have already been filtered from this data set. As such, Fig. 1 and Table 1 give already some interesting viewpoints, as it clearly shows that most malicious sources seem to be clustered in a limited number of IP blocks (or AS'es). Nevertheless, this type of global analysis does not help us to get insights into the individual attack phenomena that occurred at a large scale (such as zombie armies). Moreover, such global trends do not allow us to learn about the *modus operandi* of the attackers, which is why we need to develop a more detailed analysis.

1.2. Coordinated Attack Events

We use a classical clustering algorithm to perform a first low-level classification of the raw network traffic. Hence, each IP source observed on a honeypot sensor is attributed to a so-called *attack cluster* [21] according to its network characteristics, such as the number of IP addresses targeted on the sensor, the number of packets and bytes sent to each IP, the attack duration, the average inter-arrival time between packets, the associated port sequence being probed (e.g., if a source sends first some ICMP packets followed by an exploit on port 445/TCP, then it is associated to the port sequence $\langle$I-445T$\rangle$), and the packet payload. Therefore, all IP sources belonging to a given attack cluster have left

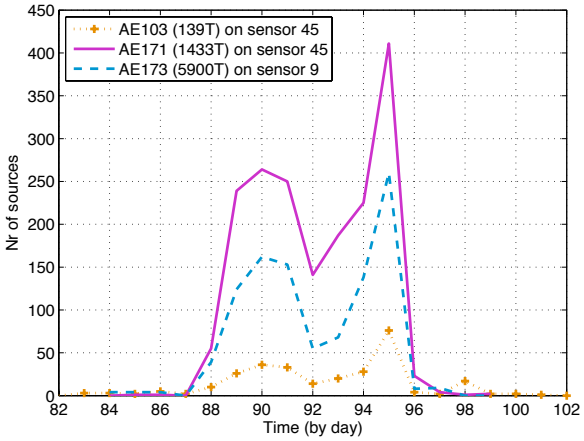

Figure 2. Illustration of 3 attack events observed on 2 different sensors, and targeting 3 different ports.

very similar network traces on a given sensor and consequently, they can be considered as having the same *attack profile*. This leads us then to the concept of attack event, which is defined as follows:

> An *attack event* refers to a subset of IP sources having the same attack profile on a given sensor, and whose coordinated activity has been observed within a specific time window.

Fig. 2 illustrates this notion by representing the time series (i.e., the number of sources per day) of three coordinated attack events observed on two different sensors in the same time interval, and targeting three different ports. The identification of those events can be easily automated by using the method presented in [20]. By doing so, we are able to extract interesting events from the spurious, nonproductive traffic collected by our sensors, and we can focus on the most important events that might originate from coordinated phenomena, such as attack activities resulting from botnet reconnaissance scans, and bot propagation. As previous botnet studies have already showed [13], it seems that the botnet scanning behavior is ingrained to the botnets because this is an effective (and low-cost) way for them to recruit new bots. Therefore, botmasters will probably not give up scanning in the near future.

By using the technique described in [20], we have extracted from the whole data set about 351 attack events that were coordinated on at least two different sensors. In the rest of this paper, we will focus on the analysis of this set of attack events, which still accounts for 282,363 unique sources (23.6 % of the original data set), or 741,349 packets (21.5%), and we will show how to take advantage of different external attack characteristics to discover knowledge, and to identify individual phenomena related to zombie armies.

2. A Framework to Identify Global Attack Phenomena

2.1. Overview

Once we have identified a set of attack events occurring at different moments, how could we know in a reliable way which events can be attributed to the same root phenomenon? That is, how can we identify which *sequences of attack events* are very likely the consequence of the same zombie army scanning or probing one or several subnets, eventually during non-contiguous intervals of time?

In the realm of threat monitoring, this problem is sometimes referred to as "attack attribution", which is the process of effectively attributing new attack events to (un)-known phenomena, based on some evidence or traces left on one or several monitoring platforms. To address this problem in a systematic way, we have developed a framework that analyzes attack events with appropriate knowledge discovery (KDD) techniques. The main components of this framework are sketched in Fig. 3. Based on a set of attack events (as defined here above), the first KDD component extracts cliques of attackers in an unsupervised way, so as to identify meaningful correlations between events. That is, we want to know whether some groups of events are strongly correlated with respect to some given characteristics. For example, we could discover which groups of attack events share the very same spatial distributions (in terms of geographical or IP subnet distributions), or which other groups of attack events are targeting the same set of sensors in the same time interval, or which groups of attacks are similar in terms of activities (e.g., the port sequences targeted by malicious sources), and so on. We motivate our choice of attack characteristics used to discover knowledge in the next subsection. Then, we evaluate the consistency of the extracted cliques (or clusters) by using dimensionality reduction techniques, which enable us to visualize on a map the cliques results for each attack dimension. We refer to this step as "semantic mapping", since the distance between each pair of events on a given mapping has a certain meaning. Indeed, the distances are related to the degree of similarity between the underlying feature vectors of the attack events (i.e., the distributions of countries, subnets, etc).

In the next component of the framework, we have implemented a multi-criteria decision-making algorithm that is based on fuzzy inference systems (FIS). The objective consists in combining intelligently the previously extracted knowledge (i.e., the cliques and the semantic mappings), so as to build sequences of attack events that can be attributed to the same global phenomena with a high degree of confidence, thanks to the combination of different statistical measurements. Interestingly, a FIS does not need any training prior making inferences. Instead, it takes only advantage of the previously extracted knowledge to make sound inferences, so as to attribute incoming attack events to a given phenomenon. Each identified attack phenomenon is then modeled with a fuzzy inference system.

2.2. Defining Attack Characteristics

In most knowledge discovery applications, we must first define salient features that may provide some meaningful *patterns* [11]. So, we start by defining different attack char-

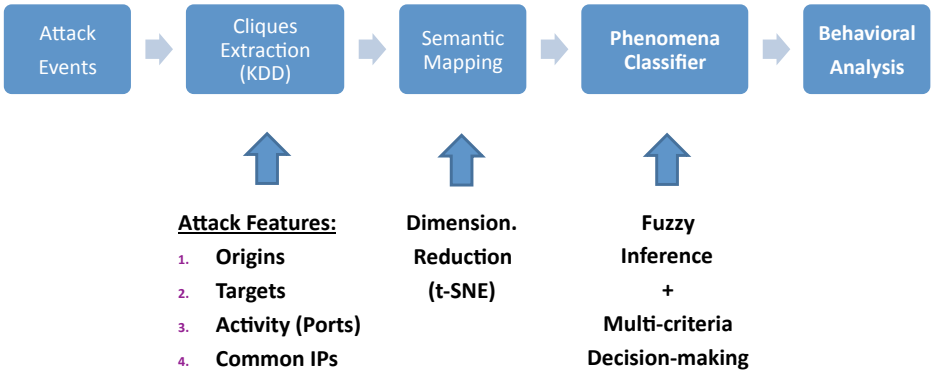

Figure 3. Components of a Knowledge Discovery Framework for Identifying Global Phenomena.

acteristics that we have used to extract knowledge from our set of attack events. In this specific case, we consider them as useful to analyze the root causes of global phenomena observed on our sensors, and as a result, to identify different zombie armies. However, we do not pretend they are the only ones that could be used in threat monitoring. Since other characteristics might prove relevant in the future, our framework is built such that additional features could be easily included when necessary (e.g., to include characteristics related to code injection attacks, shellcodes, or malware samples).

The two first characteristics retained are related to the *origins* of the attackers, i.e. their spatial distributions. First, the geographical distribution of malicious sources can be used to identify botnets that are located in a limited number of countries. Similarly, the IP network blocks provide also an interesting viewpoint on the attack phenomena, since it gives a good indication of the spatial "uncleanliness" of certain networks, i.e., the tendency for compromised hosts (e.g., zombie machines) to stay clustered within unclean networks [4]. So, for each attack event, we can create a feature vector representing either the distribution of originating countries, or of IP addresses grouped by Class A-subnet (i.e., by /8 prefix).

The next characteristic deals with the *targets* of the attackers, namely the distribution of sensors that have been targeted by the sources. Botmasters may indeed send commands at a given time to all zombies to instruct them to start scanning (or attacking) one or several IP subnets, which of course will create coordinated attack events on specific sensors. Therefore, it seems important to look at relationships that may exist between attack events and the sensors they have been observed on.

Besides the origins and the targets, the type of activity performed by the attackers seems also relevant to us. In fact, bot software is often crafted with a certain number of available exploits targeting a reduced set of TCP or UDP ports. In other words, we might think of each botnet having its own *attack capability*, which means that a botmaster will normally issue scan or attack commands only for vulnerabilities that he might exploit to expand his botnet. So, it seems to make sense to take advantage of this feature to look for similarities between the sequences of ports that have been targeted by the sources of the attack events.

Table 2. Some experimental clique results obtained from a honeynet dataset collected from Sep 06 until June 08. (1) the given patterns represent the average distributions for the most prevalent cliques, i.e. the ones lying in the upper quartile in terms of number of sources. For the IP subnets (resp. targeted platforms), the numbers refer to the distributions of originating (resp. targeted) class A-subnets.

Attack Dimension	Nr of Cliques	Max.size (nr events)	Min.size (nr events)	Volume of sources (%)	Most prevalent patterns found in the cliques(1)
Geolocation	31	40	3	84.4	⟨CN,CA,US,FR,TW⟩, ⟨IT,ES,FR,SE,DE,IL⟩, ⟨KR,US,BR,PL,CN,CA⟩ ⟨US,JP,GB,DE,CA,FR,CN,KR⟩, ⟨US,FR,JP,CN,DE,ES,TW⟩, ⟨CA,CN⟩ ⟨PL,DE,ES,HU,FR⟩
IP Subnets (Class A)	25	51	3	91.2	⟨87,82,151,83,84,81,85,213⟩, ⟨222,221,60,218,58,24,124,121,219,82,220⟩ ⟨201,83,200,24,211,218,89,124,61,82,84⟩, ⟨24,60⟩ ⟨83,84,85,80,88⟩, ⟨193,195,201,202,203,216,200,61,24,84,59⟩
Targeted platforms	17	86	2	70.1	⟨202⟩, ⟨88, 192⟩, ⟨195⟩, ⟨193⟩, ⟨194⟩ ⟨129, 134, 139, 150⟩, ⟨24, 213⟩
Port sequences	22	66	4	93.2	⟨I⟩, ⟨1433T⟩, ⟨I-445T⟩, ⟨5900T⟩, ⟨1026U⟩, ⟨135T⟩, ⟨50286T⟩ ⟨I-445T-139T-445T-139T-445T⟩, ⟨6769T⟩, ⟨1028U-1027U-1026U⟩

Finally, we have also decided to compute, for each pair of events, the ratio of common IP addresses. We are aware of the fact that, as time passes, some zombie machines of a given botnet might be cured while others may get infected and join the botnet. Additionally, certain ISPs apply a quite dynamic policy of IP address allocation to residential users, which means that bot-infected machines can have different IP addresses when we observe them at different moments (i.e., DHCP churn effect). Nevertheless, and according to our domain experience, it is reasonable to expect that if two distinct attack events have a high percentage of IP addresses in common, then the probability that those two events are somehow related to the same global phenomenon is increased (assuming that the time difference between the two events is not too large).

2.3. Clique-based Knowledge Discovery

For each attack characteristic considered here above, we have applied a clique-based clustering on our set of attack events. That is, we use a graph-based approach to formulate the problem: the vertices of the graph represent the feature vectors of each attack event (e.g., the distribution of countries, subnets, targeted sensors, etc), and the edges express the similarity relationships between those vertices. Clearly, the choice of a similarity metric is very important, as it has an impact on the properties of the final clusters, such as their size, quality, and consistency. To reliably compare the kind of empirical distributions mentioned here above, we have chosen to rely on strong statistical distances, such as Pearson's χ^2, or the Jensen-Shannon divergence (JSD) [16], which derives itself from the Kullback-Leibler divergence [12]. Finally, the clustering is performed by extracting so-called *maximal weighted cliques* (MWC) from the graph, where a maximal *clique* is defined as an induced sub-graph in which the vertices are fully connected and it is not contained within any other clique. We refer the interested reader to [27,26] for a more detailed description of this clique-based clustering technique applied to honeynet traces.

Table 2 presents a high-level overview of the cliques obtained for each attack dimension separately. As we can see, a relatively high volume of sources could be classified into cliques for each dimension. The last colon with the most prevalent patterns gives an indication of which countries or class A-subnets (e.g., originating or targeted IP

subnets) are most commonly observed in the cliques that lie in the upper quartile with respect to the number of sources. Interestingly, it seems that many coordinated attack events are coming from a given IP sub-space. Regarding the targeted platforms, several cliques involve a single class A-subnet. About the type of activities, we can observe some commonly targeted ports (e.g., Windows ports used for SMB or RPC, or SQL and VNC ports), but also a large number of uncommon high TCP ports that are normally unused on standard (and clean) machines (such as 6769T, 50286T, 9661T, ...). A non-negligeable volume of sources is also due to UDP spammers targeting Windows Messenger popup service (ports 1026 to 1028/UDP).

2.4. Visualizing Cliques - Knowledge Consolidation

In order to assess the consistency of the resulting cliques of attack events, it can be useful to see them charted on a two-dimensional map so as to *i)* verify the proximities among clique members (*intra-clique* consistency), and *ii)* understand potential relationships between *different* cliques that are somehow related (i.e. *inter-clique* relationships). Moreover, the statistical distances used to compute those cliques make them intrinsically coherent, which means also that certain cliques of events may be somehow related to each other, although they were separated by the clique algorithm.

Since most of the feature vectors we are dealing with have a high number of variables (e.g., a geographical vector has more than 200 country variables), the structure of such high-dimensional data set cannot be displayed directly on a 2D map. Multidimensional scaling (MDS) is a set of methods that can help to address this problem. MDS is based on dimensionality reduction techniques, which aim at converting a high-dimensional dataset into a two or three-dimensional representation that can be displayed, for example, in a scatter plot. The aim of dimensionality reduction is to preserve as much of the significant structure of the high-dimensional data as possible in the low-dimensional map. As a consequence, MDS allows an analyst to visualize how far observations are from each other for different kinds of similarity measures, which in turn can deliver insights into the underlying structure of the high-dimensional dataset.

Because of the intrinsic non-linearity of real-world data sets, we have applied a recent MDS technique called *t-SNE* to visualize each dimension of the data set, and to assess the consistency of the cliques results. t-SNE [29] is a variation of *Stochastic Neighbour Embedding*; it produces significantly better visualizations than other MDS techniques by reducing the tendency to crowd points together in the centre of the map. Moreover, this technique has proven to perform better in retaining both the local and global structure of real, high-dimensional datasets in a single map, in comparison to other non-linear dimensionality reduction techniques such as Sammon mapping, Isomaps or Laplacian Eigenmaps [10].

Figure 4 shows the resulting two-dimensional plot obtained by mapping the geographical vectors on a 2D map using t-SNE. Each datapoint on this map represents the geographical distribution of a given attack event. The coloring refers to the clique membership of each event, and the dotted circles indicate the clique sizes. We could easily verify that two adjacent events on the map have highly similar geographical distributions (even from a statistical viewpoint), while two distant events have clearly nothing in common in terms of originating countries. Quite surprisingly, the resulting mapping is far from being "chaotic"; it presents a relatively sparse structure with clear datapoint group-

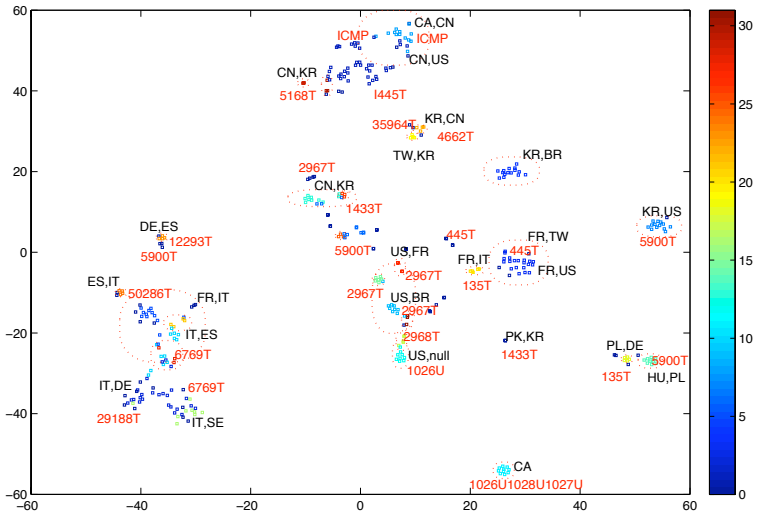

Figure 4. Visualization of geographical cliques of attackers. The coloring refers to the different cliques and the dotted circles indicate their sizes on the low-D map. The superposed text labels indicate the two first attacking countries of the distribution of certain attack events, as well as some of the targeted port sequences (in red).

ings, which means also that most of those attack events present very tight relationships regarding their origins. Due to the strict statistical distances used to calculate cliques, this kind of correlation can hardly be obtained by chance only.

Similar "semantic mapping" can naturally be obtained for the other dimensions (e.g., subnets, platforms, etc), so as to help assessing the quality of other cliques of attackers. As described in the next Section, those different mappings will be used by the multi-criteria decision-making component of our framework to identify global phenomena, i.e. by combining efficiently different sets of cliques.

2.5. Identification of Zombie Armies using Fuzzy Inferences

The final objective consists in re-constructing *sequences of attack events* that can be attributed with a high confidence to the same root phenomenon in function of multiple criteria. In other words, we want to build an inference engine that takes as input the extracted knowledge (cliques and mappings) to classify incoming attack events into either "known phenomena", or otherwise to identify a new phenomenon when needed (e.g., when we observe a new zombie army). To do this, we have implemented a multi-criteria decision-making algorithm that relies on fuzzy inferences. Our motivation is that: i) we have *a priori* zero-knowledge of the expected output, which means that we can not provide training samples showing the characteristics of the output we are looking for; and ii) we want to include some domain knowledge to specify which type of combinations we expect to be promising in the root cause identification. Also, we favor the "white-box" approach (or a transparent reasoning process), which allows an expert to understand why the system has grouped a given set of events into the same root phenomenon.

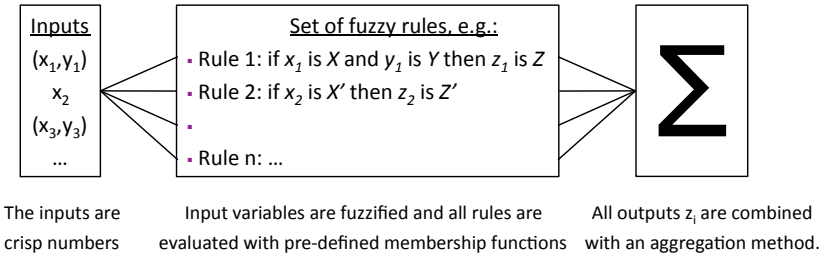

Figure 5. Main components of a Fuzzy System.

Although large-scale phenomena on the Internet are complex and dynamic, our intuition is that two **consecutive** attack events should be linked to the same root phenomenon if and only if they share at least two different attack characteristics. That is, our decision-making process will attribute two attack events to the same phenomenon when the events characteristics are "close enough" (from a statistical viewpoint) for any combination of **at least two** attack dimensions out of the complete set of criteria: $\{origins, targets, activity, common_{IP}\}$. In other words, we hypothesize that real-world phenomena may perfectly evolve over time, which means that two consecutive attack events of the same zombie army must not necessarily have all their attributes in common. For example, the bots' composition of a zombie army may evolve over time because of the cleaning of infected machines and the recruitment of new bots. From our observation viewpoint, this will translate into a certain shift in the IP subnet distribution of the zombie machines for subsequent attack events of this army (and thus, most probably different cliques w.r.t. the origins). Or, a zombie army may be instructed to scan several consecutive IP subnets in a rather short interval of time, which will lead to the observation of different events having highly similar distributions of originating countries and subnets, but those events will target completely different sensors, and may eventually use different exploits (hence, targeting different port sequences).

On the other hand, we consider that only one correlated attack dimension is not sufficient to link two attack events to the same root cause, since the result might then be due to chance only (e.g., a large proportion of attacks originate from some large or popular countries, certain Windows ports are commonly targeted, etc). However, by combining intelligently several attack viewpoints, we can reduce considerably the probability that two attack events would be attributed to the same root cause whereas they are in fact unrelated.

We still need to formally define what is the "relatedness degree" between two attack events, certainly when they do not belong to a same clique but are somehow "close" to each other. Intuitively, attack events characteristics in the real world have unsharp boundaries, and the membership to a given phenomenon can be a matter of degree. For this reason, we have developed a decision-making process that is based on a fuzzy inference system (FIS). Fuzzy Inference is a convenient way to map an input space to an output space with a flexible and extensible system, and using the codification of common sense and expert knowledge. The mapping then provides a basis from which decisions can be made. The main components of an inference system are sketched in Fig. 5. To map the input space to the output space, the primary mechanism is a list of if-then statements called rules, which are evaluated in parallel, so the order of the rules is unimportant.

Instead of using crisp variables, all inputs are *fuzzified* using membership functions in order to determine the degree to which the input variables belong to each of the appropriate fuzzy sets. If the antecedent of a given rule has more than one part (i.e., multiple 'if' statements), a fuzzy logical operator is applied to obtain one number that represents the result of the antecedent for that rule.

Concretely, we use the knowledge obtained from the extraction of cliques to build the fuzzy rules that describe the behavior of a given phenomenon. The characteristics of new incoming attack events are then used as input to the fuzzy systems that model the phenomena identified so far. In each of those fuzzy systems, the features of the *most recent* attack event shall define the current parameters of the membership function used to evaluate the following simple rules: if x_i is $close$ AND if y_i is $close$ then z_i is $related$, $\forall i \in \{geo, subnets, targets, portsequence\}$. The membership functions referred to as "is close" in the fuzzy rules are thus defined by the characteristics of the cliques to which the attack events belong. The calculation of the rule output $z_i \in [0, 1]$ is just the intersection between two curves, which quantifies the inter-relationship between the cliques (and hence, between the attack events).

The results of all rules are then combined and distilled into a single, crisp value using an appropriate multi-criteria aggregation function. In this case, we use an Ordered Weighted Average (OWA) operator, which allows to model more complex requirements such as "most of", or "at least two" criteria to be satisfied in the overall decision function [30]. We refer the interested reader to [28] for a more detailed discussion of our multi-criteria decision-making algorithm.

3. Behavioral Analysis of Zombie Armies

3.1. Global Characteristics

In this Section, we provide some experimental results obtained by applying our multi-criteria inference method to our set of attack events introduced in Section 2 (clique analysis). Over the whole collection period (640 days), we found only 32 global phenomena. In total, 348 attack events (99%) could be attributed to a large-scale phenomenon. An in-depth analysis has revealed that most of those phenomena (apart from the noisy network worm W32.Rahack.H [25], also known as W32/Allaple) are quite likely related to *zombie armies*, i.e. groups of compromised machines belonging to the same botnet(s). We conjecture this for the following main reasons: *i)* the apparent coordination of the sources, both in time (i.e., coordinated events on several sensors) and in the distribution of tasks (e.g., scanners versus attackers); *ii)* the short durations of the attack events, typically a few days only, whereas "classical" worms tend to spread over longer, continuous periods of time; *iii)* the absence of known classical network worm spreading on many of the observed port sequences; and *iv)* the source growing rate, which has a sort of exponential shape for worms and is somehow different for botnets [13].

To illustrate the results, Table 3 presents an overview of some global phenomena found in our dataset. Thanks to our method, we are able to characterize precisely the behaviors of the identified phenomena or zombie armies. Hence, we found that the largest army had in total 57 attack events comprising 69,884 sources, and could survive for about 112 days. The longest lifetime of a zombie army observed so far was still 586

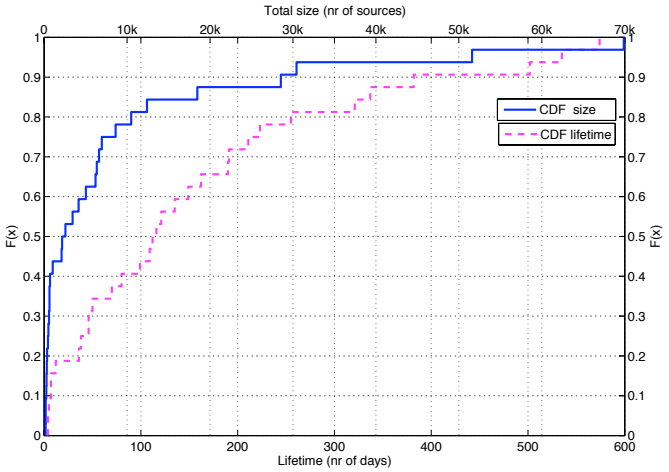

Figure 6. Empirical CDF of the size and lifetime of zombie armies.

days. Fig. 6 shows the cumulative distributions (CDF) of the lifetime and size of the identified armies. Those figures reveal some interesting aspects of their global behaviors: according to our observations, at least 20% of the zombie armies had in total more than ten thousand observable[7] sources during their lifetime, and the same proportion of armies could survive on the Internet for at least 250 days. On average, zombie armies have a total size of about 8,500 observed sources, a mean number of 658 sources per event, and their mean survival time is 98 days.

Regarding the origins, we observe some very persistent groups of IP subnets and countries of origin across many different armies. On Fig. 7, we can see the CDF of the sources involved in the zombie armies of Table 3, where the x-axis represents the first byte of the IPv4 address space. It appears clearly that malicious sources involved in those phenomena are highly unevenly distributed and form a relatively small number of tight clusters, which account for a significant number of sources and are thus responsible for a large deal of the observed malicious activities. This is consistent with other prior work on monitoring global malicious activities, in particular with previous studies related to measurements of Internet background radiation [3,18,31]. However, we are now able to show that there are still some notable differences in the spatial distributions of those zombie armies with respect to the average distribution over all sources (represented with the blue dashed line). In other words, certain armies of compromised machines can have very different spatial distributions, even though there is a large overlap between "zombie-friendly" IP subnets. Moreover, because of the dynamics of this kind of phenomena, we can even observe very different spatial distributions within a *same army* at different moments of its lifetime. This is a strong advantage of our analysis method that is more precise and enables us to distinguish *individual* phenomena, instead of global trends, and to follow their dynamic behavior over time.

[7] It is important to note that the sizes of the zombie armies given here only reflect the number of sources we could *observe* on our sensors; the actual sizes of those armies are most probably much larger.

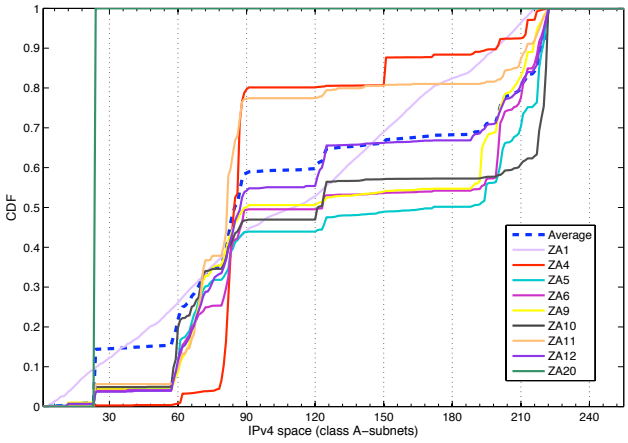

Figure 7. Empirical CDF of sources in IPv4 address space for the 9 zombie armies illustrated in Table 3.

Another interesting observation on Fig. 7 is related to the subnet CDF of ZA1 (uniformly distributed in the IPv4 space, which means randomly chosen source addresses) and ZA20 (a constant distribution coming exclusively from the subnet 24.0.0.0/8). A very likely explanation is that those zombie armies have used spoofed addresses to send UDP spam messages to the Windows Messenger service. So, this indicates that IP spoofing is still possible under the current state of filtering policies implemented by certain ISP's on the Internet.

Then, in terms of *attack capability*, we observe that about 50% of the armies could target at least two completely different ports (thus, probably two different exploits, at least), and one army had even an attack capability greater than 10. Table 4 provides additional details on the characteristics of malicious sources involved in those zombie armies. Regarding the operating systems (detected through passive OS fingerprinting with P0f), we can see that a large majority of the sources are running either Windows 2000 SP or Windows XP Pro. Finally, by analyzing the hostnames of the sources (obtained via reverse DNS lookups), we infer the ratio of home users's machines by looking for typical strings such as '%DSL%', '%PPP%', '%CABLE%'. Over all zombie armies observed so far, we found that at least 43% of the botnet population is made of residential users with high-speed Internet connections. If we take 256kbps as a lower-bound estimate of the average upstream bandwidth for this kind of connection, then we observe that most of those zombie armies could have an aggregate network capacity of several gigabits per seconds, which can easily be used to exhaust almost any type of network resources on the Internet by launching Distributed Denial of Service attacks.

3.2. Some Detailed Examples

In this Section, we further detail two zombie armies to illustrate some typical behaviors we could observe among the identified phenomena, e.g.:

i) a move (or drift) in the origins of certain armies (both geographical and IP blocks) during their lifetime;

Table 3. Overview of some large-scale phenomena found in a honeynet dataset (Sep'06 until Jun'08.

Id	Nr of events	Total size (nr sources)	Lifetime (nr days)	Targeted sensors (Class A- subnets)	Attack capability	Main origins (countries / subnets)
1	10	18,468	535	24.*,193.*,195.*,213.*	1026U	US,JP,GB,DE,CA,FR,CN,KR,NL,IT 69,128,195,60,81,214,211,132,87,63
4	82	26,962	321	202.*	12293T,15264T,18462T,25083T, 25618T,28238T,29188T, 32878T,33018T,38009T,4152T, 46030T,4662T,50286T,...	IT,ES,DE,FR,IL,SE,PL 87,82,83,84,151,85,81,88,80
5	13	9,644	131	195.*	135T,139T,1433T,2968T,5900T	CN,US,PL,IN,KR,JP,FR,MX,CA 218,61,222,83,195,221,202,24,219
6	15	51,598	>1 year	> 7 subnets	ICMP (W32.Rahack.H / Allaple)	KR,US,BR,PL,CN,CA,FR,MX,TW 201,83,200,24,211,218,89,124
9	23	11,198	218	192.*,193.*,194.*	2967T,2968T,5900T	US,CN,TW,FR,DE,CA,BR,IT,RU 193,200,24,71,70,213,216,66
10	57	69,884	112	128.*,129.*,134.*,139.*,150.*	I-I445T	CN,CA,US,FR,TW,IT,JP,DE 222,221,60,218,58,24,70,124
11	14	2,636	110	129.*,134.*,139.*,150.*	I-445T-139T-445T-139T-445T	US,FR,CA,TW,IT 82,71,24,70,68,88,87
12	14	27,442	183	192.*,193.*,194.*,195.*	1025T,1433T,2967T	US,JP,CN,FR,TR,DE,KR,GB 218,125,88,222,24,60,220,85,82
20	10	30,435	337	24.*, 129.*, 195.*	1026U,1026U1028U1027U,1027U	CA,CN 24,60

Table 4. Some detailed characteristics related to the composition of different zombie armies.

Zombie Army Id	Home Users (DSL, Cable, PPP)	Operating Systems (P0f)
1	spoofed IP's	-
4	69%	Windows 2000 SP (68%), Windows XP Pro (5%)
5	27%	Windows 2000 SP (50%), Windows XP Pro (21%)
6	38%	Windows 2000 SP (2%), unknown (98%)
9	29%	Windows 2000 SP (63%), Windows XP Pro (16%)
10	34%	Windows 2000 SP (10%), unknown (87%)
11	61%	Windows 2000 SP (56%), unknown (35%)
12	26%	Windows 2000 SP (61%), Windows XP Pro (17%)
20	spoofed IP's	-

ii) a large scan sweep by the same army targeting several consecutive class A-subnets;

iii) within a same army, multiple changes in the port sequences (or exploits) used by zombies to scan or to attack;

iv) a coordination between different armies.

Zombie army 12 (ZA12) is an interesting case in which we can observe the behaviors *ii)* and *iii)*. Fig. 8 represents the output of the fuzzy system modeling this phenomenon. Each bar graph represents the fuzzy output z_i for a given attack dimension, whereas the last plot shows the final aggregated output from which the decision to group those events together was made (i.e., $F(z_i)$). We can clearly see that the targets and the activities of this army have evolved between certain attack events (e.g., when the value of z_i is low). That is, this army has been scanning (at least) four consecutive class A-subnets during its lifetime (still 183 days), while probing at the same time three different ports on these subnetworks.

Then, the largest zombie army observed by the sensors (ZA10) has showed the behaviors *i)* and *iv)*. On Fig. 9, we can see that this army had four waves of activity during which it was randomly scanning 5 different subnets (note the almost perfect coordination among those attack events) on Windows ports (445T, 139T), preceded by ICMP. When inspecting the subnet distributions of those different attack waves, we could clearly ob-

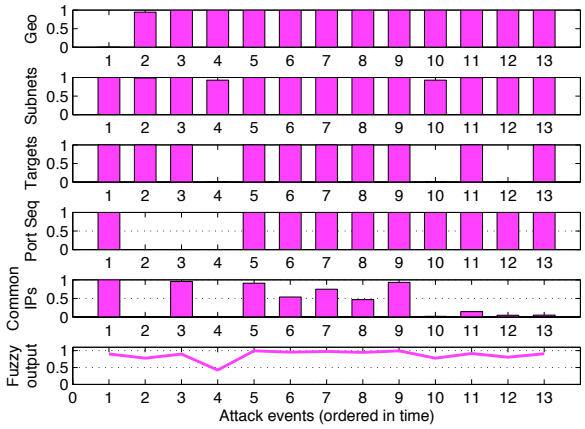

Figure 8. Output of the fuzzy inference system (z_i and $F(z_i)$) modeling the zombie army nr 12.

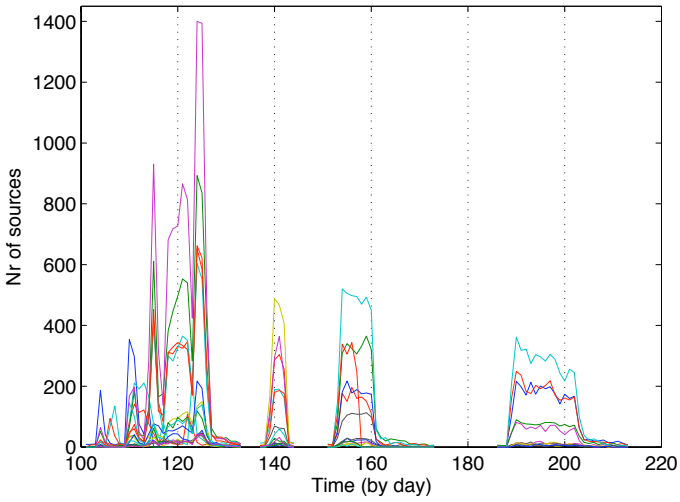

Figure 9. Time series of coordinated attack events for zombie army ZA10 (Nr of sources / day).

serve a drift in the origins of those sources, quite likely as certain machines were infected by (resp. cleaned from) the bot software. Finally, we found another smaller army (ZA11) that is clearly related to ZA10 (e.g., same temporal behavior, similar activity, same targets); but in this case, a different group of zombie machines, resulting in very different subnet CDF's on Fig. 7), was used to attack only specific IP addresses on our sensors, probably by taking advantage of the results given by the army of scanners (ZA10). The scanners were probably using some OS fingerprinting techniques to detect Windows operating systems, since only those ones were targeted by the attackers on ports 445 and

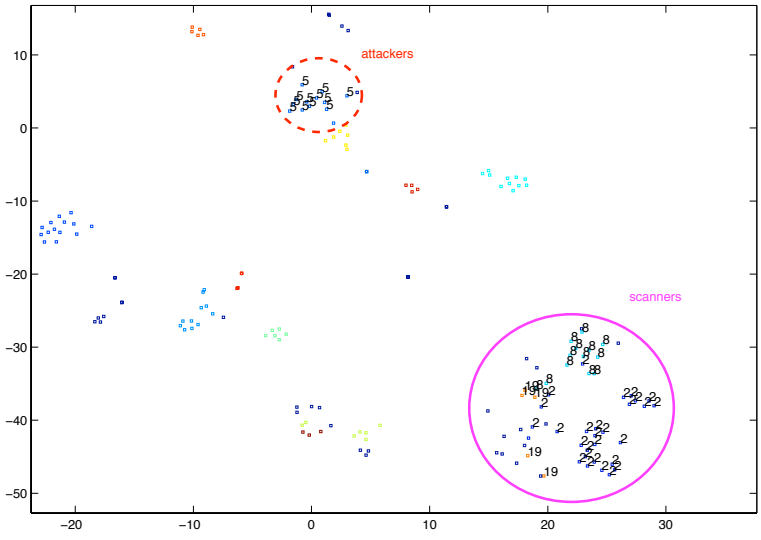

Figure 10. Visualization of the distributions of subnets of origins for Zombie armies 10 and 11, which involve two distinct communities of machines (scanners and attackers). The labels indicate the cliques' memberships of the attack events represented by the data points.

139 (and not the Linux honeypots). The distinction between scanners and attackers is even more visible on the 2D mapping (illustrated on Fig 10) obtained from the subnets distributions of these two zombie armies.

4. Conclusions

In this paper, we have introduced an analysis framework to identify, observe and characterize zombie armies on the Internet, based on the attack traces they have left on distributed sensors. Recent cyber-conflicts have showed that zombie armies and botnets can be easily turned into digital weapons and used to perform DDoS attacks against the network infrastructure of a Nation. It is thus very important to understand the long-term behavior of botnet armies, and their strategic evolution, in order to deploy effective countermeasures against those latent threats. Our analysis is based on the application of appropriate knowledge discovery techniques and a multi-criteria decision-making process. A key aspect of the proposed method is the exploitation of external characteristics of malicious sources, such as their spatial distributions in terms of countries and IP subnets. Our experiments on a set of real-world attack traces have also highlighted some interesting aspects of the global characteristics of such zombie armies, such as their high resilience and the high attack capacity that zombie machines can potentially offer. As future work, we envisage to extend our method to other data sets, such as high-interaction (client) honeypot data, or malware data sets, and to include even more relevant attack features so

as to improve further the inference capabilities of the system, and thus also our insights into malicious behaviors observed on the Internet.

Acknowledgements

This work has been partially supported by the European Commission through project FP7-ICT-216026-WOMBAT funded by the 7th framework program. The opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the European Commission.

References

[1] Paul Barford and Vinod Yegneswaran. *An Inside Look at Botnets*. Advances in Information Security. Springer, 2006.

[2] David Barroso. Botnets - the silent threat. In *European Network and Information Security Agency (ENISA)*, November 2007.

[3] Zesheng Chen, Chuanyi Ji, and Paul Barford. Spatial-temporal characteristics of internet malicious sources. In *Proceedings of INFOCOM*, pages 2306–2314, 2008.

[4] M. P. Collins, T. J. Shimeall, S. Faber, J. Janies, R. Weaver, M. De Shon, and J. Kadane. Using uncleanliness to predict future botnet addresses. In *IMC '07: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, pages 93–104, New York, NY, USA, 2007. ACM.

[5] Evan Cooke, Farnam Jahanian, and Danny McPherson. The Zombie roundup: Understanding, detecting, and disrupting botnets. In *Proceedings of the Steps to Reducing Unwanted Traffic on the Internet (SRUTI 2005 Workshop)*, Cambridge, MA, July 2005.

[6] Crime-Research. Cyberwar: Russia vs estonia, http://www.crime-research.org/articles/cyberwar-russia-vs-estonia/, [may 09].

[7] Darkreading. Botnets behind georgian attacks offer clues, http://www.darkreading.com/security/app-security/showarticle.jhtml?articleid=211201216, [may 09].

[8] G. Gu, R. Perdisci, J. Zhang, and W. Lee. BotMiner: Clustering analysis of network traffic for protocol- and structure-independent botnet detection. In *Proceedings of the 17th USENIX Security Symposium*, 2008.

[9] Guofei Gu, Junjie Zhang, and Wenke Lee. BotSniffer: Detecting botnet command and control channels in network traffic. In *Proceedings of the 15th Annual Network and Distributed System Security Symposium (NDSS'08)*, February 2008.

[10] Geoffrey Hinton and Sam Roweis. Stochastic neighbor embedding. In *Advances in Neural Information Processing Systems 15*, volume 15, pages 833–840, 2003.

[11] A.K. Jain and R.C. Dubes. *Algorithms for Clustering Data*. Prentice-Hall advanced reference series, 1988.

[12] S. Kullback and R. A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics 22: 79-86.*, 1951.

[13] Wenke Lee, Cliff Wang, and David Dagon, editors. *Botnet Detection: Countering the Largest Security Threat*, volume 36 of *Advances in Information Security*. Springer, 2008.

[14] C. Leita, V.H. Pham, O. Thonnard, E. Ramirez-Silva, F. Pouget, E. Kirda, and Dacier M. The Leurre.com Project: Collecting Internet Threats Information Using a Worldwide Distributed Honeynet. In *Proceedings of the WOMBAT Workshop on Information Security Threats Data Collection and Sharing, WISTDCS 2008*. IEEE Computer Society press, April 2008.

[15] Corrado Leita, Ken Mermoud, and Marc Dacier. Scriptgen: an automated script generation tool for honeyd. In *Proceedings of the 21st Annual Computer Security Applications Conference*, December 2005.

[16] J. Lin. Divergence measures based on the shannon entropy. *Information Theory, IEEE Transactions on*, 37(1):145–151, Jan 1991.

[17] Arbor Networks. http://asert.arbornetworks.com/2007/05/estonian-ddos-attacks-a-summary-to-date, [may 09].

[18] Ruoming Pang, Vinod Yegneswaran, Paul Barford, Vern Paxson, and Larry Peterson. Characteristics of internet background radiation. In *IMC '04: Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*, pages 27–40, New York, NY, USA, 2004. ACM.

[19] Markus Kötter Georg Wicherski Paul Bächer, Thorsten Holz. Know your enemy: Tracking botnets. In *http://www.honeynet.org/papers/bots/*.

[20] V. Pham, M. Dacier, G. Urvoy Keller, and T. En Najjary. The quest for multi-headed worms. In *DIMVA 2008, 5th Conference on Detection of Intrusions and Malware & Vulnerability Assessment, Paris, France*, Jul 2008.

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

[22] The Leurre.com Project. http://www.leurrecom.org.

[23] Niels Provos. A virtual honeypot framework. In *Proceedings of the 12th USENIX Security Symposium*, pages 1–14, August 2004.

[24] M. Abu Rajab, J. Zarfoss, F. Monrose, and A. Terzis. A multifaceted approach to understanding the botnet phenomenon. In *IMC '06: Proceedings of the 6th ACM SIGCOMM conference on Internet measurement*, pages 41–52, New York, NY, USA, 2006. ACM.

[25] Symantec Security Response. W32.rahack.h, [april 2009].

[26] Olivier Thonnard and Marc Dacier. A framework for attack patterns' discovery in honeynet data. *DFRWS 2008, 8th Digital Forensics Research Conference, August 11- 13, 2008, Baltimore, USA*, 2008.

[27] Olivier Thonnard and Marc Dacier. Actionable knowledge discovery for threats intelligence support using a multi-dimensional data mining methodology. In *ICDM'08, 8th IEEE International Conference on Data Mining series, December 15-19, 2008, Pisa, Italy*, Dec 2008.

[28] Olivier Thonnard, Wim Mees, and Marc Dacier. Addressing the attack attribution problem using knowledge discovery and multi-criteria fuzzy decision-making. In *KDD'09, 15th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Workshop on CyberSecurity and Intelligence Informatics, Paris, France*, Jun 2009.

[29] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9:2579–2605, November 2008.

[30] Ronald R. Yager. On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *IEEE Trans. Syst. Man Cybern.*, 18(1):183–190, 1988.

[31] Vinod Yegneswaran, Paul Barford, and Johannes Ullrich. Internet intrusions: global characteristics and prevalence. In *SIGMETRICS*, pages 138–147, 2003.