Code-Red: a case study on the spread and victims of an Internet worm

David Moore, Colleen Shannon, k claffy

Abstract— **On July 19, 2001, more than 359,000 computers connected to the Internet were infected with the Code-Red (CRv2) worm in less than 14 hours. The cost of this epidemic, including subsequent strains of Code-Red, is estimated to be in excess of $2.6 billion. Despite the global damage caused by this attack, there have been few serious attempts to characterize the spread of the worm, partly due to the challenge of collecting global information about worms. Using a technique that enables global detection of worm spread, we collected and analyzed data over a period of 45 days beginning July 2nd, 2001 to determine the characteristics of the spread of Code-Red throughout the Internet.**

In this paper, we describe the methodology we use to trace the spread of Code-Red, and then describe the results of our trace analyses. We first detail the spread of the Code-Red and CodeRedII worms in terms of infection and deactivation rates. Even without being optimized for spread of infection, Code-Red infection rates peaked at over 2,000 hosts per minute. We then examine the properties of the infected host population, including geographic location, weekly and diurnal time effects, top-level domains, and ISPs. We demonstrate that the worm was an international event, infection activity exhibited time-of-day effects, and found that, although most attention focused on large corporations, the Code-Red worm primarily preyed upon home and small business users. We also qualified the effects of DHCP on measurements of infected hosts and determined that IP addresses are not an accurate measure of the spread of a worm on timescales longer than 24 hours. Finally, the experience of the Code-Red worm demonstrates that wide-spread vulnerabilities in Internet hosts can be exploited quickly and dramatically, and that techniques other than host patching are required to mitigate Internet worms.

Keywords—**Code-Red, Code-RedI, CodeRedI, CodeRedII, worm, security, backscatter, virus, epidemiology**

CAIDA, San Diego Supercomputer Center, University of California, San Diego. E-mail: {cshannon,dmoore,kc}@caida.org.

Support for this work is provided by DARPA NMS Grant N66001-01-1-8909, NSF grant NCR-9711092, Cisco Systems URB Grant, and Caida members.

I. Introduction

At 18:00 on November 2, 1988, Robert T. Morris released a 99 line program onto the Internet. At 00:34 on November 3, 1988, Andy Sudduth of Harvard University posted the following message: "There may be a virus loose on the Internet." Indeed, Sun and VAX machines across the country were screeching to a halt as invisible tasks utilized all available resources [1] [2].

No virus brought large computers across the country to a standstill – the culprit was actually the first malicious worm. Unlike viruses and trojans which rely on human intervention to spread, worms are self-replicating software designed to spread throughout a network on their own. Although the Morris worm was the first malicious worm to wreak widespread havoc, earlier worms were actually designed to maximize utilization of networked computation resources. In 1982 at Xerox's Palo Alto Research Center, John Shoch and Jon Hupp wrote five worm programs that performed such benign tasks as posting announcements [3]. However, research into using worm programs as tools was abandoned after it was determined that the consequences of a worm malfunction could be dire.

In the years between the Morris worm in November 1988 and June 2001, Several other worms achieved limited spread through host populations. The WANK (Worms Against Nuclear Killers) worm of October, 1989 attacked SPAN VAX/VMS systems via DECnet protocols [4]. The Ramen worm, first spread in January of 2001 targeted the wu-ftp daemon on RedHat Linux 6.2 and 7.0 systems [5]. Finally, the Lion Worm targeted the TSIG vulnerability in BIND in March of 2001 [6].

While all of these worms caused some damage, none approached the $2.6 billion cost of recovering from the Code-Red and CodeRedII worms [7]. We can no longer afford to remain ignorant of the spread and effects of worms as information technology plays a critical role in our global economy.

II. Background

On June 18, 2001, eEye released information about a buffer-overflow vulnerability in Microsoft's IIS web servers [8]. Microsoft released a patch for the vulnerability eight days later, on June 26, 2001 [9]. Then on July

12, 2001, the Code-RedI worm began to exploit the aforementioned buffer-overflow vulnerability in Microsoft's IIS web servers.

Upon infecting a machine, the worm checks to see if the date (as kept by the system clock) is between the first and the nineteenth of the month. If so, the worm generates a random list of IP addresses and probes each machine on the list in an attempt to infect as many computers as possible. However, this first version of the worm uses a static seed in its random number generator and thus generates identical lists of IP addresses on each infected machine. The first version of the worm spread slowly, because each infected machine began to spread the worm by probing machines that were either already infected or impregnable. On the 20th of every month, the worm is programmed to stop infecting other machines and proceed to its next attack phase in which it launches a Denial-of-Service attack against `www1.whitehouse.gov` from the 20th to the 28th of each month. The worm is dormant on days of the month following the 28th.

On July 13th, Ryan Permeh and Marc Maiffret at eEye Digital Security received logs of attacks by the worm and worked through the night to disassemble and analyze the worm. They christened the worm "Code-Red" both because the highly caffeinated "Code Red" Mountain Dew beverage fueled their efforts to understand the workings of the worm and because the worm defaces some web pages with the phrase "Hacked by Chinese". There is no evidence either supporting or refuting the involvement of Chinese hackers with the Code-RedI worm. The first version of the Code-Red worm (Code-RedI v1[1]) caused little damage. Although the worm's attempts to spread itself consumed resources on infected machines and local area networks, it had little impact on global resources.

The Code-RedI v1 worm is memory resident, so an infected machine can be disinfected by simply rebooting it. However, the machine is still vulnerable to repeat infection. Any machines infected by Code-RedI v1 and subsequently rebooted were likely to be reinfected, because each newly infected machine probes the same list of IP addresses in the same order.

At approximately 10:00 UTC in the morning of July 19th, 2001, we observed a change in the behavior of the worm as infected computers began to probe new hosts. At this point, a random-seed variant of the Code-RedI v1 worm began to infect hosts running unpatched versions of Microsoft's IIS web server. The worm still spreads by probing random IP addresses and infecting all hosts vulnerable to the IIS exploit. Unlike Code-RedI v1, Code-RedI v2 uses a random seed in its pseudo-random number generator, so each infected computer tries to infect a different list of randomly generated IP addresses at an observed rate of roughly 11 probes per second (pps). This seemingly minor change had a major impact: more than 359,000 machines were infected with Code-RedI v2 in just fourteen hours [10][11].

Because Code-RedI v2 is identical to Code-Red v1 in all respects except the seed for its pseudo-random number generator, the only direct damage to the infected host is the "Hacked by Chinese" message added to top level web pages on some hosts. However, Code-RedI v2 had a greater impact on global infrastructure due to the sheer volume of hosts infected and probes sent to infect new hosts. Code-RedI v2 also wreaked havoc on some additional devices with web interfaces, such as routers, switches, DSL modems, and printers [12]. Although these devices were not susceptible to infection by the worm, they either crashed or rebooted when an infected machine attempted to send them the unusual http request containing a copy of the worm.

Like Code-RedI v1, Code-RedI v2 can be removed from a computer simply by rebooting it. However, rebooting the machine does not prevent reinfection once the machine is online again. On July 19th, the number of machines attempting to infect new hosts was so high that many machines were infected while the patch for the vulnerability was being applied.

On August 4, 2001, an entirely new worm, CodeRedII began to exploit the buffer-overflow vulnerability in Microsoft's IIS web servers [13] [14]. Although the new worm is completely unrelated to the original Code-RedI worm, the source code of the worm contained the string "CodeRedII" which became the name of the new worm.

Ryan Permeh and Marc Maiffret analyzed CodeRedII to determine its attack mechanism. When a worm infects a new host, it first determines if the system has already been infected. If not, the worm initiates its propagation mechanism, sets up a "backdoor" into the infected machine, becomes dormant for a day, and then reboots the machine. Unlike Code-RedI, CodeRedII is not memory resident, so rebooting an infected machine does not eliminate CodeRedII.

Initial intuition might lead one to believe that this twenty-four hour delay will retard the spread of the worm so severely that it will never compromise a large number of machines, this is not the case. The delay adds a layer of subterfuge to the worm, since perusal of logs showing connections to the machine around the time that the ma-

[1] Although the initial Code-Red worm did not carry a suffix denoting its temporal position, we have added the suffix "I" in the interest of clarity, in the same manner as The Great War later came to be known as World War I.

chine begins to demonstrate symptoms of the infection (i.e. when it starts to actively spread the worm) will not yield any unusual activity.

After rebooting the machine, the CodeRedII worm begins to spread. If the host infected with CodeRedII has Chinese (Taiwanese) or Chinese (PRC) as the system language, it uses 600 threads to probe other machines. On all other machines it uses 300 threads. CodeRedII uses a more complex method of selecting hosts to probe than Code-RedI. CodeRedII generates a random IP address and then applies a mask to produce the IP address to probe. The length of the mask determines the similarity between the IP address of the infected machine and the probed machine. CodeRedII probes a completely random IP address 1/8th of the time. Half of the time, CodeRedII probes a machine in the same /8 (so if the infected machine had the IP address 10.9.8.7, the IP address probed would start with 10.), while 3/8ths of the time, it probes a machine on the same /16 (so the IP address probed would begin with 10.9.). Like Code-RedI, CodeRedII avoids probing IP addresses in the 224.0.0.0/8 (multicast) and 127.0.0.0/8 (loopback) address spaces. The bias toward the local /16 and /8 networks means that an infected machine may be more likely to probe a susceptible machine, based on the supposition that machines on a single network are more likely to be running the same software as machines on unrelated IP subnets.

The CodeRedII worm is much more dangerous than Code-RedI because CodeRedII installs a mechanism for remote, administrator-level access to the infected machine. Unlike Code-RedI, CodeRedII neither defaces web pages on infected machines nor launches a Denial-of-Service attack. However, the backdoor installed on the machine allows any code to be executed, so the machines could be used as "zombies" for future attacks (Denial-of-Service or otherwise).

III. Methodology

In this section, we detail our trace collection methodology, how we validated that the traffic we trace is from the spread of the worms, and describe our approaches for characterizing the type of hosts infected and their geographics locations.

Our analysis of the Code-RedI worm covers the spread of the worm between July 4, 2001 and August 25, 2001. Before Code-RedI began to spread, we were collecting data in the form of a packet header trace of hosts sending unsolicited TCP SYN packets into our /8 network. When the worm began to spread extensively on the morning of July 19, we noticed the sudden influx of probes into our network and began our monitoring efforts in earnest.

The data used for this study were collected from two locations: a /8 network and two /16 networks. Two types of data from the /8 network are used to maximize coverage of the expansion of the worm. Between midnight and 16:30 UTC on July 19, a passive network monitor recorded headers of all packets destined for the /8 research network. After 16:30 UTC, a filter installed on a campus router to reduce congestion caused by the worm blocked all external traffic to this network. Because this filter was put into place upstream of the monitor, we were unable to capture IP packet headers after 16:30 UTC. However, a backup data set consisting of sampled netflow [15] output from the filtering router was available for the /8 throughout the 24 hour period. The data from the /16 networks were collected with Bro between 10:00 UTC on July 19 and 7:00 on July 20 [16]. We merged these three sources into a single dataset. Hosts were considered to be infected if they sent at least two TCP SYN packets on port 80 to nonexistent hosts on these networks during this time period. The requirement of two packets helps to eliminate random source denial-of-service attacks from the Code-Red data.

Early on July 20, the filter was removed and we resumed packet header data collection. Although we collected data through October, we include data through August 25, 2001 in this study. No significant changes were observed in Code-RedI or CodeRedII activity between August 2001 and the pre-programmed shutdown of CodeRedII on October 1, 2001.

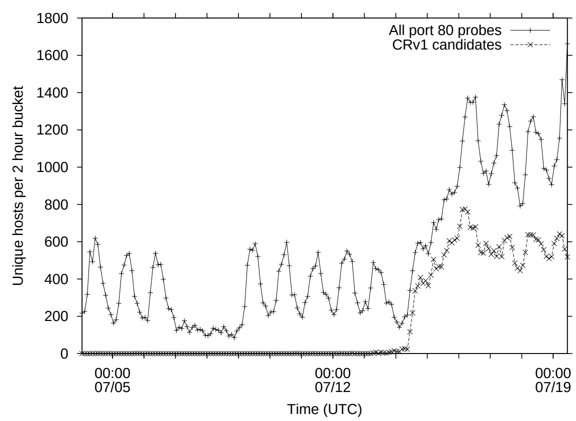


Fig. 1. Background level of unsolicited SYN probes and the beginning of the spread of the Code-RedI worm.

A constant background level of unsolicited TCP SYN packets, most likely port scans seeking to identify vulnerable machines, target the IPv4 address space. In our /8, this rate fluctuates between 100 and 600 hosts per two hour period, with diurnal and weekly variations. On July 12, the static-seed version of the Code-RedI worm began to spread. We noticed that the hosts that appeared clearly infected with Code-RedI v1 probed the same set of 23 IP addresses within our /8 research network. In Figure 1, we used the criterion of probing these 23 addresses to separate the Code-RedI v1 probes from the background port scans.

To confirm that the 23 addresses were actually among those probed by the worm, we reverse engineered the exploit to extract the IP addresses probed by its static-seed pseudo-random number generator. We obtained a disassembled version of the worm from eEye [17] and identified the code responsible for spreading the worm. The worm creates one hundred threads, each with its own static-seed and thus its own distinct, although not disjoint, set of IP addresses probed sequentially.

We examined the PRNG (Pseudo-Random Number Generator) code used to generate the target sequences and wrote a C implementation to generate the first one thousand IP addresses probed by each thread (approximately the first one million IP addresses). We extracted the IP addresses that fell in our /8 and found the same 23 address sequence we predicted from our packet trace data. The 23 addresses that fall in our research network actually occur very early in the generated sequences[2]. A machine newly infected with Code-RedI v1 probes our /8 network 23 times in the first fifteen minutes of propagation.

Once we had identified the IP addresses initially probed by the worm, we compared this sequence to the hosts we observed probing the 23 target addresses in our research network. We discovered that the first three hosts that probed our /8 research network were not contained in the IP address sequence probed by any thread. We believe that the individual (or individuals) responsible for the Code-RedI worm compromised these machines and seeded them with the worm to initiate the epidemic. The first two machines both appear to be located in the United States, one in Cambridge, Massachusetts and the other in Atlanta, Georgia. The third address appears to be in the city of Foshan in China's GuangDong province. However there remains no evidence linking Chinese hackers to the development or deployment of the Code-RedI worm.

We classify infected hosts using the DNS name of each host and a hand-tuned set of regular expression matches[3] (e.g. DNS names with "dialup" represent modems, "dsl" or "home.com" identifies broadband, etc.) into the following categories: mail servers, name servers, web servers, IRC servers, firewalls, dial-up, broadband, other (unclassifiable) hosts, and hosts with no hostname. The prevalence of each type of host is discussed in Section IV-B.4.

We also used Ixia's IxMapping [18] service to determine the latitude, longitude, and country of each IP address infected with the worm. IxMapping uses public data sources such as WHOIS and DNS, as well as specialized measurement to geographically place IP addresses. We identified a rough approximation of the timezone of each infected host based on this longitude.

IV. Results

In this section of the paper, we present the results of our trace analyses. We first characterize the spread of the Code-RedI and CodeRedII worms, then examine the properties of the infected host population, and finally determine the rate at which infected hosts are repaired.

A. Worm Spread

In this section, we examine the dynamics of the spread of the Code-RedI and CodeRedII worms.

A.1 Host Infection Rate

We detected more than 359,000 unique IP addresses[4] infected with the Code-RedI worm between midnight UTC on July 19 and midnight UTC on July 20. To determine the rate of host infection, we recorded the time of the first attempt of each infected host to spread the worm. Because our data represent only a sample of all probes sent by infected machines, the number of hosts detected provides a lower bound on the number of hosts that have been compromised at any given time.

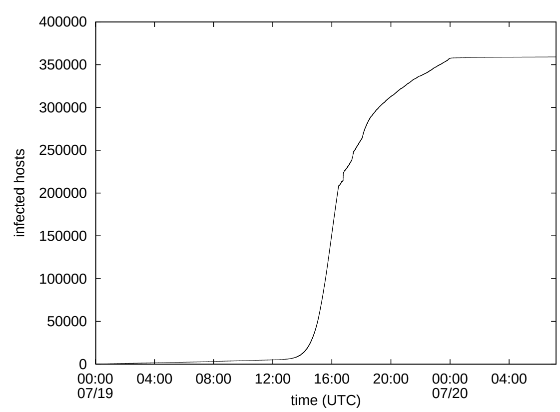


Fig. 2. Cumulative total of unique IP addresses infected by the first outbreak of Code-RedI v2.

[2] IP addresses in the monitored class A network occurred early in each of the 100 threads started on Code-RedI v1 infected machines. Probe sequence numbers within their threads included: 8, 12, 14, 20, 22, 25, 26, 29, 32, 34, 36, 40, 41, 41, 43, 43, 44, 45, 45, 51, 56, 57, 59. Thus we are able to detect the compromise of a new host almost instantly as we receive many probes from the host in the first minute following infection.

[3] The regular expressions are available at http://www.caida.org/tools/measurement/misc/HostClassify

[4] We required at least 2 probes from each host to two different addresses before we conclusively identified it as infected.

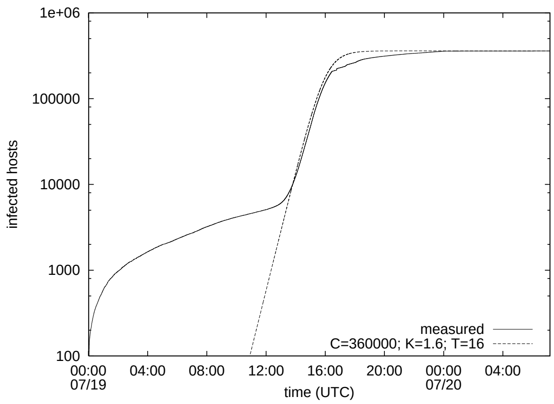


Fig. 3. Comparison of the growth rate of the first outbreak of Code-RedI v2 with infection model.

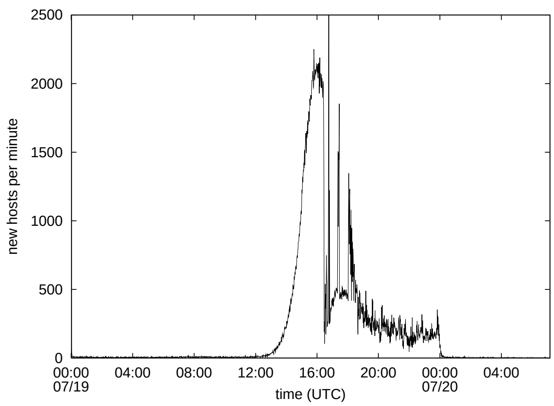


Fig. 4. One minute infection rates for Code-RedI v2.

Figure 2 shows the number of infected hosts over time. The growth of the curve between 11:00 and 16:30 UTC is exponential, as can be seen in the logarithmic scale plot (Figure 3). On the surface, the data seems to fit reasonably with the growth model for the worm infection proposed by Stuart Staniford [11]. Discrepancies between the upper ranges of the growth model and our data are caused both by the fixed cutoff time of the worm itself and by hosts repaired or isolated throughout the day.

Figure 4 provides a more detailed view of the spread of the worm in terms of the number of newly infected hosts seen in 1 minute periods throughout the day. In the figure, we see that the infection rate peaked at 2,000 host/minute. Unfortunately, the peak of the initial curve occurs at about the same time that the passive monitor data became unavailable, so the duration of the 2,000 host/minute infection rate is unknown. In particular, the large spike corresponds to 7,700 hosts; it is an anomaly caused by a small gap in the collected netflow data that resulted in detection of all hosts infected during the down time when collection resumed. Thus the spike in the number of hosts infected is actually representative of all the hosts infected between 16:51 and 17:21 UTC. We believe that in actuality the infection rate from 16:30 to 18:00 UTC tapered smoothly.

Although the growth was slowing, had the worm not been programmed to stop spreading at midnight, additional hosts would have been compromised. The infection rate would have continued to decrease once the vast majority of vulnerable machines were infected. We speculate that the memory resident status of this worm would have allowed reinfection of many hosts after a reboot cleared the initial infection..

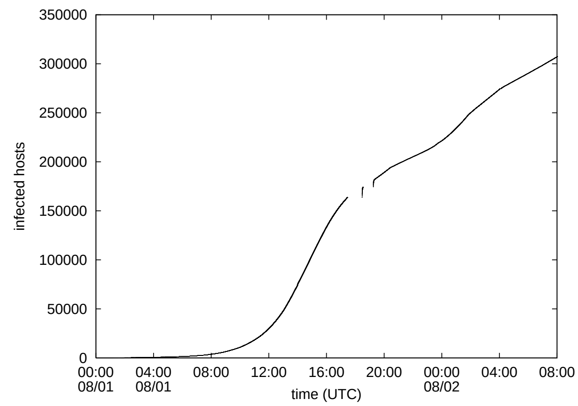


Fig. 5. Cumulative total of unique IP addresses infected during the first day of the second outbreak of Code-RedI v2.

On August 1, the Code-RedI v2 worm began to spread again in earnest. By midnight, we had observed approximately 275,000 unique IP addresses spreading the Code-RedI v2 worm, as seen in Figure 5. The difference between the infected host count at 24 hours for the first and second outbreaks of Code-RedI v2 is likely caused by the patching of hosts, which removed them from the susceptible population.

Figure 6 shows the rate at which new hosts were infected with the Code-RedI v2 worm. The spread of the outbreak peaked in the early afternoon of August 1, with 29710 hosts infected in the hour following 14:00 UTC and 28583 following 15:00 UTC. A rate of more than twenty thousand new hosts per hour was sustained from 13:00 through 17:00 UTC. After this point, the host population approached saturation with the worm – when almost all susceptible hosts are already infected by the worm, it becomes increasingly difficult to locate new hosts.

A.2 Deactivation rate

During the course of the day on July 19, a few initially infected machines were patched, rebooted, or filtered and consequently ceased to probe networks for vulnerable hosts. We consider a host that was previously infected to be inactive after we have observed no further unsolicited

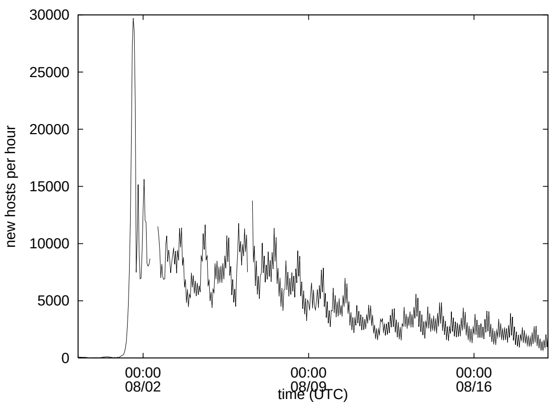


Fig. 6. Hourly infection rates for the second outbreak of Code-RedI v2 between August 1 and August 19, 2001.

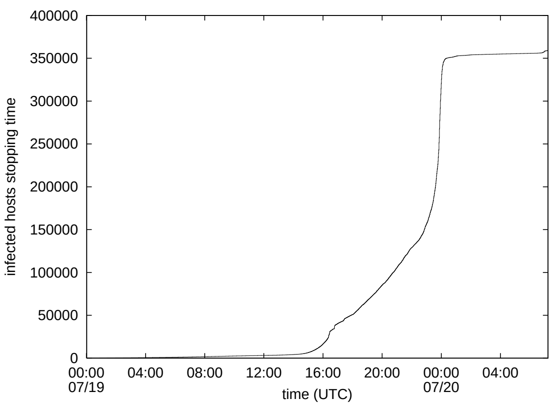


Fig. 7. Cumulative total of deactivated Code-RedI v2. infected hosts.

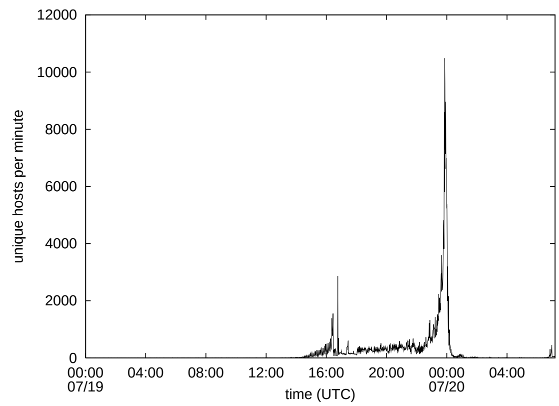


Fig. 8. Rate of infected host deactivation in one minute periods.

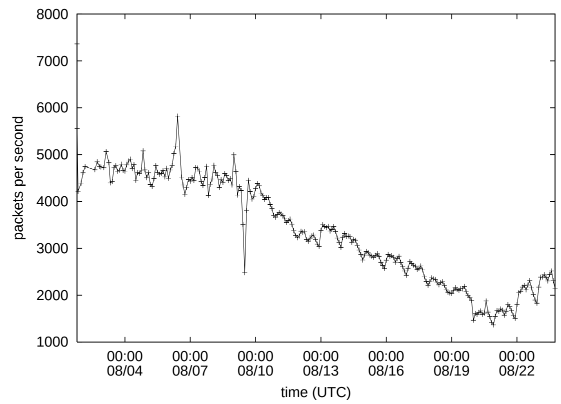


Fig. 9. The raw probe rate observed in our /8 network.

traffic from it. Figure 7 shows the total number of inactive hosts over time. The majority of hosts stopped probing in the last hour before midnight UTC on July 20. At midnight, the worm was programmed to switch from an "infection phase" to an "attack phase", so the large rise in host inactivity is due to this design. The end of day phase change can be seen clearly in Figure 8, which shows the number of newly inactive hosts per minute. As in previous graphs, the spike near 16:30 is caused by a gap in data collection.

A.3 CodeRedII

Because the CodeRedII worm infects the same host population as Code-RedI v2, we neither expected nor measured an increase in the number of hosts probing our network once the CodeRedII worm began to spread. We also monitored no significant difference in the overall number of unsolicited TCP SYNs. Figure 9 shows the raw probe rate (including both worm spread and port scans) into our /8 network for every 2 hours between August 1 and August 22. The spike in probes on August 6 shows backscatter from a Denial-of-Service attack [19], while the dip on August 9 was caused by a gap in data collection. No change in probe rate is apparent following the spread of CodeRedII. Although CodeRedII uses six times as many threads to spread as Code-RedI v2, only one probe in eight is sent to a random IP address, with the rest sent to local networks as described in Section III. Because our /8 network contained no susceptible hosts, the net probe rate we observe from CodeRedII is the same as that of Code-RedI v2. Thus, we cannot distinguish hosts infected with CodeRedII from those infected with Code-RedI v2 without collecting packet payloads. In their October 2001 study, Arbor Networks measured the ratio between Code-RedI and CodeRedII probes to be 1:3 [20]. This 1:3 ratio may indicate the ratio between hosts infected with CodeRedII versus CodeRedI. However, we expect that the bias towards hosts on the same subnet causes wide variations in the actual probe rates measured at different locations across the Internet.

B. Host Characterization

In this section, we look at the properties of the host population infected by the Code-RedI and CodeRedII worms.

Top 10 Countries		
Country	hosts	hosts(%)
United States	157694	43.91
Korea	37948	10.57
China	18141	5.05
Taiwan	15124	4.21
Canada	12469	3.47
United Kingdom	11918	3.32
Germany	11762	3.28
Australia	8587	2.39
Japan	8282	2.31
Netherlands	7771	2.16

TABLE I
TOP TEN COUNTRIES WITH CODE-RED INFECTED HOSTS ON JULY 19.

Top 10 Domains		
Domains	hosts	hosts(%)
Unknown	169584	47.22
home.com	10610	2.95
rr.com	5862	1.63
t-dialin.net	5514	1.54
pacbell.net	3937	1.10
uu.net	3653	1.02
aol.com	3595	1.00
hinet.net	3491	0.97
net.tw	3401	0.95
edu.tw	2942	0.82

TABLE III
TOP TEN DOMAINS WITH CODE-RED INFECTED HOSTS ON JULY 19.

Top 10 Top-Level Domains		
TLD	hosts	hosts(%)
Unknown	169584	47.22
net	67486	18.79
com	51740	14.41
edu	8495	2.37
tw	7150	1.99
jp	4770	1.33
ca	4003	1.11
it	3076	0.86
fr	2677	0.75
nl	2633	0.73

TABLE II
TOP TEN TOP-LEVEL DOMAINS WITH CODE-RED INFECTED HOSTS ON JULY 19.

B.1 Countries

To understand the demography of the Code-RedI v2 epidemic on July 19, we examined the domains, geographic locations, and top level domains (TLDs) of the infected hosts. Table I shows the breakdown of hosts by country, as placed by IxMapping [18]. Surprisingly, Korea is the second most prevalent source country of compromised machines, with 10.57% of all infected hosts.

B.2 Top-Level Domains

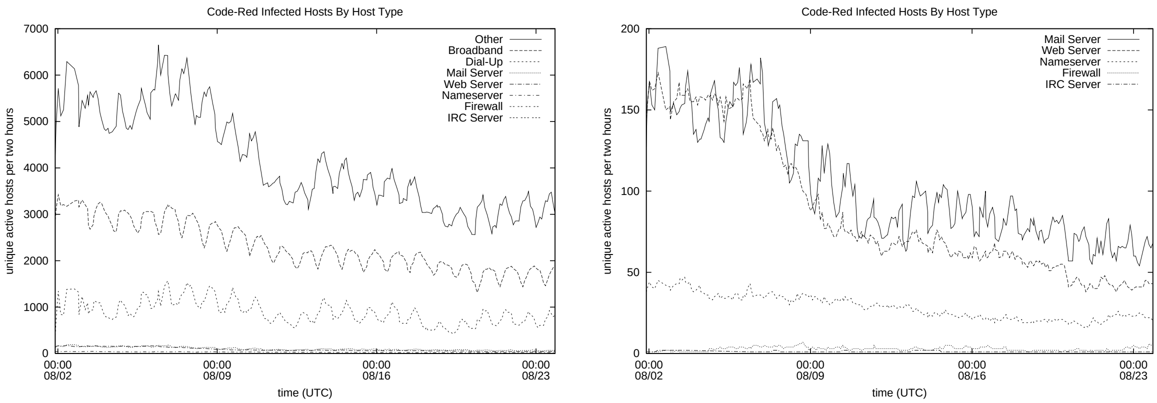
Table II provides a breakdown of machines infected on July 19th by top-level domain (TLD). NET, COM, and EDU are all represented in proportions roughly equivalent to their estimated share of all existing hosts, as estimated by NetSizer [21]. We also observed 136 MIL and 213 GOV hosts infected by the worm. Approximately 50% of all July 19th infected hosts had no reverse DNS records, so they could not be classified by their domain names. These included, for example, 390 addresses in the reserved network space 10.0.0.0/8. These machines were probably on private networks and were infected via either an external interface or another machine accessible via both internal and external networks. This suggests that many more hosts on internal networks may have been compromised in a manner transparent to our monitor.

B.3 Domain Names

Table III shows the top ten domains in terms of the number of infected hosts. We note that the top domain names are providers of home and small business connectivity, suggesting that hosts maintained by individuals at home are an important aspect of global Internet health.

B.4 Host Classification

We utilized the reverse DNS records for the August Code-Red infected hosts to identify the function of the compromised machines. While reverse DNS records did not exist for 55% of the hosts infected in August 2001, we did manage to identify about 22% of the host types. Computers without reverse DNS records are less likely to be running major services (such as those demonstrated in the other host types). Broadband and dial-up services represented the vast majority of identifiable hosts, as shown in Figure 10(a). Furthermore, the large diurnal variations in the number of infected hosts suggest that these machines are unlikely to be running production web servers of any kind, a surprising result given that the worm attacks a vulnerability in web servers. This periodicity con-



(a) All hosts with reverse DNS records.

(b) A closer look at the lower ranges.

Fig. 10. Reverse DNS Record-based classification of Code-Red hosts.

DNS-based host types		
Type	Average Hosts	Hosts(%)
Unknown	88116	54.8
Other	37247	23.1
Broadband	19293	12.0
Dial-Up	14532	9.0
Web	846	0.5
Mail	731	0.5
Nameserver	184	0.1
Firewall	9	0.0
IRC	2	0.0

TABLE IV

THE CLASSIFICATIONS OF HOSTNAMES BASED ON REVERSE-DNS LOOKUPS OF THE IP ADDRESSES OF CODE-RED INFECTED HOSTS BETWEEN AUGUST 1 AND AUGUST 8, 2001. SHOWN HERE ARE THE AVERAGE NUMBER OF ACTIVE HOSTS IN EACH TWO HOUR INTERVAL AND THE OVERALL PERCENTAGE OF EACH TYPE OF HOST ACROSS THE WHOLE SEVEN DAY INTERVAL. UNKNOWN HOSTS HAD NO REVERSE DNS RECORDS.

trasts with the limited diurnal variation seen in the number of infected web and DNS servers in Figure 10(b), which show limited fluctuations from day to day. We do observe significant diurnal changes in the number of infected mail servers, indicating that we may be mis-identifying the function of a number of these computers. Overall, the number of broadband and dial-up users affected by this random-source worm seems to significantly exceed those affected by random-source denial-of-service attacks [19]. While 21% of all hosts compromised by Code-Red were home and small business machines, only 13% of random-source denial-of-service attack targets shared this characteristic. And while web servers, mail servers, and nameservers were the target of 5% of all denial-of-service attacks, they represent only 1.1% of the computers infected by the Code-Red worm.

B.5 Timezones

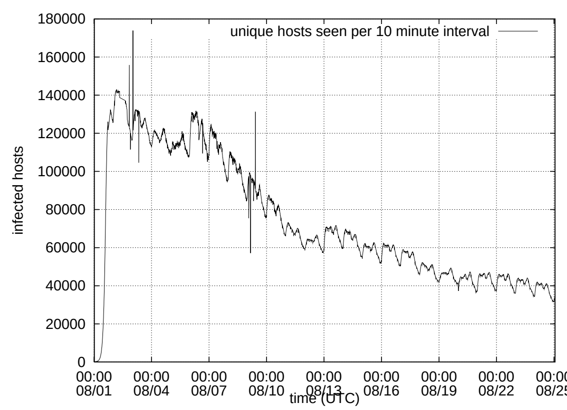


Fig. 11. Unique IP addresses infected with Code-RedI v2 in ten minute intervals.

Figure 11 shows the number of unique IP addresses infected with Code-RedI v2 in ten minute intervals. From the figure, we see that the number of infected hosts follows both diurnal and weekly variations. While the slight decrease in infected hosts on the weekends (Aug 4-5, 11-12, and 18-19) is immediately apparent, the origins of the rather strangely shaped daily variations proved perplexing.

Suspecting that the varying local times of day obscured the infection pattern, we identified the longitude of each infected host via IxMapping and mapped each host to an approximate timezone. We ignored minor variations in

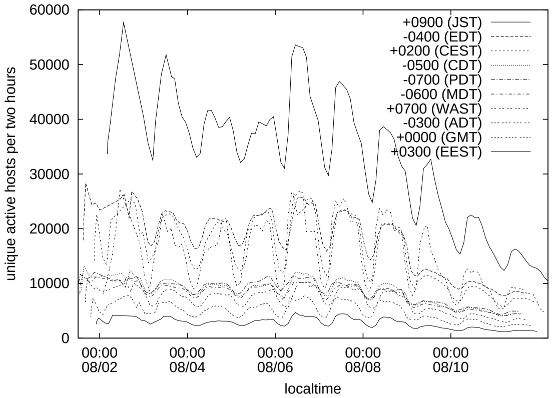


Fig. 12. Unique IP addresses infected with Code-RedI v2 in two hour intervals, localtime.

the longitudinal boundaries of timezones, as well as differences in observation of daylight saving within a timezone.

We then recreated the interval of Figure 11 between August 2 and August 10, differentiating infected hosts according to local time, rather than UTC.

Figure 12 shows the results of this differentiation for the top ten timezones in terms of number of infected hosts. The diurnal pattern clearly follows the pattern of the business day, with the number of infected hosts rising sharply around 8 am and falling off in the afternoon and evening hours as people shut down their computers to go home for the night. The Code-RedI worm attacks a vulnerability in Microsoft web server software, yet production web servers are not usually shut down at the end of the day. We suspect that these machines are office desktop computers whose users are not aware that they are running an active web server. This calls into question both the wisdom and the security of automatically enabling software unbeknownst to the end user.

B.6 Subnets

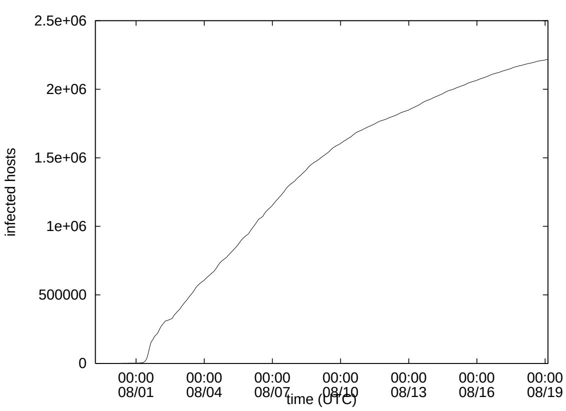


Fig. 13. Cumulative total of unique IP addresses infected with the second outbreak of Code-RedI v2.

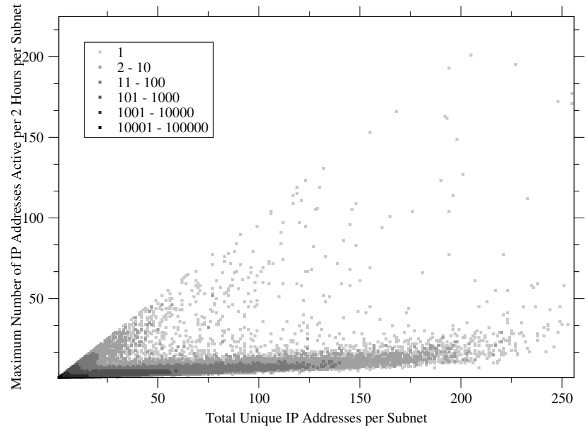


Fig. 14. The DHCP effect: the relationship between total active IP addresses in a subnet and the maximum number of IP addresses active simultaneously, August 2-16.

Between August 2 and August 16, we observed two million IP addresses actively transmitting the worm (Figure 13), yet only 143000 active hosts in the most active ten minute period (Figure 11). This order of magnitude discrepancy leads us to question whether there were actually around two million infected hosts, or whether the use of DHCP is sufficiently extensive that it artificially inflates IP address-based estimates of the extent of the Code-RedI epidemic.

To answer this question, we compared two measures of the infection within a subnet: the number of total unique IP addresses in each subnet active at any time between August 2 and August 16; and the 2 hour period in between August 2 and August 16 in which the greatest number of infected hosts were actively spreading the worm simultaneously. We plotted total unique IP addresses on the X axis, maximum IP addresses per two hour window on the Y axis, and then colored the data points based on the number of subnets with the same X and Y values (Figure 14). The resulting graph is surprisingly bimodal: one line of hosts stretches along the $y = x$ intercept, representing subnets in which the total number infected and the maximum number infected were the same – no shift in the IP addresses of infected machines was detected. A far more populous arm stretches just above the X axis, showing many subnets with as many as fifteen times as many total IP addresses infected as were infected simultaneously. This suggests that without accounting for this "DHCP effect," counting the number of IP addresses infected by a pathogen grossly overestimates the actual number of infected machines.

While the vast majority of the subnets had fewer than fifteen machines infected, a few subnets had as many as two hundred simultaneously infected machines. We inves-

Patch Rate in Top 10 Countries		
Country	patched (%)	unpatched (%)
United Kingdom	65.65	34.34
United States	59.59	40.41
Canada	57.57	42.42
Germany	55.55	44.44
Netherlands	46.46	53.53
Japan	39.39	60.61
Australia	37.37	62.62
Korea	20.20	79.79
Taiwan	15.15	84.84
China	13.13	86.86

TABLE V

PATCHING RATE SEEN ON AUGUST 14TH FOR THE TEN COUNTRIES WITH CODE-RED INFECTED HOSTS ON JULY 19. PERCENTAGES ARE OF INFECTED HOSTS IN EACH COUNTRY THUS EACH ROW ADDS UP TO 100%

tigated the ownership of the subnets with the most infected machines and discovered that they belonged to Microsoft.

While DHCP use may artificially inflate the number of infected hosts as measured by IP addresses, the use of Network Address Translation may artificially deflate the number of compromised IP addresses that we measured. While many infected machines can sit behind a NAT router, it appears to the rest of the Internet as only a single machine. We attempted to use the observed probe rate of a host as a way of identifying NAT IP addresses. However, the wide variation in machine load and network connection speed of individually infected IP addresses masks all but the most blatant evidence of NAT use. Further work on quantifying the effects of NAT on epidemiological study of worm spread is in progress.

C. *Repair rate*

We performed a follow-up survey to determine the extent to which infected machines were patched in response to the Code-RedI worm. Every day between July 24 and August 28, we chose ten thousand hosts at random from the 359,000 hosts infected with Code-RedI on July 19 and probed them to determine the version number and whether a patch had been applied to the system. Using that information, we assessed whether they were still vulnerable to the IIS buffer overflow exploited by Code-RedI.

Although this data does not show the immediate response to Code-RedI, it does characterize the efficacy over time of user response to a known threat. Between July 24 and July 31, the number of patched machines increased an average of 1.5% every day. Despite unprecedented levels of local and national news coverage of the Code-RedI worm and its predicted resurgence on August 1, the response to the known threat was sluggish. Only after Code-RedI began to spread again on August 1 did the percentage of patched machines increase significantly, rising from 32% to 64%.

Improvement is needed in the communication of information about present threats to non-English speaking countries. As shown in Table V, there is a significant gap between the patch rate in English speaking countries and non-English speaking countries.

We observed a wide range in the response to Code-RedI exhibited by the top ten most frequently infected domains, as shown in Table VI. While many of these domains contain IP addresses that are assigned dynamically via DHCP, the percentages of unpatched machines remain valid whether or not the machines we reached in our survey were known previously to be infected. Some ISPs took aggressive action to prevent the spread of the worm, including temporarily blocking both inbound and outbound traffic on port 80 and rapid notification of customers who were observed to be spreading the worm.

The EDU top-level domain exhibited a much better patching response to Code-RedI than did COM or NET – 81% of infected hosts were patched by August 14. COM (56%) and NET (51%) did respond well, ranked third and sixth, respectively.

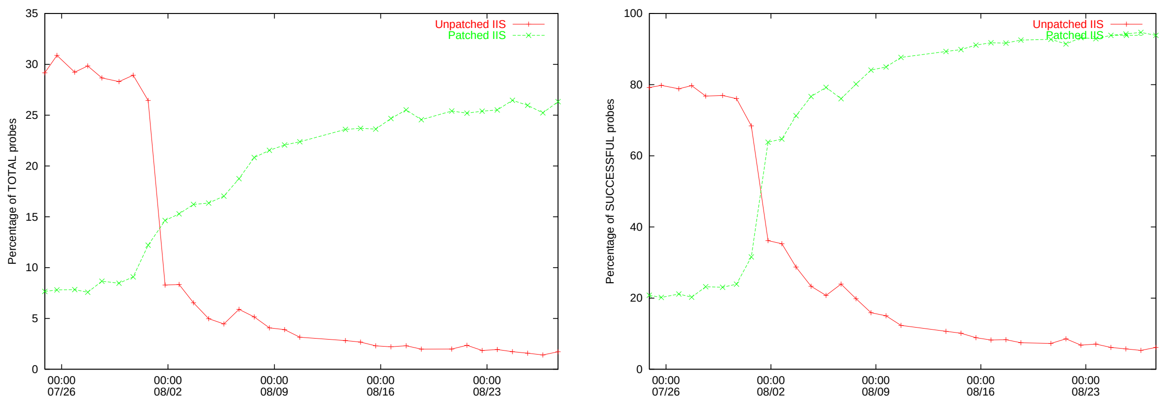
V. CONCLUSION

The primary observation to make about the Code-RedI worm is the speed at which a malicious exploit of a ubiquitous software bug can incapacitate host machines. In particular, physical and geographical boundaries are meaningless in the face of a virulent attack. In less than 14 hours, 359,104 hosts were compromised.

This assault also demonstrates that machines operated by home users or small businesses (hosts less likely to be maintained by a professional systems administrators) are integral to the robustness of the global Internet. As is the case with biologically active pathogens, vulnerable hosts can and do put everyone at risk, regardless of the significance of their role in the population.

Care must be taken in estimating the extent of the spread of Internet pathogens. The effects of DHCP on IP address counts lead to gross overrepresentation of the cumulative number of hosts infected over time. The majority of subnets show a discrepancy between the total number of IP addresses observed to be transmitting the worm and the maximum number active in a two hour period.

Finally, we should all be concerned that it seems to take a global, catastrophic incident to motivate us to respond



(a) All survey attempts. (b) Hosts which responded.

Fig. 15. Patching rate of IIS servers following initial Code-RedI v2 outbreak on July 19th.

Domain	Unpatched IIS (%)	Patched IIS (%)	Conn. Timeout (%)	Conn. Refused (%)
in-addr.arpa	40	7	30	11
home.com	44	5	30	8
rr.com	44	5	27	10
t-dialin.net	0.4	0	81	16
aol.com	0.3	0	39	61
pacbell.net	29	8	24	23
uu.net	0.6	0.2	51	47
hinet.net	20	0	46	25
net.tw	32	1	46	13
edu.tw	60	2	20	5

TABLE VI

PERCENTAGE BREAKDOWN OF PATCHING SURVEY RESPONSES BY CATEGORY FOR THE TOP DOMAINS ORIGINALLY INFECTED WITH CODERED V2. ROWS ADDING TO LESS THAN 100% ARE DUE TO RESPONSES NOT BEING CLEARLY CATEGORIZABLE AS PATCHED IIS OR UNPATCHED IIS. MOST DOMAINS SHOW A LARGE PERCENTAGE OF CONNECTION REFUSED OR CONNECTION TIMEOUT SUGGESTION FILTERING OF TRAFFIC, DISABLING OF PREVIOUSLY RUNNING IIS SERVERS OR DHCP.

to a known threat. The exploit was discovered on June 18, 2001 and the first version of the Code-RedI worm emerged on July 12, 2001. The especially virulent strain of the worm (Code-RedI v2) began to spread on July 19, a full 29 days after the initial discovery of the exploit and four days after the detection of the first (static seed) attack. As the economies of many nations become increasingly dependent on wide area network technologies, we must critically assess and remedy the economic consequences of the current lack of adequate network and host security measures.

VI. ACKNOWLEDGMENTS

We would like to thank Pat Wilson and Brian Kantor of UCSD for data and discussion; Vern Paxson (LBL and ACIRI) for providing an additional view point of data; Jeffrey Mogul and Compaq Research for additional data; Jeff Brown (UCSD/CSE) for producing animations of worm spread; Ken Keys (CAIDA) for development of graphs and discussion; Bill Fenner (AT&T Research) for useful comments and fli2gif; and Stefan Savage (UCSD) and Geoff Voelker (UCSD) for suggestions. Support for this work was provided by DARPA ITO NGI and NMS programs, NSF ANIR, and CAIDA members. This work would not have been possible without the generous support of Cisco Systems.

REFERENCES

[1] E. Spafford, “The internet worm: Crisis and aftermath,” 1989.
[2] Charles Schmidt and Tom Darby, “The

Morris Internet Worm," Tech. Rep., http://www.software.com.pl/newarchive/misc/Worm/darbyt/pages/history.html.

[3] John Shoch and Jon Hupp, "The 'Worm' Programs – Early Experience with a Distributed Computation," *Communications of the ACM*, vol. 25, no. 3, pp. 172–180, Mar. 1982.

[4] CERT Coordination Center, "CERT Advisory CA-1989-04 WANK Worm On SPAN Network," `http://www.cert.org/advisories/CA-1989-04.html`.

[5] Max Vision, "Ramen Internet Worm Analysis," `http://www.whitehats.com/library/worms/ramen/`.

[6] SANS Global Incident Analysis Center, "Lion Worm," `http://www.sans.org/y2k/lion.htm`.

[7] Computer Economics, "2001 economic impact of malicious code attacks," `http://www.computereconomics.com/cei/press/pr92101.html`.

[8] eEye Digital Security, "Advisories and Alerts: AD20010618," `http://www.eeye.com/html/Research/Advisories/AD20010618.html`.

[9] Microsoft, "A Very Real and Present Threat to the Internet," http://www.microsoft.com/technet/treeview/default.asp?url=/technet/security/topics/codealrt.asp.

[10] eEye Digital Security, "Advisories and Alerts: .ida "Code Red" Worm," July 2001, `http://www.eeye.com/html/Research/Advisories/AL20010717.html`.

[11] Silicon Defense, "Code Red Analysis page," `http://www.silicondefense.com/cr/`.

[12] Cisco Systems, Inc, "Cisco Security Advisory: "Code Red" Worm - Customer Impact," `http://www.cisco.com/warp/public/707/cisco-code-red-worm-pub.shtml`.

[13] eEye Digital Security, "CodeRedII Worm Analysis," August 2001, `http://www.eeye.com/html/Research/Advisories/AL20010804.html`.

[14] SecurityFocus, "SecurityFocus Code Red II Information Headquarters," `http://aris.securityfocus.com/alerts/codered2/`.

[15] "Cisco NetFlow," `http://www.cisco.com/warp/public/732/netflow/`.

[16] Vern Paxson, "Bro: A System for Detecting Network Intruders in Real-Time," *Computer Networks*, vol. 31, no. 23-24, pp. 2435–2463, 1999.

[17] eEye Digital Security, "eEye Code Red analysis, commented disassembly, full IDA database, and binary of the worm," `http://www.eeye.com/html/advisories/codered.zip`.

[18] Ixiacom IxMapping, "Ixmapping," `http://www.ipmapper.com`.

[19] David Moore, Geoffrey M. Voelker, and Stefan Savage, "Inferring Internet Denial-of-Service Activity," *Usenix Security Symposium*, 2001.

[20] Dug Song and Rob Malan and Robert Stone, "A Snapshot of Global Internet Worm Activity," http://research.arbor.net/up_media/up_files/ snapshot_worm_activity_f.ps.

[21] Netsizer, "Evaluating the size of the internet," `http://www.netsizer.com`.