Mapping Condor

Douglas Thain, University of Wisconsin
Paradyn/Condor Week, April 2004

Mapping Condor. How can we monitor the impact of a software package on the world at large? To answer this question, we have collected a large amount of data on the deployment of Condor over five years. This handout is just a preview of this fertile data set. Expect to see more results in the future!

Data Sources. Each deployed Condor pool in the world periodically reports back to the University of Wisconsin by sending one email per week and one UDP packet every fifteen minutes. These messages give high level details such as the number of machines in each pool, along with their operating system and CPU type. Of course, we don't receive all messages: some are blocked by firewalls, some are lost due to mis-configured email systems, and some are disabled by users. Email messages have been archived since January 2000, while UDP messages have only been archived since November 2003. The first two pages used combined email and UDP data for a current view, while the last two pages used only email for a historical view.
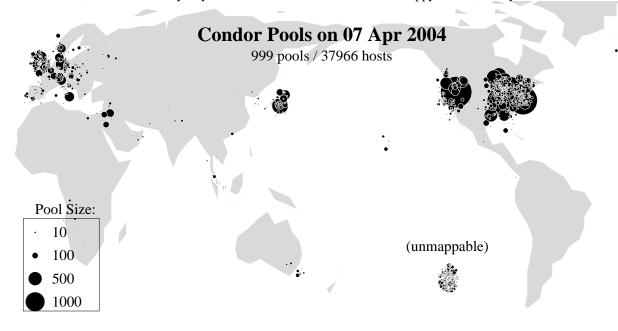
Mapping Technique. Each point on the map is located by way of this public WHOIS database. The domain name of each pool is used to retrieve a WHOIS record, extract a postal address, and then use a public U.S. postal database to obtain the coordinates of the nearest city. Outside the U.S., the top-level country domain name is used to identify the host country, and the point is plotted at the national capital. In both cases, a small random factor is added to provided visual distinction as well as anonymity.

Tidbits of Data
- Before 2003, Condor grew at 48 CPUs/week.
- Since 2003, Condor grew at over 250 CPUSs/week.
- The largest pools have grown: 500-2000 CPUs.
- Five to ten percent of pools exceed 100 CPUs.
- About one third of known pools notified by email only, one third by UDP only, and one third both ways.
- WinNT machines are far less likely to use email.
- 86% of Linux machines gave a valid DNS name.
- 56% of WinNT machines gave a valid DNS name.

Trends. Notice the increase in downloads during 2002 and the increase in deployed machines in 2003. We believe this extraordinary growth worldwide was the result of several factors: the long-awaited release of the 6.4 stable series, the introduction of production support for Windows machines, and the increase in size of large pools.

Privacy. Users are free to decline participation in this data collection technique. Both UDP and email updates can be turned off in the standard Condor configuration file. No information identifying individual users or jobs is submitted to Wisconsin, nor will we reveal details of individual systems, except in anonymized ways such as the figures in this handout. It should be noted that the purpose and role of WHOIS in the Internet and the legal system is a matter of continuing debate. Anyone with concerns about privacy is encouraged to contact us at condor-admin@cs.wisc.edu. We would be happy to work with you.
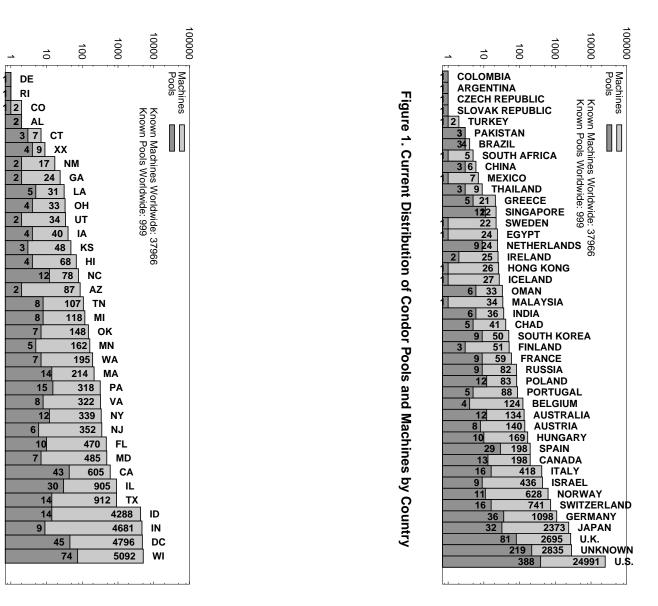

Condor Pools on 07 Apr 2004
999 pools / 37966 hosts

Pool Size:
· 10
• 100
● 500
● 1000

(unmappable)

Figure 1. Current Distribution of Condor Pools and Machines by Country

Machines
Pools

Known Machines Worldwide: 37966
Known Pools Worldwide: 999

Country	Pools	Machines
COLOMBIA	1	
ARGENTINA	1	
CZECH REPUBLIC	1	
SLOVAK REPUBLIC	1	
TURKEY		2
PAKISTAN		3
BRAZIL		34
SOUTH AFRICA	1	5
CHINA	3	6
MEXICO	1	7
THAILAND	3	9
GREECE	5	21
SINGAPORE	1	22
SWEDEN	1	22
EGYPT	1	24
NETHERLANDS	9	24
IRELAND	2	25
HONG KONG	1	26
ICELAND	1	27
OMAN	6	33
MALAYSIA	1	34
INDIA	6	36
CHAD	5	41
SOUTH KOREA	9	50
FINLAND	3	51
FRANCE	9	59
RUSSIA	9	82
POLAND	12	83
PORTUGAL	5	88
BELGIUM	4	124
AUSTRALIA	12	134
AUSTRIA	8	140
HUNGARY	10	169
SPAIN	29	198
CANADA	13	198
ITALY	16	418
ISRAEL	9	436
NORWAY	11	628
SWITZERLAND	16	741
GERMANY	36	1098
JAPAN	32	2373
U.K.	81	2695
UNKNOWN	219	2835
U.S.	388	24991

Figure 2. Current Distribution of Condor Pools and Machines by State

Machines
Pools

Known Machines Worldwide: 37966
Known Pools Worldwide: 999

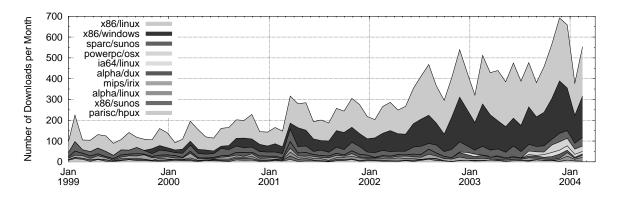| State	Pools	Machines
DE	1	
RI	1	
CO	1	2
AL		2
CT	3	7
XX	4	9
NM	2	17
GA	2	24
LA	5	31
OH	4	33
UT	2	34
IA	4	40
KS	3	48
HI	4	68
NC	12	78
AZ	2	87
TN	8	107
MI	8	118
OK	7	148
MN	5	162
WA	7	195
MA	14	214
PA	15	318
VA	8	322
NY	12	339
NJ	6	352
FL	10	470
MD	7	485
CA	43	605
IL	30	905
TX	14	912
ID	14	4288
IN	9	4681
DC	45	4796
WI	74	5092

Figure 3. Downloads of Condor Software Over Time.

The number of successful downloads of Condor software each month over time. The three dominant platforms are, from the top, x86/linux, x86/windows, and sparc/sunos. Notice the significant increase in downloads in 2002. Downloads are very bursty due to periodic releases of the software.
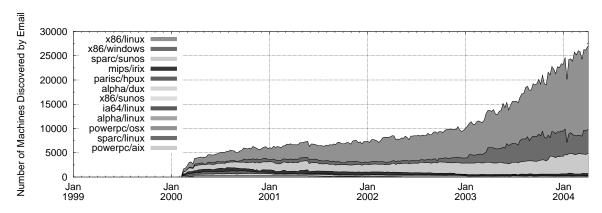

Figure 4. Condor Machines Reported by Email Over Time.

The number of Condor machines reporting to Wisconsin since data collection began in February 2000. The three dominant platforms are, from the top, x86/linux, x86/windows, and sparc/sunos. Notice the rapid expansion beginning in January 2003. *Unlike the first page, this figure only includes email reports, not UDP reports, and thus underestimates the total number.*
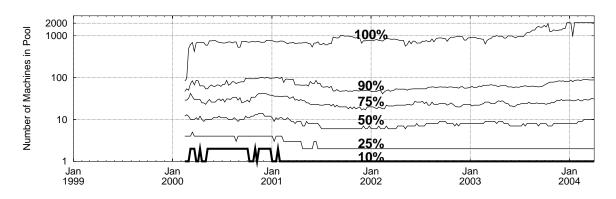

Figure 5. Size of Condor Pools Reported by Email Over Time.

Each line represents a percentile of all Condor pools. For example, 100% line shows the largest pool at any given time while the 50% line shows the median pool size. Notice that about half of all Condor pools are small, less than ten hosts, while the number of large pools has grown steadily since January 2002.

Figure 6. Spread of Condor Over Time.

The left column shows the location of downloads. One dot represents ten downloads. The right column shows each Condor pool reported by e-mail. In both columns, each dot is resolved to the national capital, or, in the United States, to the nearest city. *Unlike the first page, the right column only includes email reports, not UDP reports, and thus underestimates the total number.*

Downloads in 2000 — 2919 total downloads

Condor Pools on 01 Jan 2001 — 352 pools / 5582 hosts

Downloads in 2001 — 4300 total downloads

Condor Pools on 01 Jan 2002 — 329 pools / 7392 hosts

Downloads in 2002 — 6684 total downloads

Condor Pools on 01 Jan 2003 — 397 pools / 10037 hosts

Downloads in 2003 — 8300 total downloads

Condor Pools on 01 Jan 2004 — 548 pools / 24619 hosts

Downloads in 2004 — 2531 total downloads

Condor Pools on 01 Apr 2004 — 522 pools / 26671 hosts

Pool Size: 10 100 500 1000

(unmappable)