Analiza plików rejestrowych w systemie GNU/Linux

Grzegorz Jacek Nalepa

20.02.2000, Kraków, *Revision* : 1.5

Streszczenie

Artykuł porusza tematykę związaną z analizowaniem plików rejestrowych. Zawiera przegląd ogólnie dostępnych w sieci Internet narzędzi służących do szeroko pojętego analizowania i przetwarzania zawartości plików rejestrowych – podstawowego źródła danych o pracy systemów GNU/Linux czy Unix.

Spis treści

[1]Tekst ukazał się w: *Magazynie Netforum*, nr 4/2000, wydawanym przez Lupus.

[2]Kontakt z autorem: mail:gjn@agh.edu.pl

[3]Tytuł angielski: *Analyzing logfiles in GNU/Linux system environment*

1. Wstęp

Monitorowanie współczesnego systemu komputerowego jest zadaniem złożonym. W literaturze fachowej można znaleźć informacje na temat konfigurowania oprogramowania umożliwiającego gromadzenie danych dotyczących pracy systemu. Nieco rzadziej pisze się natomiast o tym, co z tym ogromem danych robić. Trzeba bowiem pamiętać, że dla administratora systemu ważne są przecież nie tyle same *dane*, ile *informacje*, jakie można z nich uzyskać.

2. Pliki rejestrowe

Programy monitorujące pracę systemu zapisują gromadzone dane w tak zwanych plikach rejestrowych, nazywanych potocznie „logami" (ang. *log files*).

Najważniejszym oprogramowaniem monitorującym system jest Syslog, będący narzędziem zcentralizowanym, mogącym rejestrować wiele kategorii zdarzeń. W praktyce dowolna aplikacja pracująca w środowisku systemu GNU/Linux może być źródłem komunikatów przesyłanych do Syslog, dotyczy to szczególnie tak zwanych demonów systemowych. Ponieważ możliwości grupowania komunikatów przez Syslog są w pewnym stopniu ograniczone, poszczególne programy często używają własnych mechanizmów zbierania danych o własnej pracy i zapisywania ich we własnych plikach rejestrowych. Jest tak chociażby w przypadku serwerów WWW czy FTP.

2.1. Struktura plików rejestrowych

Pliki rejestrowe tworzone przez różne programy mają najczęściej znormalizowaną postać. Ich strukturę można opisać następująco:

- są plikami tekstowymi,
- każda linia zawiera osobny komunikat,
- komunikat jest opatrzony informacją o dacie jego wysłania,
- każdemu komunikatowi najczęściej towarzyszy informacja o jego kategorii i programie-nadawcy, jest tak przynajmniej w przypadku plików rejestrowych Syslog.

To, że pliki rejestrowe mają postać tekstową znacząco upraszcza ich analizowanie przy pomocy dowolnych narzędzi i na dowolnej platformie.

2.2. Zagadnienia

Z korzystaniem z plików rejestrowych wiążą się następujące zagadnienia:

- Sortowanie komunikatów – jest najczęściej realizowane przez Syslog, lub konkretną aplikację, na przykład Apache.
- Rotacja plików – czyli dzielenie ich na chronologicznie posortowane fragmenty i ich archiwizacja. To zadanie jest przeważnie wykonywane przez specjalistyczne oprogramowanie, na przykład program LogRotate, współpracujące z systemem Cron.
- Bezpieczeństwo danych – analizując dane z plików rejestrowych należy się upewnić, że nie zostały one zmodyfikowane w nieuprawniony sposób. Z bezpieczeństwem wiąże się spójność danych, której kontrolą zajmują się pakiety takie jak Tripwire, oraz bezpieczne przesyłanie komunikatów przez sieć, na przykład z wykorzystaniem Ssh.

- Analiza i wizualizacja informacji – z danych zgromadzonych w plikach rejestrowych można uzyskać informacje o działaniu systemu. Nawet w niewielkich systemach dobre przeprowadzanie tej analizy jest trudne jedynie w oparciu o ręczne przeglądanie zawartości plików rejestrowych. Stworzono specjalne narzędzia, ułatwiające proces analizy i wizualizację jej wyników.
- Śledzenie zmian – ponieważ pliki rejestrowe są modyfikowane na bieżące w trakcie pracy systemu, administratorzy spędzający dużo czasu przy konsolach systemów mogą wykorzystywać narzędzia ułatwiające obserwowanie danych dopisywanych do plików rejestrowych.

Informacje dotyczące sortowania komunikatów na etapie ich otrzymywania, oraz dotyczące rotacji plików rejestrowych można znaleźć w literaturze na temat Syslog. Zalecenia związane z bezpieczeństwem i spójnością danych są również często opisywane i dyskutowane na łamach prasy fachowej. Warto natomiast skupić się na śledzeniu zmian w plikach rejestrowych, oraz ich analizie i wizualizacji.

3. Wizualizacja i śledzenie zmian

Złożone analizy plików rejestrowych wykonuje się najczęściej cyklicznie, po zebraniu odpowiedniej ilości danych. Są jednak sytuacje kiedy administrator chce mieć możliwość monitorowania pracy systemu na bieżąco. Wtedy wykorzystywane są narzędzia wyświetlające dane dopisywane do plików rejestrowych.

Najprostszą metodą jest oczywiście użycie polecenia `tail -f`, dublowanie informacji zapisywanych przez Syslog na konsolę, lub użycie programu `xconsole` w środowisku X Window.

Aby ułatwić przeglądanie plików rejestrowych warto użyć narzędzi pozwalających na wyróżnianie fragmentów plików przy pomocy różnych kolorów. Mają one najczęściej postać skryptów w języku Perl.

Przykładem takiego prostego skryptu jest LogColorise. Umożliwia on wyróżnienie linii zawierających podany łańcuch znaków przy pomocy jednego z ośmiu kolorów, oraz jej podkreślenie i rozjaśnienie. Pozwala na kolorowanie nie tylko całych linii, lecz również ich fragmentów.

Jeżeli pracuje się w środowisku X Window warto używać program XLogMaster do obserwowania zmian w plikach rejestrowych. Jest napisany z wykorzystaniem biblioteki Gtk+ i umożliwia równoczesne monitorowanie zmian w wielu plikach rejestrowych.

Każdy plik może być przeglądany w osobnym oknie. Dla każdego z monitorowanych plików możliwe jest zdefiniowanie reguł, w postaci wyrażeń regularnych, określających zachowanie programu w przypadku pojawienia się podanego komunikatu. Linie zawierające wyrażenie mogą zostać wyróżnione odpowiednimi kolorami, oprócz tego program może tworzyć okienka informacyjne, czy wykonywać zewnętrzne polecenia.

Program jest bardzo prosty w obsłudze i ma duże możliwości konfiguracji. Umożliwia przede wszystkim szybkie tworzenie nowych okien dla różnych plików rejestrowych, wraz z odpowiednimi filtrami.

4. Analiza plików

Klasycznym przykładem programu, analizującego dowolne pliki rejestrowe jest Swatch. Pozwala na przeglądanie plików rejestrowych i wyszukiwanie w nich odpowiednich wzorców. Duże możliwości konfiguracji programu pozwalają na grupowanie powtarzających się komunikatów, kolorowanie fragmentów plików, wysyłanie informacji pocztą elektroniczną i uruchamianie odpowiednich programów. Program jest napisany w języku Perl i pracuje w trybie tekstowym. Może być również wykorzystywany do śledzenia na bieżąco zmian w plikach.

Bardziej zaawansowanym narzędziem jest LogSurfer. Ma praktycznie wszystkie funkcje Swatcha. Oprócz tego umożliwia kontekstowe filtrowanie plików, gdzie przez kontekst rozumie się odpowiednie sekwencje linii (komunikatów). Pozwala na dynamiczne dodawanie i usuwanie reguł filtrowania. Reguły, których używa, są również bardziej zaawansowane od tych, których używa Swatch.

4.1. Pliki serwerów FTP

Osobną grupę stanowią programy analizujące dane o transferach z serwera FTP. Warto wspomnieć przynajmniej o trzech. Wszystkie są napisane w języku Perl, pracują jako skrypty CGI i wykonują wizualizację statystyk FTP w postaci plików html – stron WWW.

FtpLog_CGI jest prostym narzędziem analizującym pliki popularnego serwera Wu-Ftpd. Wyświetla podstawowe informacje dotyczące najczęściej kopiowanych plików, posortowane według podstawowych kryteriów.

Bardziej zaawansowanym narzędziem jest FtpLoger, przystosowany do pracy z serwerem Wu-Ftpd. Pozwala na wyświetlanie szczegółowych statystyk, również dotyczących pojedynczych plików.

Najbardziej rozbudowany jest FtpWebLog. Ma wszystkie funkcje powyższych programów lecz oprócz tego umożliwia wizualizację wyników w postaci tekstu i kolorowych wykresów. Podaje rankingi dotyczące kopiowanych plików i domen, których dotyczył transfer.

4.2. Pliki Sendmaila

Prostym programem umożliwiającym tworzenie statystyk dotyczących poczty elektronicznej jest Mreport. Pracuje z plikami programu Sendmail, umożliwiając sortowanie informacji według rozmiaru, liczby, lub adresów nadawców i odbiorców wiadomości.

Bardziej zaawansowany jest program Anteater. Oprócz wymienionych powyżej funkcji umożliwia między innymi sortowanie według domen oraz obsługę plików aliasów pocztowych.

4.3. Pliki serwerów WWW

Ponieważ WWW jest jednym z najważniejszych serwisów, ten rodzaj plików rejestrowych jest najczęściej analizowany. Ten temat jest często omawiany na łamach prasy fachowej i wykracza poza ramy artykułu, tu jest miejsce tylko na wymienienie najważniejszych programów. Należą do nich: Analog, Http-Analyze, Webalizer.

W tabeli są również wspomniane programy analizujące pliki programu Squid.

5. Podsumowanie

Gromadzenie danych w plikach rejestrowych nie ma sensu, jeżeli dane te nie są analizowane. Jak widać istnieje szereg programów, które są w tym pomocne. Ich umiejętne wykorzystanie pozwala na uzyskiwanie informacji, które umożliwiają optymalne skonfigurowanie systemu, czy podniesienie jego bezpieczeństwa. W przypadku statystyk z serwerów FTP czy WWW takie dane mogą również mieć znaczenie marketingowe.

Nazwa	Typ	Adres i Możliwości	Licencja
Analog	analiza (WWW)	http://www.statslab.cam.ac.uk/~sret1/analog • zaawansowane możliwości, • wizualizacja z użyciem wykresów graficznych, • pracuje na wielu platformach.	open source
Colorlogs	wizualizacja	http://www.resentment.org/projects/colorlogs • kolorowanie fragmentów dowolnych plików, • praca jako filtr.	open source
Anteater	analiza (Sendmail)	http://www.profzone.ch/anteater • generowanie zaawansowanych statystyk	GNU GPL
Calamaris	analiza (Squid)	http://calamaris.cord.de statystyki dotyczące: • wydajności, • transferów, • użytkowników, • domen.	GNU GPL
FtpLog_cgi	analiza (FTP)	http://www.curtisonline.net/software/ • pracuje z Wu-Ftpd, • proste statystyki.	open source
FtpLoger	analiza (FTP)	http://www.geocities.com/SiliconValley/8236/unix.html • przystosowany do pracy z Wu-Ftpd, • różne statystyki.	GNU GPL
FtpWebLog	analiza (FTP)	http://www.nihongo.org/snowhare/utilities/ftpweblog • bardzo różnorodne statystyki, • wizualizacja w postaci wykresów graficznych.	open source
Http-Analyze	analiza (WWW)	http://www.netstore.de/Supply/http-analyze • bardzo szybki, • uproszczone statystyki.	open source

LogColorise	wizualizacja	http://www.linuxsupportline.com/~pgp/linux • kolorowanie fragmentów dowolnych plików, • duże możliwości kolorowania, • praca jako filtr.	GNU GPL
LogSurfer	analiza, monitorowanie, wizualizacja	http://www.cert.dfn.de/eng/logsurf • praca w trybie tekstowym, • bardzo zaawansowane możliwości konfiguracji, • pracuje z dowolnymi plikami.	open source
MReport	analiza (Sendmail)	ftp://ftp.datrix.co.za/pub/mreport • proste sortowanie komunikatów	GNU GPL
SquidSites	analiza (Squid)	http://www.cineca.it/~nico/squidclients.html • informacje o najczęściej odwiedzanych stronach	GNU LGPL
SquidTaild	analiza (Squid)	http://trailer.linuxatwork.at • zaawansowane filtrowanie, • sortowanie według URL, • sortowanie według użytkowników.	GNU GPL
Swatch	analiza, monitorowanie, wizualizacja	ftp://sierra.stanford.edu/swatch • praca w trybie tekstowym, • duże możliwości konfiguracji, • pracuje z dowolnymi plikami.	GNU GPL
Webalizer	analiza (WWW)	http://www.mrunix.net/webalizer • zaawansowane możliwości, • wizualizacja z użyciem wykresów graficznych.	GNU GPL
XLogMaster	monitorowanie	http://www.gnu.org/software/xlogmaster • praca w środowisku X Window z wykorzystaniem Gtk+. • monitorowanie dowolnych plików, • duże możliwości konfiguracji,	GNU GPL

Tablica 1: Narzędzia do przetwarzania plików rejestrowych.