dr inż. Marian Wrzesień
Przemysłowy Instytut Automatyki i Pomiarów

WEKTOROWA ARCHIWIZACJA DANYCH WYKORZYSTUJĄCA PRZYROSTOWE KOPIE BEZPIECZEŃSTWA

Zaprezentowano implementację systemu wektorowej archiwizacji danych informatycznych generowanych w serwerach pracujących pod systemami operacyjnymi Windows. Jako miejsce docelowe archiwizowanych zasobów zastosowano serwer pracujący pod systemem operacyjnym Linux Fedora Core 9. System archiwizacji zapewnia transfer przyrostowych kopii bezpieczeństwa tworzonych codzienne w lokalnych serwerach aplikacji do serwera docelowego, w których jest stosowana metoda archiwizacji wektorowej transferowanych zasobów. Dla zapewnienia właściwej integracji systemów informatycznych pracujących w różnych platformach software-owych zastosowano oprogramowanie samba posadowione w serwerze Linux. Transfer pomiędzy serwerami oraz synchronizacja przeznaczonych do archiwizowania zasobów są sterowane z wykorzystaniem oprogramowania SyncBack, rsync, rsnapshot oraz cron. Podkreślono efekty wybranych metod archiwizacji minimalizującej czasy trwania transferów danych oraz minimalizujących przestrzeń mediów przeznaczonych do przechowywania zarchiwizowanych zasobów informatycznych.

VECTOR DATA ARCHIVING BASED ON THE INCREMENTAL BACKUP COPIES

The implementation of the vector archiving of the information data created by servers based on the Windows operating system is presented. As the target destination for the archived resources the Linux Fedora Core 9 server is used. The archiving system provides the transfer of the incremental backup copy formed everyday by the local application servers to the destination server, where the vector archiving method of the transferred resources is applied. For the assurance an appropriate integration of the information systems which are working under various software platforms, the samba software implemented on the Linux server is used. The transfer between servers as well as the synchronization of the information resources intended to be archived are controlled by such software as SyncBack, rsync, rsnapshot and cron. It is emphasized, that the indicated method leads to the minimization of the data transfer duration as well as the minimization of the space that is required for the archived data in the store medium.

1. WSTĘP

Obecnie wiele organizacji wdraża System Zarządzania Bezpieczeństwem Informacji (SZBI) (ang. ISMS Information Security Management System). W Polsce SZBI stał się popularny zwłaszcza od czasu opublikowania polskich norm (styczeń 2007 r.): **PN-ISO/IEC 27001** ujmującej zagadnienia objęte SZBI oraz **PN-ISO/IEC 17799** normującej Praktyczne Zasady Zarządzania Bezpieczeństwem Informacji. Ze względu na łatwość integracji SZBI z innymi normami dotyczącymi zarządzania, jeden odpowiednio zaprojektowany zintegrowany system zarządzania może spełnić wymagania wszystkich norm tej klasy.

Najważniejsze cechy SZBI to:

- Poufność (ang. **C**ONFIDENTIALITY)
  - właściwość polegająca na tym, że informacja nie jest udostępniana lub wyjawiana nieupoważnionym osobom, podmiotom lub procesom
- Integralność (ang. **I**NTEGRALITY)
  - właściwość polegająca na zapewnieniu dokładności i kompletności aktywów
- Dostępność (ang. **A**CCESSIBILITY)
  - właściwość bycia dostępnym i użytecznym na żądanie upoważnionego podmiotu

oraz towarzyszące im:

- Autentyczność (ang. **A**UTHENTICITY)
  - właściwość polegająca na potwierdzeniu autorstwa informacji
- Rozliczalność (ang. **A**CCOUNTABILITY)
  - właściwość polegająca na uzasadnieniu korzystania z dostępu do informacji
- Niezaprzeczalność (ang. **I**NCONTROVERTIBILITY)
  - właściwość polegająca na uzyskaniu dowodu wystąpienia lub nie wystąpienia zdarzenia lub działania
- Niezawodność (ang. **R**ELIABILITY)
  - właściwość polegająca na wiarygodności aktywów

Niezależnie od formalnego wdrożenia SZBI, często stosuje się zalecenia SZBI intuicyjnie, a jednym z priorytetowych działań pozwalających zwiększyć bezpieczeństwo informacji określone ww. cechami systemu jest wykonywanie kopii bezpieczeństwa (ang. backup copy) i/lub archiwizowania zasobów informatycznych. Zapewnienie bezpieczeństwa informacji w sposób istotny przyczynia się do uzyskania właściwego wizerunku organizacji, w tym jej konkurencyjnej pozycji rynkowej, płynności finansowej i zyskowności, oraz postrzegania jej jako działającej w zgodzie z lokalnymi wymogami prawnymi tym samym godnej zaufania.

2. POLITYKA ARCHIWIZACJI

Koncentrując się na działaniach prowadzących do archiwizacji należy zdefiniować politykę tego postępowania, której najistotniejszymi elementami są:

- Wybór informacji, które powinny być archiwizowane oraz oszacowanie środków finansowych przeznaczonych na bezpieczeństwo informacji
- Właściwy dobór odpowiedniego terminu i cyklu archiwizacji;
- Odpowiedni wybór nośnika oraz oprogramowania służącego do archiwizacji;
- Właściwy wybór metody przeprowadzania archiwizacji oraz liczby kopii;
- Ustalenie czasu i sposobu przechowywania archiwów;
- Właściwy dobór metody w celu zapewnienia poufności i integralności archiwizowanych danych;
- Czas trwania transferu oraz czas trwania tworzenia archiwizacji wektorowej

2.1. Wybór archiwizowanych zasobów i oszacowanie nakładów na bezpieczeństwo

O zakresie zasobów poddanych archiwizacji decyduje szacowanie ryzyka dotyczącego organizacji, przy uwzględnieniu całościowej strategii biznesowej i ocena zagrożeń oraz podatności zagrożeń dla aktywów. Ten ogólny pogląd implikuje stwierdzenie, że nakłady na bezpieczeństwo powinny być porównywalne z ewentualnymi stratami powodowanymi potencjalną utratą zasobów informatycznych. To w następstwie jest kluczem do przygotowania listy zasobów informatycznych objętych koniecznością archiwizowania,

i wreszcie wyznaczenie serwerów, a następnie posadowionych na nich katalogów i plików wyselekcjonowanych jako objętych backup-em. W PIAP archiwizuje się dane stanowiące elementy zasobów informatycznych:

- Serwera Windows Serwer 2003,
- Serwera Windows Serwer 2003 SI,
- Serwera Windows Serwer 2008, MS Project,
- Serwera oprogramowań sieciowych,
- Serwera Linux Fedora Core 9

2.2. Dobór czasu i cyklu archiwizacji

Uwzględniając częstość zmian źródeł danych wyznaczonych do archiwizowania oraz przydatność uprzednio zarchiwizowanych danych, w PIAP zastosowano podejście polegające na archiwizowaniu codziennym, przechowywaniu zestawu kopii z ostatnich 30 dni wszystkich kopii codziennych oraz zachowywaniu kopii codziennej każdego z katalogów prywatnych użytkowników z prawem dostępu do nich przez ich właścicieli. Moment archiwizacji wypada w czasie, gdy z dużym prawdopodobieństwem żaden z użytkowników sieci informatycznej nie pracuje.

2.3. Wybór metody przeprowadzania archiwizacji

Spośród trzech metod archiwizacji:
archiwizowanie pełne, podczas którego tworzone są pełne kopie zabezpieczanych zasobów (pliki, katalogi) podczas kolejnych sesji archiwizacyjnych.
archiwizowanie przyrostowe, podczas którego tworzone są kopie tylko tych elementów zabezpieczanych zasobów, które zostały zmienione od czasu ostatniej sesji archiwizacyjnej.
archiwizowanie różnicowe, podczas którego tworzone są kopie tylko tych elementów zabezpieczanych zasobów, które zostały zmienione od czasu utworzenia pełnej kopii podczas pierwszej sesji archiwizacyjnej,
w PIAP wybrano – w odniesieniu do katalogów - **archiwizowanie przyrostowe**, co jest podyktowane wygodą przy korzystaniu z zachowanych kopii oraz wydajnością wykorzystania przeznaczonych na kopie bezpieczeństwa mediów. Istnieje ponadto sposób wektoryzowania przyrostowych zapisów dziennych, nazwane tu **archiwizowaniem wektorowym**. Jak to opisano poniżej, zastosowana metoda umożliwia w każdym kolejnym zapisie archiwizacyjnym prezentację pełnego zbioru danych źródłowych tworzonych z danych ostatnio zmodyfikowanych oraz odniesień do pozostałych danych zarchiwizowanych pierwotnych bez ich kopiowania, co oszczędza miejsce na nośniku i daje wrażenie PEŁNOŚCI

2.4. Wybór nośnika oraz oprogramowania służącego do archiwizacji

Spośród takich nośników jak: dyskietki FD, compact-disc CD, DVD, BluRay-DVD, HD-DVD, Streamer, Dyski Twarde wybrano te ostatnie, jednak w szczególnej konfiguracji serwerowej. Zastosowano tu bowiem macierz dyskową RAID 5 + dysk hot-spare (dysk zapasowy). Dodatkowo zastosowano system hot-swap (wymiana dysku „w biegu"), umożliwiający modyfikację zestawu dysków bez wyłączania serwera. Tak więc poza kontrolą parzystości ECC charakteryzującą RAID 5, zastosowanie dysku hot-spare zwiększa dopuszczalną liczbę jednocześnie uszkodzonych dysków do 2, z zachowaniem funkcjonalności systemu przechowującego dane. Zastosowanie kontroli parzystości w serwerze oraz umieszczenie serwera w pomieszczeniu chronionym spełnia wymóg integralności i poufności archiwizowanych danych.

Jako system operacyjny OS serwera zastosowano Linux Fedora Core 9. Ten OS jest wyposażony w narzędzia dedykowane backupowi.

2.4.1. Usytuowanie nośnika służącego do archiwizacji

Archiwizowanie nie powinno być realizowane na serwerze będącym źródłem danych. Z tego względu backup dzienny jest transferowany do serwera docelowego, który przetwarza dane do formatu backupu długoterminowego. Serwer służący jako docelowe miejsce przechowywania archiwizowanych danych usytuowano w PIAP w miejscu znacznie odległym od serwera będącego źródłem archiwizowanych danych. Stąd też, w procesie archiwizacji uwzględniono komunikację sieciową w relacjach:

- Serwer źródłowy ⇔ serwer backupowy,
- Serwer backupowy ⇔ użytkownicy zarchiwizowanych zasobów

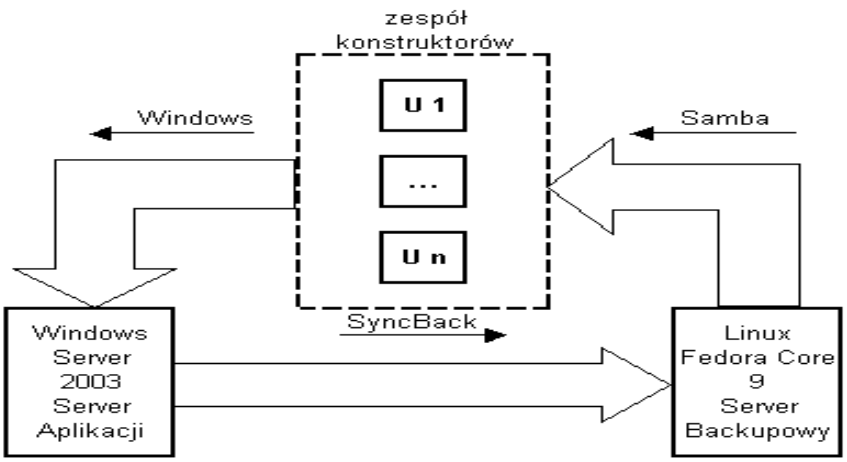
2.4.2. Oprogramowanie stosowane podczas archiwizacji

Wykonywanie bazowych, dziennych, przyrostowych kopii bezpieczeństwa realizują programy posadowione na kilku serwerach objętych archiwizacją. Te kopie – zestawiane zwykle w określonych katalogach – podlegają następnie właściwemu przetworzeniu, w wyniku którego uzyskuje się **archiwizację wektorową**.

3. System archiwizacji wektorowej na przykładzie serwera CAD ProEngineer

3.1. Serwer CAD ProEngineer

Oprogramowanie ProEngineer zostało posadowione na serwerze **ServerZsm** z zaimplementowanym systemem operacyjnym Windows Server 2003. Służy ono do wspomagania projektowania podzespołów wyrafinowanych konstrukcji elektro-mechanicznych. Realizowane projekty wymagają bezpiecznego przechowywania opracowań oraz starannej archiwizacji. Jest również spełnione wymaganie łatwego, autoryzowanego dostępu do przechowywanych danych w przypadku ewentualnego odtwarzania kopii zasobów. Te wymogi zainspirowały opracowanie systemu archiwizacji wektorowej zasobów informatycznych generowanych podczas korzystania z oprogramowania ProEngineer. Rys. 1 pokazuje środowisko informatyczne pracy konstruktorów, miejsce tworzenia opracowań oraz przepływ archiwizowanych danych.



Rys. 1. Wytwarzanie i obieg informacji w systemie CAD ProEngineer

3.2. Metodyka Systemu archiwizacji serwera SerwerZsm

Backup zasobów serwera **ServerZsm** składa się z trzech faz:

- Backup lokalny na serwerze **SerwerZsm**,
- Transfer Codziennej kopii bezpieczeństwa do serwera Linuxowego,
- Archiwizacja wektorowa na serwerze Linuxowym

Istotnym elementem systemu archiwizacji jest także

- Zarządzanie dostępem do zarchiwizowanych danych

3.2.1. Archiwizacja lokalna

Backup lokalny jest realizowany przez podsystem archiwizacji stanowiący element aplikacji pakietu oprogramowania ProEngineer zaimplementowanego w serwerze **SerwerZsm**. Podczas procesu archiwizacji lokalnej, przeprowadzany jest backup codzienny, w którego wyniku tworzona jest bazowa, dzienna, przyrostowa kopia bezpieczeństwa. Jest ona lokowana na serwerze macierzystym w wybranym katalogu, w tym przypadku o ścieżce dostępu **D:\backupPro\wtbackup.** Ten katalog jest przewidziany do archiwizowania. Zawiera on projekty opracowane przez konstruktorów – użytkowników pakietu oprogramowania ProEngineer. Ponadto, do archiwizowania są przeznaczone inne określone katalogi robocze uzupełniane na bieżąco przez – będących ich właścicielami - użytkowników aplikacji ProEngineer, o ścieżkach dostępu **D:\pub\katalog1… D:\pub\katalogN**, oraz o ścieżkach dostępu **D:\home\użytkownik1… D:\home\użytkownikN**. Ważność wymienionych katalogów trwa 1 dzień. Kolejna kopia bezpieczeństwa na serwerze macierzystym zawiera nowy zbiór aktualnych plików i katalogów, z zachowaniem przyrostowej metody ich tworzenia. Kolejne kopie bezpieczeństwa stanowią więc bazę do utworzenia właściwej, zarchiwizowanej, wektorowej kopii bezpieczeństwa.

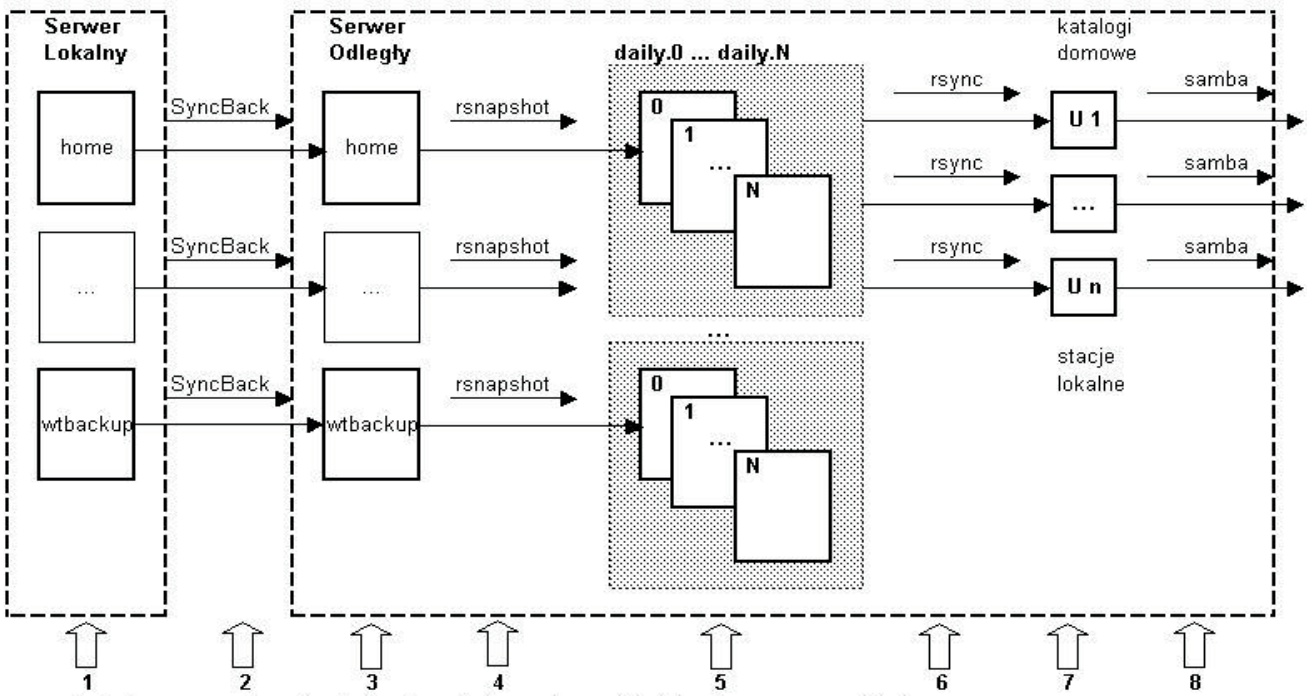
3.2.2. Transfer Backupu Dziennego do serwera linuxowego

Do transferu zestawu katalogów przeznaczonych do archiwizowania jest wykorzystywany program narzędziowy SyncBack. Umożliwia on tworzenie kopii plików we wskazanej lokalizacji – na dysku lokalnym lub sieciowym, na serwerze FTP lub nośniku wymiennym. Przy jego użyciu transfer może być wykonywany automatycznie, bez interwencji użytkownika, zgodnie ze zdefiniowanym harmonogramem. Przesył każdego z katalogów może być sterowany niezależnie, zgodnie ze zdefiniowanym przez użytkownika profilem. W omawianym systemie wprowadzono profile **ProfKatalog1,…, ProfKatalogN**. Program SyncBack posadowiono w katalogu **C:\backup\SyncBack** na serwerze **SerwerZsm.** W profilach wskazano jako miejsce docelowe w serwerze Linuxowym **BackPro** katalog o nazwie **/home/samba/snapshot.**

Warunkiem umożliwiającym transferowanie danych do serwera Linuxowego jest uruchomienie komunikacji pomiędzy serwerami o odmiennych platformach software-owych. Do realizacji tego celu wykorzystano omówiony dalej pakiet oprogramowania **Samba**.

Istotą transferu katalogów z serwera **ServerZsm** do serwera **BackPro** za pomocą narzędzia **SyncBack** jest wykorzystanie informacji o ostatnich zmianach plików tych zasobów, dzięki czemu transfer obejmuje jedynie pliki zmodyfikowane od czasu ostatniego przekazu. Ta właściwość pozwala na przyjęcie metody archiwizowania codziennego, bez narażania sieciowych systemów informatycznych na przeciążanie dużym ruchem w sieci w znacznych interwałach czasowych.

Rys. 2 przedstawia transfer i przetwarzanie archiwizowanych danych w serwerze źródłowym, serwerze archiwizującym oraz w stacjach lokalnych użytkowników systemu.



1. dzienne przyrostowe kopie bezpieczeństwa wybranych katalogów na serwerze lokalnym
2. transfer przyrostowy katalogów (SyncBack) do serwera odległego
3. dzienne przyrostowe kopie bezpieczeństwa wybranych katalogów na serwerze odległym
4. tworzenie wektorowej kopii bezpieczeństwa
5. zbiór przyrostowych wektorów kopii bezpieczeństwa
6. transfer przyrostowy najnowszych katalogów domowych użytkowników do serwera odległego
7. katalogi domowe użytkowników na serwerze odległym
8. udostępnianie zawartości katalogów domowych użytkownikom autoryzowanym

Rys. 2. Funkcjonowanie systemu archiwizowania serwera Windows 200x

3.2.2.1. Komunikacja serwerów Linux - Windows Server 2003

Dla zapewnienia właściwej współpracy serwera **ServerZsm**z serwerem **BackPro,** wykorzystano oprogramowanie **samba** stanowiące moduł opracowany dla platformy Linux/Unix dla potrzeb zarządzaniu serwerami plików oraz drukarek, współpracujący z platformą Windows. Oprogramowanie to obsługuje protokół Server Message Block (SMB) - od którego pochodzi nazwa pakietu, przy czym jako protokół transportowy niższej warstwy używany jest TCP/IP. W Windows protokołem transportowym może być również NetBEUI i IPX, ale w nowszych wersjach TCP/IP jest domyślną opcją. Protokół SMB jest też nazywany Common Internet File System (CIFS). Samba jest serwerem plików Windows SMB/CIFS dla UNIX-a. Samba pozwala na tworzenie heterogenicznego (mieszanego) środowiska, w którym nie tylko w ramach jednej sieci lokalnej mogą działać obok siebie komputery z systemem operacyjnym Linux oraz Windows, ale także mogą wzajemnie korzystać ze swoich zasobów – plików i drukarek. Serwer Samby może na przykład systemom Windows udostępniać drukarkę linuksową lub systemom Unix dawać dostęp do plików środowiska Windows NT.

Poza zapewnieniem komunikacji pomiędzy serwerami, samba umożliwia także właściwe zarządzanie autoryzacją dostępu użytkowników serwera **ServerZsm** do zasobów informatycznych serwera **BackPro.**

3.2.2.2. Konfiguracja serwera samba

Podczas konfiguracji **samby** ustanowiono w serwerze Linuxowym udziały

/home/samba/backup

/home/samba/snapshot

przeznaczone odpowiednio: jako miejsce docelowe transferowanego Backupu Dziennego oraz do przechowywania odpowiednio przetworzonych danych zarchiwizowanych. Plik konfiguracyjny modułu samba **/etc/samba/smb.conf** ma postać:

```
[global]
        workgroup =             zsm
        server string =         Samba Server Version %v
        hosts allow =           127.  10.1.1.124   195.187.100.111
        log file =               /var/log/samba/log.%m
        max log size =          50
        server(deprecated)
        security =              server
        passdb backend =         tdbsam
        password server =       10.1.1.124
[homes]
        comment =                Home Directories
        browseable =            no
        writable =              yes
[backup]
        path =                  /home/samba/backup
        writeable =              yes
        comment =               Kopia Przyrostowa
        valid users =           backmaster, marian, tkrakowka
        create mask =           0777
        directory mask =        0777
[snapshot]
        comment =               Zrzut jednorazowy - dzienny
        path =                  /home/samba/snapshot
        writeable =              yes
        valid users =           backmaster, marian
        create mask =           0777
        directory mask =         0777
```

Powyższa konfiguracja jest przeznaczona do zarządzania transferem zestawu katalogów przeznaczonych do archiwizowania z serwera **ServerZsm** do serwera **BackPro.**

Zastosowanie **samby** ma również na celu właściwe zarządzanie dostępem właścicieli do ich prywatnych katalogów domowych w serwerze BackPro z poziomu systemu operacyjnego Windows. W katalogach tych, co opisano dalej, są bowiem przechowywane ostatnie kopie dzienne zasobów prywatnych poszczególnych użytkowników.

3.2.3. Archiwizacja wektorowa na serwerze Linuxowym

Do długoterminowej archiwizacji wektorowej zaprojektowano serwer (o nazwie **BackPro)**. z zaimplementowanym systemem operacyjnym Linux Fedora Core. Ten wybór został podyktowany szerokimi możliwościami komunikowania się systemu Linux z innymi systemami operacyjnymi stosowanymi w PIAP (Windows, NetWare) oraz wyposażeniem go w programy narzędziowe umożliwiające programową realizację archiwizowania. Jak wskazano wyżej, serwer ten charakteryzuje się wysokim stopniem bezpieczeństwa sprzętowego, którego podstawą jest autonomiczny system redundancyjny RAID. Do archiwizacji wektorowej zastosowano oprogramowanie rsnapshot.

Konfiguracja i funkcjonowanie rsnapshot

Oprogramowanie rsnapshot jest narzędziem umożliwiającym zarządzanie dziennymi, przyrostowymi kopiami bezpieczeństwa. Ten program jest uruchamiany okresowo, zarządzany przez moduł wywołań czasowych crontab ( )

- 08 06 * * * **/usr/bin/rsnapshot** daily 2>/dev/null

Podczas tworzenia wektora zapisów kopii kolejnych, przyrostowych kopii bezpieczeństwa, program oszczędza przeogromną ilość przestrzeni dyskowej dzięki temu, że miejsce wykorzystywane przez ten program, to jeden katalog pełnego backupu archiwizowanych zasobów plus kolejne katalogi dzienne z plikami zmienionymi od czasu ostatniego backupu. Pozostałe, nie zmienione pliki są reprezentowane odniesieniami (tzw. linki twarde) do plików wcześniej archiwizowanych. Liczba katalogów archiwizacji dziennych zależy od konfiguracji programu. W PIAP przyjęto 30.

Plik konfiguracyjny /etc/rsnapshot

```
config_version      1.2
snapshot_root       /home/samba/backup/
cmd_cp              /bin/cp
cmd_rm              /bin/rm
cmd_rsync           /usr/bin/rsync
cmd_logger          /usr/bin/logger
cmd_du              /usr/bin/du
interval            daily  30
verbose             2
loglevel            3
logfile             /var/log/rsnapshot
lockfile            /var/run/rsnapshot.pid
backup              /home/samba/snapshot/home/          ./
backup              /home/samba/snapshot/pub/           ./
```

Z powyższego pliku konfiguracyjnego wynika, że w katalogu docelowym przeznaczonym na backup długoterminowy (archiwum wektorowe) będą zawarte katalogi: **daily.0,…daily.30**, oraz, że po każdej dziennej operacji rsnapshot, katalog **daily.30** zostanie usunięty, przypisane nazwy katalogów ulegną rotacji o +1, oraz zostanie utworzony nowy katalog **daily.0** zgodnie z zasadą tworzenia kolejnego elementu wektora archiwum, tj. będzie zawierał zmodyfikowane od czasu ostatniego backupu pliki oraz linki twarde do pozostałych plików, zapisanych wcześniej.

3.2.4. Zarządzanie dostępem do zasobów informatycznych serwera Linux

O ile archiwizacją w systemach informatycznych zarządza użytkownik z największymi prawami dostępu – backupmaster, i jego działania mogą zapewnić ewentualne odzyskanie utraconych danych, to – poza archiwizacją systemową - uwzględniono w PIAP metodę dostępu do katalogów domowych użytkowników serwera **SerwerZsm**. W tym celu, za pomocą narzędzia rsync, jest realizowana synchronizacja pomiędzy ostatnimi przekazanymi z serwera **ServerZsm** do serwera **BackPro** kopiami dziennymi katalogów domowych użytkowników, a takimi samymi zasobami posadowionymi wcześniej w katalogach domowych na serwerze Linuxowym. Tak więc, użytkownicy stanowiący grupę konstruktorską Proengineer mogą wykorzystać serwer archiwizujący niezależnie od zarządzanego poza ich kontrolą systemu archiwizacji wektorowej. Dostęp do zasobów jest również realizowany za pomocą oprogramowania **samba**, zarządzającym w tym przypadku dostępem do plików.

4. LITERATURA

1. W. Curtis Preston, 2008 r., Archiwizacja i odzyskiwanie danych, Helion
2. B. Danowski, 2004 r. , Norton Ghost, ćwiczenia, Helion
3. M. Hart, R.G. Freeman, 2008 r., Oracle Database 10g RMAN. Archiwizacja i odzyskiwanie danych. Helion
4. P. Czarny, 2002 r., Odzyskiwanie danych w praktyce, Helion
5. William von Hagen, B.K. Jones, 2007 r., 100 sposobów na Linux Server. Wskazówki i narzędzia dotyczące integracji, monitorowania i rozwiązywania problemów, Helion