The Sysadmin's Daily Grind: Sarg

PEDRO'S ANALYSIS TOOL

A busy proxy server is something that no self-respecting admin should leave to its own devices. The Squid logfile analyzer, dubbed Sarg by its author, helps you keep your Squid servers on track.

BY CHARLY KÜHNAST

really enjoy browsing sites such as Sourceforge or Freshmeat for interesting software packages. Of course,

packages with interesting sounding names are more likely to catch my eye. I couldn't help noticing a tool by the Brazilian software developer, Pedro Orso. I'm sure Pedro wasn't aware of the slightly morbid connotations of his Squid Analysis Report Generator (Sarg) for German Linux users. (Sarg is the German word for coffin.) But this didn't put me off in the least, and it's just as well it didn't, because Sarg is exactly the kind of tool that I like: lean and quick. And it gets on very efficiently with the task for which it was intended: creating reports based on Squid logs.

Sarg source code and binary packages for various Linux distributions, *BSD, MacOS, and even OS/2, are available from [1]. Sarg takes Squid logfiles and uses the data to generate a useful statistical overview, like the overview shown in Figure 1. But in contrast to the Squid add-on Calamaris [2], which we looked at in a previous issue, Sarg generates user-specific statistics.

You can pass the most important parameters to Sarg at the command line, and the *sarg.conf* file (Sarg comes with

SYSADMIN

Admin Workshop 6	2
The Filesystem Hierarchy Standard	
defines the directory structure for Unix	
based sytems.	
DM-Crypt 6	5

DIM-Crypt 6
Hard disk encryption offers an extra measure of protection.

an example) gives you more options, such as modifying the output design. Pedro obviously put a lot of thought into

2	MAG -		-	ert Gene	erator.			
	Squid Uter Access Reported 2004(40) 200							
ALUGNO AFTE	COMMITT			THESE	-	NAMES OF	MILITARY.	-
model of the earlier com-	-	1991	27.79%	100	100,00%	-	299	1 80%
dental fatire safe on		14000	21.00%	10.50%	25 m/s	min si	- None	27 5676
ene sendo un	dea	Age	18719	him	WET	micro to	9000	ENERGY.
CORP. PARTY SET SAL	Land	- 0.0	1117	0.00%	Library	mul b	- Am	6.00%
marked the telephone red	71	MH	1.10%	sien	100,005	DALT	911	3.30%
copile solid on test	584	_ per	589	8.90%	18000	mere	- 84	1 Minu
end graph on the		1.00	1,00%	1.70	20.00	MIN N	- 34	1800
9744	- In		580%	100	38 0075	min no	1044	_83em
CHARLES CONTINUE.	- 16	- 10	1605	11 em	80,30%	B100 (8)	156	3.85
toma maid? wise &	141	401	211%	8,09%	10 TOO	01.41 W	410	TWTN
the telefoliorists of	1 tim	441	2.67%	a any	HEIGHT	MI 10 21	- 000	toms
Reference or a	- 1	401	649	at arm	78 print	-WIEN	den	100%
COMPLETE TOT	400	101	0.07%	127%	30,775	00/17 (m	34	8.36%
exercise condi-			9.96%	25%	XXX	.eex	eth	
codumbination and	100		2112	2,500	Mitto	MARK	- the	1.000
CTC AND USE	100	_8	5.00%	LECT	JERS	5010.00	- 100	3400
eer Jilde der	H		5215	9.30%	96,795	5000.17	594	2.00%
een brouget	784	281	8.00%	140%	11.95	2006.30	304	1000

Figure 1: Sarg creates clear, user-specific reports, keeping you up to date on what Squid is doing.

what most users expect from the Sarg tool and has provided meaningful defaults for most settings. This means that to generate a report, you can simply specify the source file, that is, the Squid *access.log* file, and the target directory where you would like Sarg to put the results.

sarg -1 2
/var/log/squid/access.log 2
-o /www/sarg/

Sarg and DNS

For more convenience, Sarg has a -n command line option that enables DNS resolution of addresses. This is fine for a small Squid with just a few users, but if you have a large cache that processes billions of requests a day, you will not want to enable Sarg name resolution

because the analysis could take all day. Apart from this, most DNS admins would not be too pleased about the involuntary stress test this puts their servers through.

The ability to restrict analysis to a specific period of time, using the -d TT/MM/YYYY-TT/MM/YYYY option, is another useful feature. Time has always been an issue with Squid, which stores time in seconds past the eon with a resolution of

one thousandth of a second in its *access.log* file. Although you can stop Squid from doing this by telling the tool to use the legacy common logfile format, you do lose some information in the process. Sarg is a big help here. Entering

sarg -convert 2
/var/log/squid/access.log

will output the logfile on STDOUT – and it gives you a readable date format. .

A value such as 1126705707.537 is thus converted to 09/14/2005 15:48:27. And losing the thousandths doesn't faze me in the least

Thanks for the app, Pedro, I haven't had this much fun in ages, but you should rethink the acronym for the benefit of all those German Linux hackers.

INFO

[1] Sarg

http://sarg.sourceforge.net/sarg.php

[2] Charly Künast, "The Sysadmin's Daily Grind: Calamaris," Linux Magazine 12/03, p. 60.

THE AUTHOR

Charly Kühnast is a Unix System Manager at the data-cen ter in Moers, near Germany's famous River Rhine. His tasks include ensuring firewall security



and availability and taking care of