Towards a Taxonomy of Network Scanning Techniques

Richard J Barnett
Security and Networks Research Group (SNRG)
Department of Computer Science
Rhodes University
Grahamstown, South Africa
barnettrj@acm.org

Barry Irwin
Security and Networks Research Group (SNRG)
Department of Computer Science
Rhodes University
Grahamstown, South Africa
b.irwin@ru.ac.za

ABSTRACT

Network scanning is a common reconnaissance activity in network intrusion. Despite this, it's classification remains vague and detection systems in current Network Intrusion Detection Systems are incapable of detecting many forms of scanning traffic.

This paper presents a classification of network scanning and illustrates how complex and varied this activity is. The presented classification extends previous, well known, definitions of scanning traffic in a manner which reflects this complexity.

Categories and Subject Descriptors

C.2.0 [**Computer- Communication Networks**]: General—*Security and protection*; C.2.3 [**Computer- Communication Networks**]: Network Operations—*Network monitoring*

General Terms

Security

1. INTRODUCTION

As we become ever more reliant on the Internet, the need to protect our networks from security threats becomes an essential aspect of network management. As a result, firewall technologies have become an integral part of network security. Despite this, intrusions are still attempted and frequently succeed [5].

Network Intrusion Detection has become an important technology in the detection of intrusions and intrusion attempts. Scan-detection is a useful metric in intrusion detection as it permits monitoring of precursors to intrusion attempts.

The Open-Source NIDS Snort [3] and Bro [1] both have the ability to perform scan-detection. However, despite their use in the monitoring of reconnaissance traffic, they are unable to detect all forms of scanning traffic.

As will be presented in Section 7, we are actively engaged in the development of more sophisticated scan-detection modules in Snort and Bro. In this paper, we present a precursor to that research in which we have classified network scanning techniques. This classification is presented in the form of a taxonomy which will later be used to facilitate the development of these new modules.

1.1 Paper Organisation

The remainder of this paper is constructed as follows. Section 2 discusses previous work which is related to our research. Section 3 presents the methodology followed in the construction of our taxonomy. Thereafter, Section 4 highlights the findings of our scanning traffic analysis. We then present our taxonomy in Section 5 and conclusions which we have drawn from this process in Section 6. Lastly, we discuss our ongoing research and how it relates to this work in Section 7.

2. BACKGROUND AND RELATED WORK

This paper investigates network scanning and forms a taxonomy of common techniques. A significant amount of work has been performed in this area and the related field of scan-detection.

We present, in this section, related work as it pertains to our research. We present those aspects which have formed a basis from which our work has developed.

We have investigated work on scanning techniques, network telescopes, intrusion detection and related taxonomic classification. The relation of these to this work will be discussed in the subsections which follow.

2.1 Scanning Techniques

The concept of network scanning as a reconnaissance exercise is a well known and defined concept, and has been in existence for a long period [7]. For example, Allman *et al.* [5] present a study of scanning from their observed onset of such traffic in the late 1990s. They studied a vast collection of scans over 12.5 years and present a number of statistics of scanning over that period. Their research shows that scanning is still prevalent many years after its inception.

Of particular interest to this research is Allman *et al.*'s def-

SAICSIT 2008, 6 - 8 October 2008, Wilderness Beach Hotel, Wilderness, South Africa

inition of what scanning traffic is. They state that connections can be classified as good, bad or unknown based on their success. Successful connections are good, unsuccessful are bad and in cases where the monitoring system is unable to determine success, they are flagged as unknown. Hosts are defined as scanning hosts if they are the source of a majority of bad traffic.

Scanning traffic has been categorised into a number of broad groups by different authors in the past. Yegneswaran *et al.* [20] classified scanning traffic into four distinct groups:

Vertical Scans are scans that target at least five ports on a single IP address from a single scanning host. Yegneswaran *et al.* note that these are usually performed by fully automated tools.

Horizontal Scans are, similarly, scans which originate from a single host to multiple target IP addresses on a single port. These scans are usually performed to identify hosts with a specific vulnerability.

Coordinated Scans or **Distributed Scans** are more complex scans which originate from multiple scanning hosts. These could target a number of ports over a number of machines. These usually come from more aggressive scanners and target an entire subnet at one time.

Stealth Scans are scans which attempt to avoid detection and are usually horizontal or vertical scans which are performed over an extended temporal reference frame.

A similar classification was discussed by Staniford *et al.* in [16]. In related work in the Department of Computer Science, Irwin and van Riel [10] also discuss the above scanning techniques. They, however, use the terms *Port Scans* and *Port Sweeps* as alternates for Vertical and Horizontal Scans. This is done to conform to definitions available in the Snort documentation [11, 15].

These categorisations will be useful as a starting point for our taxonomy, as we will see in Section 5.

2.2 Network Telescopes

Network telescopes are network monitoring tools which exist at otherwise unused IP addresses. The concept was introduced by Moore [13] who explains that larger telescopes are more effective at rapid monitoring of Internet traffic.

Due to the fact that telescopes make use of unused IP space and that they are passive monitoring devices [14], traffic which is targeted at them never leads to an established connection. As a result, all traffic which reaches a network telescope (by Allman *et al.*'s definition) could be classified as scanning traffic.

However, as Pang *et al.* [14] observe, network telescopes are subject to *background radiation* traffic, which may not be of a scanning nature. This traffic comprises malicious traffic, backscatter and traffic resulting from network misconfiguration. Scanning activity falls into the category of malicious traffic.

Due to their nature, network telescopes can generate a lot of data. The analysis of this data can be a time consuming and difficult exercise. This is because traditional tools, such as *Wireshark* [4] are unable to perform effective analysis such a large volume of data [19]. van Riel and Irwin [10, 18, 19] have developed a graphical visualisation tool, *InetVis*, for the rapid review of network traffic in a 3D scatter plot.

2.3 Network Security Taxonomies

There are a number of related taxonomies available in Network Security and Intrusion Detection in particular. Taxonomies are hierarchical classifications of the entities within a specific domain [9]. They attempt to place each entity into mutually exclusive categories.

In the construction of our taxonomy, we have investigated a number of taxonomies in somewhat related fields. This analysis has offered insight into how taxonomies in these related fields have been constructed.

Deber *et al.* [6] have constructed a relatively simple taxonomy for the classification of Intrusion Detection Systems. Mirkovic and Reiher [12] have illustrated the complexities of constructing a taxonomy for Distributed Denial of Service attacks, which is structurally relevant to our own taxonomy. Taylor [17] presents a very sophisticated mathematical taxonomy for the classification of packet data.

Our taxonomy draws on the concepts from each of these, despite the dissimilarity of the fields to our own.

3. METHODOLOGY

We have followed a multifaceted approach in constructing our taxonomy of network scanning techniques. The approach we have taken is to derive a classification from existing research and from our own study. This work was performed for both theoretical and observed Internet traffic.

The process which was taken in forming our taxonomy is discussed in some detail in the following subsections.

3.1 Identification of Scanning Techniques

In order to effectively construct a taxonomy of scanning techniques, we have had to identify and evaluate a number of actual scans. To this end, we have evaluated scans which were constructed using the *Nmap* [2] network scanning tool and scans which were identified in network telescope data.

The large majority of our analysis was performed using the traffic analysis tool *InetVis* [19]. This permitted us to view scans spatially in a "real time" review. This tool permitted us to evaluate scanning traffic at a high level. More detailed analysis of the traffic was performed by placing packet header information into a database and using SQL to extract specific information.

Scans constructed using *Nmap* were targeted at a FreeBSD Software Discard Interface (disc), and was captured using tcpdump. Before analysing the tcpdump packet capture (cap) with *InetVis*, we were required to rewrite these files with Ethernet frames as *InetVis* is unable to visualise disc interface traffic. These scans were evaluated using both *InetVis* and by making use of the *Nmap* documentation.

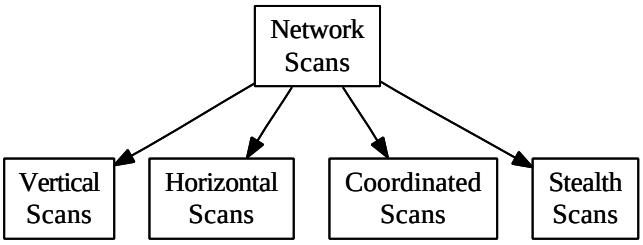


Figure 1: Prior Classification

Most of the scanning traffic analysed, however, was obtained from network telescope captures. These captures were obtained between 2005 and 2007 from a small (Class C /24) network telescope at Rhodes University.

Due to the immense volume of scanning traffic available, we were forced to analyse it significantly more rapidly than in real time. We analysed the traffic with a speedup of approximately 8640× to evaluate one months worth of data in roughly five minutes. This process permitted the isolation of scans from the traffic captures, and pattern information to be identified.

This process resulted in significant information on the nature of network scanning techniques. The next subsection discusses how this information was used in the construction of a suitable taxonomy.

3.2 Constructing a Taxonomy

The approach we took in constructing our taxonomy was to derive information from existing classifications and from our own research. We attempted to group traffic without overlap, however, as will be seen in Section 5, this was not strictly possible.

It was our intention to classify traffic into "buckets" which derived from the classification presented by Yegneswaran *et al.* [20] (and illustrated in Figure 1), but this was not strictly possible. We have, however, investigated methods of integrating our classification with this existing, well known classification.

4. ANALYSIS

The analysis of scanning traffic produced the bulk of the information which was used in the construction of our taxonomy. We have already discussed the approach which was taken in this process, and this section discusses the results of that work.

4.1 Theoretical Scanning Traffic

There are a number of automated tools available for scanning hosts and networks. One of the best known of these is *Nmap* which was first developed by Fyodor in 1997 [7].

Nmap has a plethora of options to direct the type of scan it creates. These extend from simple host discovery scans using pings, to complex scans which can even perform remote operating system detection [8]. Each of these scans is also available with a wide variety of timing options [7].

Figure 2 highlights three very distinct scans which are possible with *Nmap* and presents graphically one of the major flaws with the classification in Figure 1. This figure (2(a)), and similar figures in this paper were produced using *InetVis* and show source and destination IP addresses on the two horizontal axes (red and blue respectively) and destination port on the vertical axis (green) of the cube. The plane below the cube represents ICMP traffic.

Figure 2(a) shows a simple TCP `connect()` scan. It can be seen that packets (represented by points in the cube) form a single vertical line. Figure 2(b) shows an almost identical scan except that the packets are UDP, rather than TCP. Both scans are clearly Vertical scans in the sense of Figure 1, however they are very distinct in function. This highlights the first of the problems with such a simplistic classification. Different Layer Four protocols should be grouped separately.

Similarly, Figure 2(c) shows a *Nmap* IP Protocol Scan. From the figure, it would appear that only two IP protocols are being scanned (ICMP and one other), but this is largely due to *InetVis*'s inability to differentiate between TCP and UDP traffic, as it is plotting a point for TCP and UDP in the same place. This traffic appears as if it is not scanning traffic, but is, in fact, a Layer Two scan.

Furthermore, *Nmap* offers the ability to perform scans which are all TCP, but are sent with differing flags set. Specifically, it permits scanning with one of the SYN, ACK, or FIN flags set.

To assist *Nmap* to avoid detection by tools such as Snort's *sfPortscan* [15] module, it offers a number of options with affect the timing of scans. These timing options permit scans to be performed very rapidly (possibly with packets being dropped) to very slowly (with a packets being sent with a gap of at least five minutes between them).

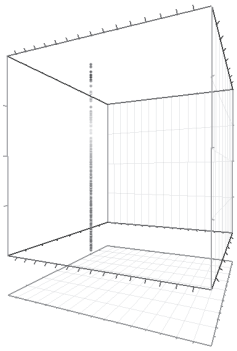
Whilst we have only discussed *Nmap* in this section, we feel that, due to it being a well known and mature product, the information obtained is representative of automated scanning tools.

4.2 Observed Scanning Traffic

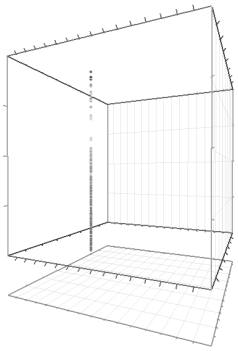
Traffic destined for our telescope has exhibited a number of interesting characteristics. Figure 3 shows a selection of scans identified from the telescope data. The images are, however, not exhaustive and a number of other scans have been identified.

The most common form of scanning traffic identified was one-to-many host scanning on multiple ports. This is illustrated by Figure 3(a) which shows several concurrent scans. Conversely, we found that "full" one-to-one scans (such as the TCP `connect()` scan of Figure 2(a)) are uncommon.

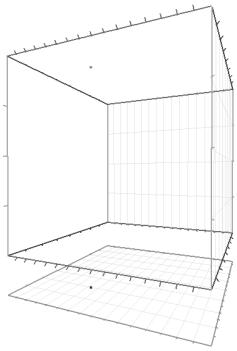
A number of scans with patterns other than horizontal and vertical were also noted. Figure 3(b) shows some scans which form a diagonal pattern. These were noted as one-to-many scans, where the scanned port increased as the scanned host changed. Whilst not shown, stepped scans were also noted in packet captures, where the port changed every few hosts scanned.



(a) TCP `connect()` Scan

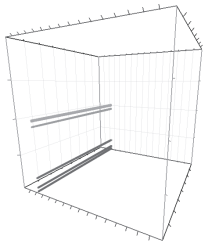


(b) UDP Scan

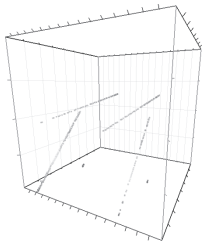


(c) IP Protocol Scan

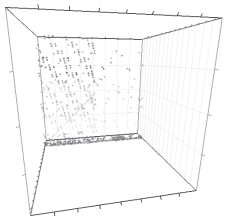
Figure 2: *Nmap* Single Host Scans



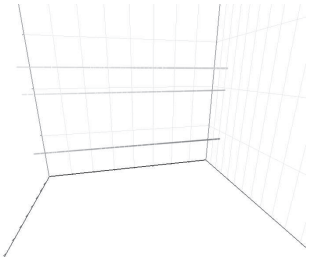
(a) Concurrent Scanning



(b) Anomalous Diagonals



(c) Pseudo Random Traffic



(d) Creepy Crawly Scan

Figure 3: Observed Scan Examples

Timing was also noted as being used. Figure 3(d) shows long running scans. The points to the left of the diagram have begun to fade from the image with age. Similarly, we found that a number of these scans would often follow each other, "chasing" the previous scan.

We discovered a lot of seemingly random traffic in the packet captures which, over time, proved to be less random than we had first thought. An example of this pseudo random traffic is shown in Figure 3(c). This traffic was usually seen in the higher port ranges, and often formed sets of short diagonals.

Our analysis of telescope traffic has shown that the majority of scans are found in the first half of the /24 subnet. Additionally, most of the observed scans were TCP.

These observations have formed the basis of our taxonomy, which is presented in the next section.

5. TAXONOMY

Having investigated work on network scanning and performed an extensive analysis on two years of network telescope traffic, this section introduces our taxonomy into network scanning. Figure 4 shows our classification graphically.

In our endeavour to restrict scans to a single grouping in the tree, we came to the realisation that numerous aspects of the classification could apply to different scans. In this regard we present our taxonomy with attributes.

In Figure 4 the root of the classification is represented by the inverted node labelled "TCP/IP Scan". Actual scans are traced through arrows down the tree to solid rectangular nodes. These scans are attributed with dashed lines. Scans and attributes are grouped together using mushroom shaped nodes. An attribute link also applies to all nodes below it in the tree.

We will discuss the scan types individually, along with a further discussion on each of the attributes.

5.1 Scan Types

We found three levels of overall scan type, which encompass seven different scans. These scans all reference different attributes, which will be discussed in Section 5.2.

5.1.1 Layer Two Scans

Layer two scans are those which scan hosts at layer two of the TCP/IP Model. These scans are typically performed to identify which Layer *Three* protocols a host will respond to. Figure 2(c) shows a theoretical example of such a scan from *Nmap*.

Due to their nature, Layer Two scans target a number of different Layer Three protocols. You do not find such a scan targeting just one destination protocol as this would be a ICMP, TCP or UDP scan. As with all scans, they can occur at differing speeds and with differing numbers of source hosts and destination hosts.

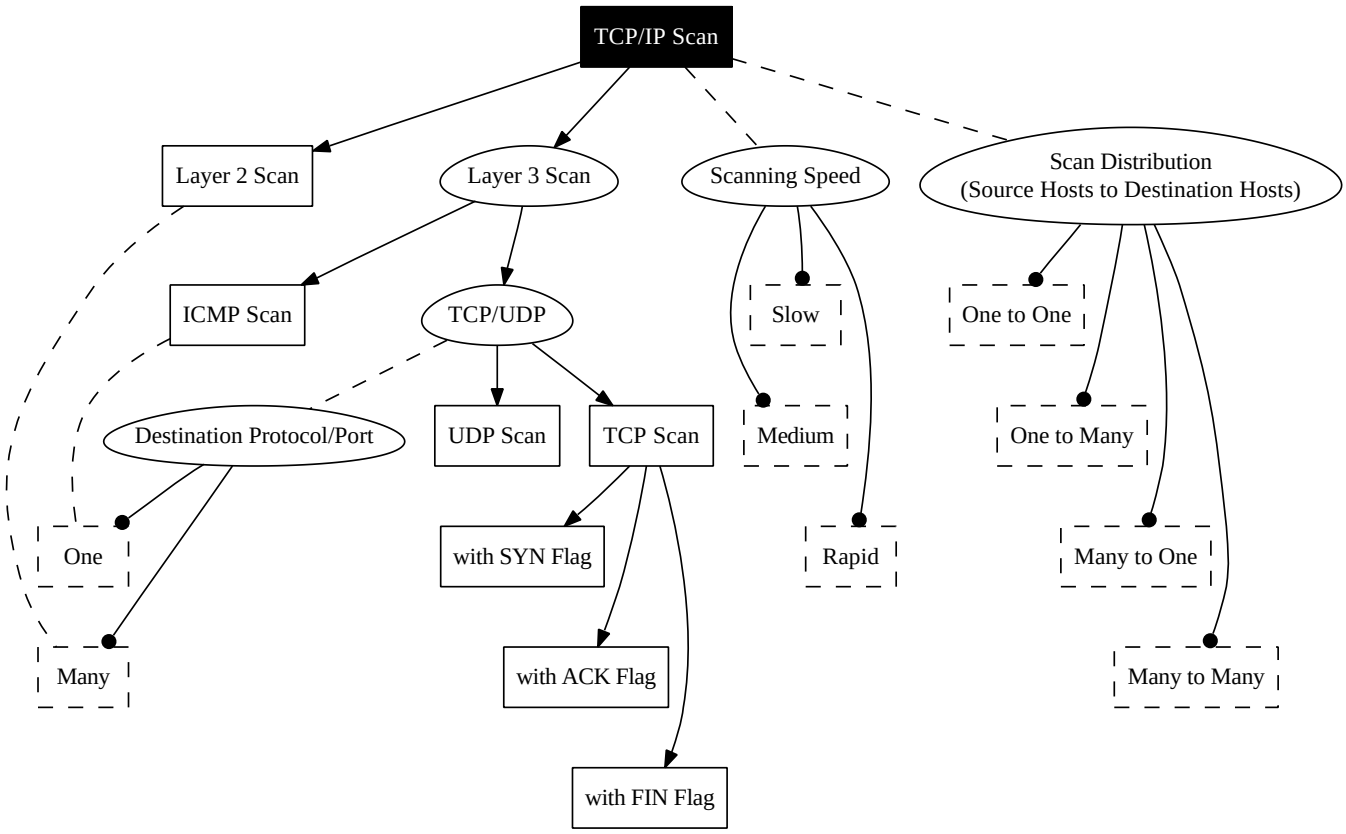


Figure 4: Taxonomy of Network Scanning Techniques

5.1.2 ICMP Scans

ICMP scanning is the simplest of the scanning types as it usually involves just a single ICMP Ping packet. Because of this we classify a single ping as a scan, but in practice, this type of scan is unlikely to use the "One to One" and "Many to One" Distribution attributes.

5.1.3 UDP Scans

UDP scans are the second most common scan after the wide variety of TCP scans. They are available in numerous options due to their ability to target a single or multiple UDP ports on the target host.

The most common combinations of these are multiple scanning hosts targeting a single destination port and a single scanning host targeting multiple destination ports.

5.1.4 TCP Scans

TCP scans are the most complex as they are available in a wide variety of options. A TCP scan can be performed without any explicit flag options (a TCP `connect()` scan, for example) or with the explicit crafting of packets with specific flags. The flags which were identified as being set specifically (also corresponding to *Nmap* scans) were the SYN, ACK and FIN flags.

All these scans are attributed in a similar way to UDP scans, and are commonly found in similar combinations to the UDP scan. However, we did find that "One to Many" scans targeting a number ("Many") of ports were the most common TCP scans (similar to the one shown in Figure 3(a)).

5.2 Attributes

Three groups of attributes have been identified. These correspond to a total of nine different attributes. In this context, attributes represent features which are common to a number of different scan types.

5.2.1 Scanning Speed

Scans can be performed at a variety of speeds. The differing speeds usually occur in order to avoid detection by Intrusion Detection Systems. Because of this, all types of scans can reflect these speeds. *Nmap* offers six different speed options, but we have condensed these to three. Slow, Medium (which would be an average speed) and Rapid. Scans avoiding detection would normally be slow scans.

5.2.2 Scan Distribution

Scans scan be distributed in four ways. This is the permutations of number of scanning hosts and number of scanned hosts (with the options being "one" and "many"). One-to-One and One-to-Many scans are far more common than their distributed counterparts. Many-to-One is only seen infrequently.

5.2.3 Destination Protocol/Port

This category of attributes actually has two different meanings dependant on which type of scan it is attributed to. In

Layer Two and ICMP scanning, this reflects the protocol, and for TCP and UDP scans, this reflects the destination port.

This reflects the number of protocols/ports scanned, and is the one category which is attributed in a complex manner. This is due to the variety of how scans use protocols and ports.

5.3 Taxonomy vs. Prior Classification

The taxonomy presented in Figure 4 presents a fairly detailed look at scan classification. It encompasses all the features of the classification presented in Figure 1, but takes a significantly more detailed look at the problem. In particular, it incorporates protocol level classification, which was not considered as a component of the common, simplistic model.

However, the classification of Figure 1 presents some interesting challenges to our taxonomy. It was our intention to use it as the starting point for our work, however, we have found that each of the groups in that classification corresponds to a variety of options in our own.

Figure 5 presents that classification again, with the attributes of our taxonomy. This shows how many possible scan types are classified into a single of these categories. It also shows the extensive overlap of these groups. The nodes and edges of this diagram have the same meanings as they do in Figure 4.

In particular, note that Stealth and Coordinated Scans consist of Horizontal and Vertical scans, albeit with different timing and distribution options. Further note that whilst Horizontal and Vertical scans each have a fairly well defined structure (each with specific immutable attributes), this does not reflect the variety of such scans. Figures 2(a) and 2(b) are a good example of this.

This shows that whilst the structure of Figure 1 is well known and used, it does not offer the flexibility of our taxonomy.

6. CONCLUSIONS

We have seen that there are a number of aspects to network scanning. This paper has highlighted our analysis of network scanning traffic and how it can be used to construct a taxonomy of scanning types.

We have presented a possible taxonomy for scanning, but realise that there are numerous approaches which can be taken to the effective classification of scanning traffic. We do not presume that our classification is either the best, or even most complete method of describing scanning traffic. It is simply a reflection of the data available to us at the time of construction.

We do, however, observe that the well known simplistic classification does not offer the flexibility of designing systems which deal with network scanning traffic. We are actively using our taxonomy in our ongoing research as is described in the next section.

7. FUTURE WORK

This paper has investigated techniques for network scanning. This classification forms part of our ongoing research into scanning and scan-detection in particular.

It has been shown by Irwin and van Riel [10] that there are flaws in the scan-detection engines in the Network Intrusion Detection Systems Snort and Bro. Our current research is investigating which scanning techniques are effectively handled by Snort and Bro, and which could be further improved.

Our intention is to use the results of this classification in the development of new scan-detection modules for both Snort and Bro, so as to make their use in Scan Detection viable in large scale networks.

8. ACKNOWLEDGEMENT

The authors would like to acknowledge the support of the National Research Foundation and the financial support of Telkom SA, Business Connexion, Comverse SA, Stortech, Tellabs, Amatole, Mars Technologies, openVOICE and THRIP through the Telkom Centre of Excellence in the Department of Computer Science at Rhodes University.

9. REFERENCES

[1] Bro intrusion detection system - bro overview. Online: `http://www.bro-ids.org/`, Accessed: 28/01/2008.

[2] Nmap - free security scanner for network exploration & security audits. Online: `http://nmap.org/`, Accessed: 29/05/2008.

[3] Snort - the de facto standard for intrusion detection/prevention. Online: `http://www.snort.org/`, Accessed: 28/01/2008.

[4] Wireshark: Go deep. Online: `http://www.wireshark.org/`, Accessed: 21/06/2008.

[5] M. Allman, V. Paxson, and J. Terrell. A brief history of scanning. In *IMC '07: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, pages 77–82, New York, NY, USA, 2007. ACM.

[6] H. Debar, M. Dacier, and A. Wespi. Towards a taxonomy of intrusion-detection systems. *Comput. Netw.*, 31(9):805–822, 1999.

[7] Fyodor. The art of port scanning. *Phrack Magazine*, 7(51):Article 11 of 17, September 1997.

[8] Fyodor. Remote os detection via tcp/ip stack fingerprinting. *Phrack Magazine*, 8(54):Article 09 of 12, December 1998.

[9] D. GaŽevic, D. Djuric, and V. Devedđic. *Model Driven Architecture and Ontology Development*. Springer, 2006.

[10] B. Irwin and J.-P. van Riel. Inetvis: a graphical aid for the detection and visualisation of network scans. In *Conference on Vizualisation Security (VizSec2007)*, 2007.

[11] T. Kohlenberg, R. Alder, J. Dr. Everett F.Carter, J. C. Foster, M. Jonkman, R. Marty, and M. Poor. *Snort Intrusion Detection and Prevention Toolkit*. Syngress Publishing Inc., 2007.

[12] J. Mirkovic and P. Reiher. A taxonomy of ddos attack and ddos defense mechanisms. *SIGCOMM Comput. Commun. Rev.*, 34(2):39–53, 2004.

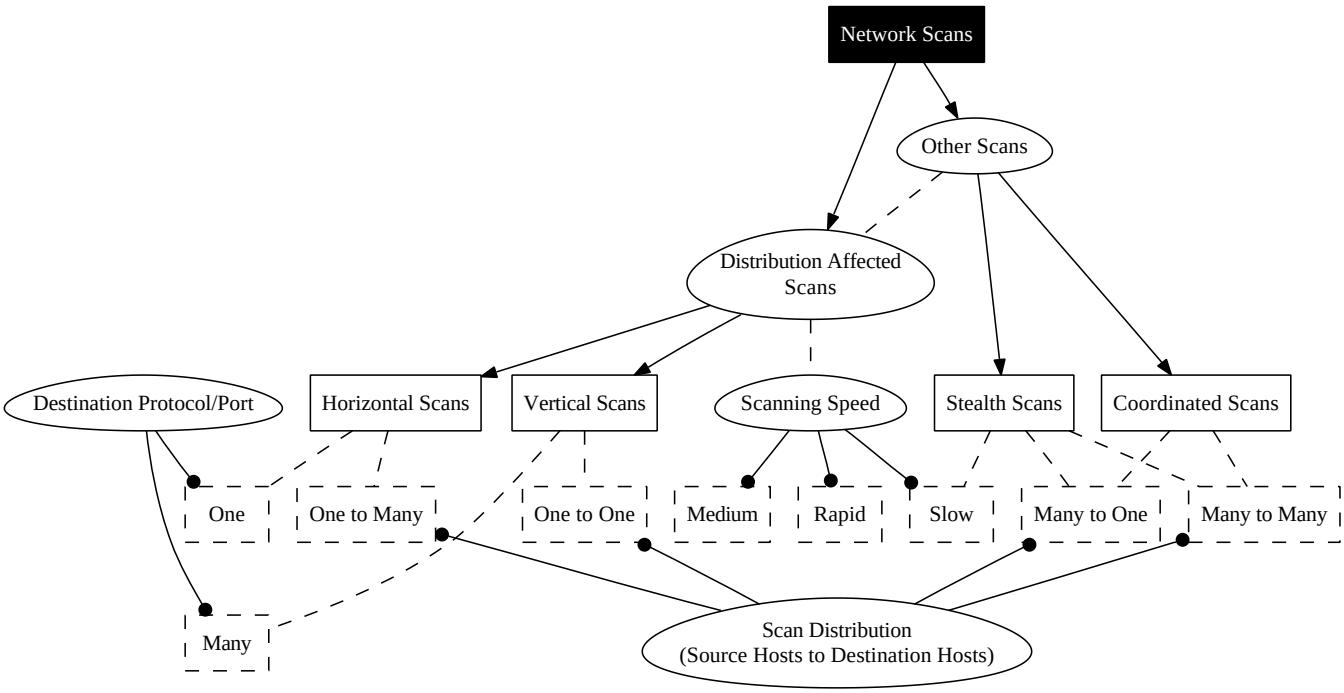


Figure 5: Prior Classification with Attributes

[13] D. Moore. Network telescopes: Observing small or distant security events. Online: `http://www.caida.org/publications/presentations/2002/usenix_sec/`, Accessed: 16/06/2008.

[14] R. Pang, V. Yegneswaran, P. Barford, V. Paxson, and L. 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.

[15] D. Roelker, M. Norton, and J. Hewlett. *sfPortscan*.

[16] S. Staniford, J. A. Hoagland, and J. M. McAlerney. Practical automated detection of stealthy portscans. *J. Comput. Secur.*, 10(1-2):105–136, 2002.

[17] D. E. Taylor. Survey and taxonomy of packet classification techniques. *ACM Comput. Surv.*, 37(3):238–275, 2005.

[18] J.-P. van Riel and B. Irwin. Identifying and investigating intrusive scanning patterns by visualizing network telescope traffic in a 3-d scatter-plot. In H. Venter, J. Eloff, L. Labuschagne, and M. Eloff, editors, *Proceedings of 6th Annual Information Security South Africa (ISSA) 2006*. Information Security South Africa, 2006.

[19] J.-P. van Riel and B. Irwin. Inetvis, a visual tool for network telescope traffic analysis. In *Afrigaph '06: Proceedings of the 4th international conference on Computer graphics, virtual reality, visualisation and interaction in Africa*, pages 85–89, New York, NY, USA, 2006. ACM.

[20] V. Yegneswaran, P. Barford, and J. Ullrich. Internet intrusions: global characteristics and prevalence. In *SIGMETRICS '03: Proceedings of the 2003 ACM SIGMETRICS international conference on Measurement and modeling of computer systems*, pages 138–147, New York, NY, USA, 2003. ACM.