Zeszyty Naukowe Wydziału Elektrotechniki i Automatyki Politechniki Gdańskiej Nr 20

XIV Seminarium
ZASTOSOWANIE KOMPUTERÓW W NAUCE I TECHNICE' 2004
Oddział Gdański PTETiS

WYKORZYSTANIE DYSTRYBUCJI SYSTEMU LINUX TYPU LIVECD DO BUDOWY KLASTRÓW OBLICZENIOWYCH

Jerzy KACZMAREK[1], Michał WRÓBEL[2]

1. Politechnika Gdańska, ul. G. Narutowicza 11/12, 80-952 Gdańsk
   tel: (058) 347 2682 fax: (058) 347 2727 e-mail:jkacz@eti.pg.gda.pl
2. Politechnika Gdańska, ul. G. Narutowicza 11/12, 80-952 Gdańsk
   tel: (058) 348 6085 fax: (058) 347 1006 e-mail:wrobel@task.gda.pl

W artykule opisano metodę wykorzystywania mocy obliczeniowych zgromadzonych w sieciach lokalnych poprzez tworzenie klastrów obliczeniowych. Metoda pozwala na wykorzystywanie systemów operacyjnych przechowywanych na nośnikach trwałych, dzięki czemu nie jest konieczna żadna zmiana konfiguracji istniejących systemów. Wieloprocesorowy klaster obliczeniowy zbudowany zgodnie z zaproponowaną metodą ma możliwości równoważenia obciążeń poszczególnych procesorów. W artykule opisano mechanizm dynamicznego przyrostowego konfigurowania klastra. Przedstawiono także interfejs użytkownika pozwalający kontrolować i monitorować pracę klastra. Metoda może być wykorzystana w dowolnej sieci lokalnej, kiedy potrzebne jest zwiększenie możliwości obliczeniowych.

1. WSTĘP

Współczesne badania w wielu dziedzinach nauki i techniki wymagają dużych mocy obliczeniowych. Obecnie użytkownicy wykorzystują zasoby zgromadzone w centrach obliczeniowych, takich jak Centrum Informatyczne TASK. Zasobami takimi są głównie superkomputery składające się z wielu procesorów i dużej ilości pamięci oraz grupy komputerów osobistych połączonych szybkimi sieciami w klastry obliczeniowe. Obecnie największy klaster w Polsce, składający się z 128 dwuprocesorowych komputerów osobistych, znajduje się w Gdańsku. Pomimo ciągłej rozbudowy zasobów, centra obliczeniowe nie są w stanie obsłużyć na bieżąco wszystkich użytkowników. Niektóre zadania mogą czekać nawet kilka tygodni zanim zostaną zrealizowane.

Istnieją jednak liczne, nie wykorzystywane powszechnie zapasy mocy obliczeniowych w postaci komputerów osobistych w sieciach lokalnych firm i uczelni. Komputery takie zazwyczaj nie są używane poza godzinami zajęć laboratoryjnych lub pracy w firmach. Jednym ze sposobów wykorzystania tej mocy jest połączenie komputerów w klaster obliczeniowy. Najprostszym sposobem na stworzenie takiego klastra jest wykorzystanie dystrybucji systemu Linux typu LiveCD.

Wszystkie programy potrzebne do obliczeń rozproszonych oraz interfejs użytkownika dostarczane są na płycie CD-ROM i nie wymagają instalacji na dysku.

W artykule zostanie przedstawiony sposób działania klastrów zbudowanych za pomocą dystrybucji typu LiveCD, które będą nazywane klastrami LiveCD. Omówione zostaną zalety i wady rozwiązań tego typu. Pokazane będą przykłady możliwych zastosowań klastrów LiveCD.

2. ZASADA DZIAŁANIA KLASTRA LIVECD

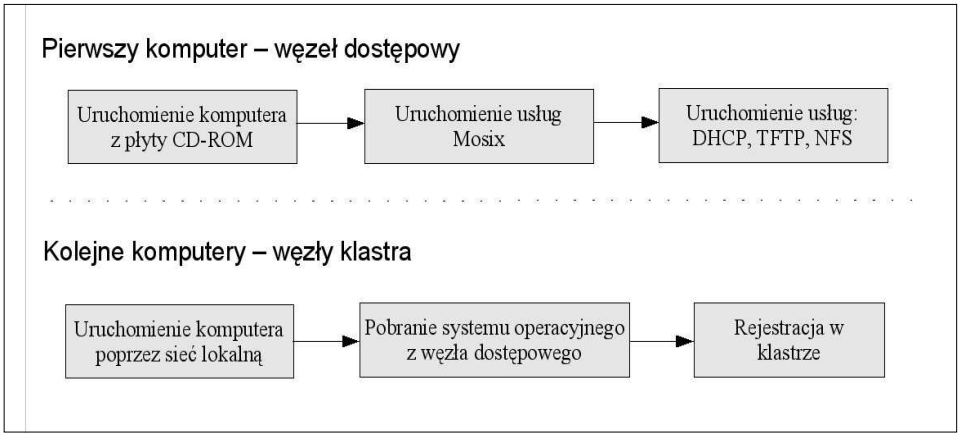
Klastrem (ang. cluster) nazywany jest system komputerowy składający się z połączonych komputerów, zwanych węzłami (ang. nodes). Węzły klastra powinny działać pod tym samym systemem operacyjnym [1,2]. Dla użytkownika klaster jest widoczny jako jedna maszyna. Ta kluczowa z punktu widzenia użytkownika funkcjonalność realizowana jest za pomocą specjalnych programów.

Obecnie klastry dzielą się na dwie grupy pod względem sposobu realizowania wielozadaniowości i równoległego wykonywania programu. Jedną grupę stanowią rozwiązania oparte na mechanizmie zwanym Beowulf, które wymagają napisania aplikacji obliczeniowej z wykorzystaniem bibliotek do programowania równoległego i rozproszonego, takich jak PVM lub MPI.

Druga grupa klastrów oparta jest o mechanizm Mosix. Można na nich wykonywać aplikacje obliczeniowe nie uwzględniające rozproszenia, które jest realizowane za pomocą migracji procesów programu na różne węzły klastra. Obecnie najpopularniejszą implementacją tego mechanizmu jest projekt o nazwie OpenMosix [3].

Klastry LiveCD mogą być oparte zarówno o mechanizm Beowulf, jak i Mosix. Tworzenie klastrów za pomocą dystrybucji LiveCD nie wymaga żadnej ingerencji w oprogramowanie zainstalowane na komputerach. System operacyjny oraz wszystkie aplikacje są uruchamiane bezpośrednio z płyty CD-ROM. Jedynie wyniki obliczeń są zapisywane na dysku twardym komputera.

Proces budowania klastra typu LiveCD zostanie przedstawiony na przykładzie zastosowania dystrybucji klaster.cdlinux.pl, która została przygotowana na Wydziale Elektroniki, Telekomunikacji i Informatyki w ramach projektu cdlinux.pl [4]. Na rysunku 1 przedstawiono schematy uruchamiania klastra z wykorzystaniem dystrybucji klaster.cdlinux.pl.



Rys. 1 Schemat uruchamiania klastra LiveCD

2.1. Uruchamianie pierwszego węzła klastra

W pierwszym kroku należy uruchomić dystrybucję klaster.cdlinux.pl na jednym wybranym komputerze. Uruchomienie dystrybucji odbywa się poprzez włożenie płyty CD-ROM do czytnika. System operacyjny jest ładowany do pamięci z tego nośnika. System sam wykryje cały zainstalowany sprzęt na komputerze i skonfiguruje go. Pod koniec procesu startowania zostaną uruchomione usługi mechanizmu OpenMosix. Dzięki nim jest możliwe wykonywanie zadań rozproszonych na wielu komputerach, które zostaną podłączone do klastra. Ostatnim krokiem jest wystartowanie szeregu usług, które umożliwią dołączenie pozostałych węzłów klastra. Usługami tymi są:

- serwer DHCP – umożliwiający dynamiczne przydzielanie numerów IP w sieci wewnętrznej klastra,
- serwer TFTP – udostępniający pozostałym węzłom klastra pliki potrzebne do wystartowania systemu operacyjnego, dzięki niemu nie jest konieczne uruchamianie systemu z płyty CD-ROM na każdym węźle,
- serwer NFS – służący do wymiany danych pomiędzy węzłami klastra.

Wraz z zakończeniem startu pierwszego węzła, klaster jest gotowy do działania. W przypadku, gdy nie zostanie podłączony żaden inny węzeł system zachowuje się tak, jak zwykły komputer.

2.2. Dołączanie kolejnych węzłów

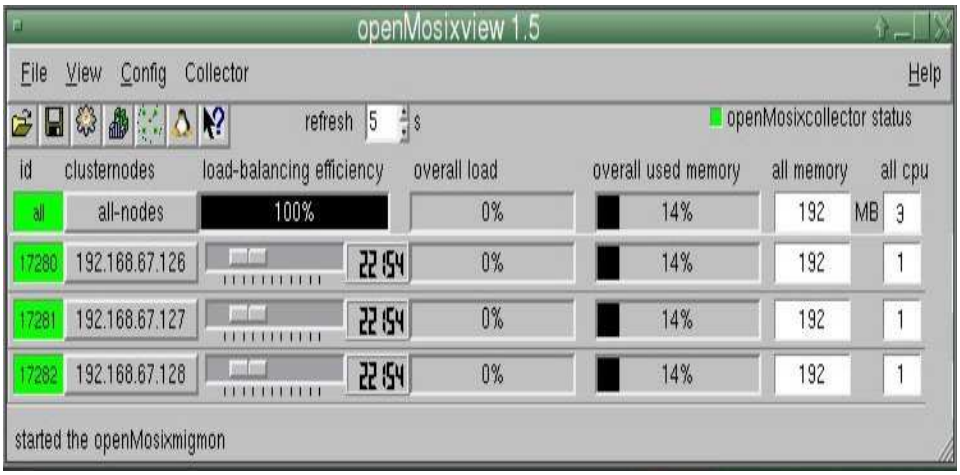
Istnieje możliwość dołączania kolejnych węzłów do klastra. Komputery należy uruchomić i w systemie BIOS wybrać opcję, aby system operacyjny został pobrany z sieci lokalnej (ang. network boot). Komputer będzie szukał w sieci lokalnej serwera DHCP, który będzie w stanie przekazać mu za pomocą protokołu TFTP jądro systemu operacyjnego. Jeżeli wcześniej w sieci lokalnej został uruchomiony komputer z dystrybucją klaster.cdlinux.pl, zostanie on znaleziony przez nowy komputer. Następnie serwer DHCP, uruchomiony na węźle dostępowym klastra, przydzieli unikalny w wewnętrznej sieci klastra numer IP. Kolejnym krokiem jest pobranie jądra systemu operacyjnego z serwera TFTP. Po jego załadowaniu zostanie przekopiowana poprzez protokół NFS reszta plików potrzebna do uruchomienia systemu operacyjnego.

Ilość węzłów klastra, które mogą być dołączone do niego jest teoretycznie nieograniczona. W praktyce, po przekroczeniu pewnej ilości węzłów, zależnej od rodzaju zadania, wydajność obliczeń zacznie spadać [5,6].

Po uruchomieniu komputera w opisany sposób zostaje on automatycznie zarejestrowany w klastrze. Użytkownik nie musi wykonywać żadnych dodatkowych operacji.

2.3. Kontrola obciążeń węzłów klastra

W celu kontroli nad dołączonymi węzłami oraz ich obciążeniem można wykorzystać program openmosixviewer, opracowany w ramach projektu OpenMosix. Program pokazuje aktualne wykorzystanie zarówno procesora jak i pamięci na każdym z dołączonych węzłów. Za pomocą programu można regulować priorytet obciążenia dla poszczególnych węzłów. Widok okna programu przedstawiono na rysunku 2.



Rys. 2 Okno programu openmosixviewer

Na rysunku 2 można zauważyć, że klaster składa się z trzech węzłów jednoprocesorowych o unikalnych identyfikatorach i numerach IP. Interfejs pokazuje obciążenie każdego z węzłów (ang. overall load), zużycie pamięci (ang. used memory). Za pomocą suwaka regulacji obciążenia (ang. load-balancing) można prowadzić optymalizację obciążenia poszczególnych węzłów.

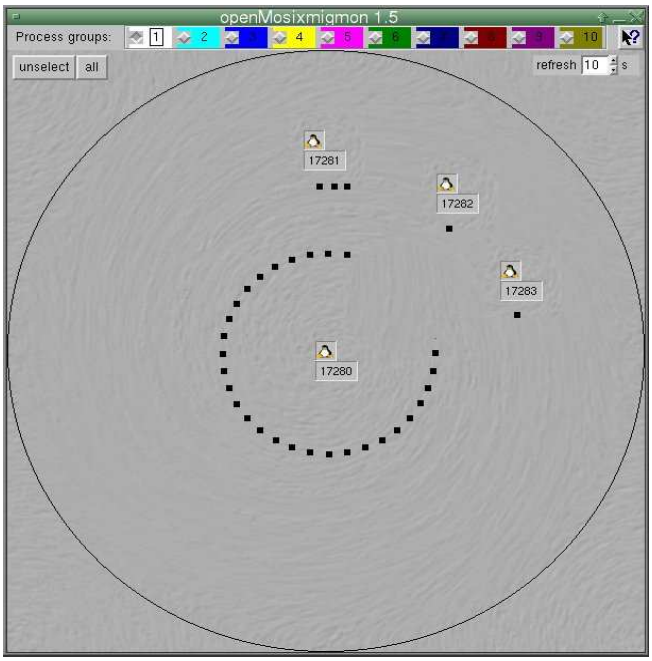
3. WYKONYWANIE ZADAŃ OBLICZENIOWYCH

Zadania obliczeniowe powinny być uruchamiane z pierwszego węzła klastra, zwanego węzłem dostępowym. Użytkownik uruchamiając zadanie nie musi wiedzieć, że jest ono wykonywane na klastrze. System sam zadba o to, żeby procesy były rozpowszechniane pomiędzy różne węzły klastra. Sumaryczne wyniki uzyskane z poszczególnych węzłów zostaną przedstawione użytkownikowi w taki sposób, jakby pracował on na pojedynczej maszynie.

W czasie wykonywania obliczeń równoległych istotne jest zapewnienie możliwości równomiernego obciążenia wszystkich dostępnych procesorów. Mechanizm ten zwiększa wydajność klastra.

Mechanizm migracji procesów w ramach zbudowanego klastra.cdlinux.pl jest ilustrowany z wykorzystaniem programu openmosixviewer. Przykładowy chwilowy stan obciążenia komputerów, wraz z oznaczeniem procesów, jakie są aktualnie na nich wykonywane, przedstawiono na rysunku 3.

Ikony z symbolem pingwina reprezentują poszczególne komputery wchodzące w skład klastra. Centralny komputer (na rysunku oznaczony numerem 17280) jest węzłem dostępowym. Symbole kwadratów przy każdym komputerze oznaczają procesy, jakie są na nich wykonywane. Migracja procesów pomiędzy węzłami odbywa się automatycznie i jest kontrolowana przez węzeł dostępowy. Użytkownik może również ręcznie wymusić migrację procesu na wybraną maszynę poprzez przeciągnięcie odpowiedniego kwadratu reprezentującego proces do innego komputera w klastrze.



Rys. 3 Openmosixviewer – migracja wątków

4. ZALETY I WADY KLASTRÓW LIVECD

Zastosowanie dystrybucji systemu Linux typu LiveCD jest najprostszym sposobem budowania klastrów. Klaster taki można zbudować z dowolnej ilości maszyn połączonych siecią lokalną. W czasie, gdy laboratoria komputerowe nie są wykorzystywane w ramach planowanych zajęć, można je wykorzystać do przeprowadzenia obliczeń.

Największą zaletą rozwiązań tego typu jest fakt, że nie jest konieczne dokonywanie żadnych zmian w oprogramowaniu, zainstalowanym na poszczególnych komputerach. Nie jest również istotny system operacyjny, jaki jest na nich zainstalowany. Wszystkie programy oraz dane są przechowywane na płycie CD-ROM lub w pamięci RAM. Jedynie wyniki można zapisać na dysku twardym.

Głównym zastosowaniem klastrów LiveCD są obliczenia, które nie wymagają dużych zasobów dyskowych, a poszczególne procesy nie wymieniają dużej ilości danych pomiędzy sobą. Przykładami takich obliczeń są projekty wizualizacyjne wykorzystujące grafikę trójwymiarową oraz przetwarzające plików dźwiękowe i video.

Są dwie główne wady klastrów typu LiveCD. Pierwszą jest brak dedykowanych sieci komputerowych. Większość sieci lokalnych działa w technologii Fast Ethernet, która oferuje przepustowość rzędu 100Mbit/s. Jest to zdecydowanie za mało dla obliczeń, które przekazują duże ilości danych pomiędzy węzłami klastra. Współczesne klastry wykorzystują przynajmniej sieci Giga Ethernet (o przepustowości 1Gbit/s), a coraz częściej są wykorzystywane specjalne, dedykowane rozwiązania, takich jak Myrinet, które oferują znacznie mniejsze opóźnienia przesyłania danych.

Drugą wadą jest brak przestrzeni dyskowej. Część aplikacji obliczeniowych przetwarza dane o objętości kilku, a nawet kilkunastu gigabajtów. W biurowych

i laboratoryjnych sieciach lokalnych takie przestrzenie dyskowe są często nieosiągalne. W takim przypadku zawsze istnieje jednak możliwość dołączenia macierzy dyskowej lub podłączenia się do zdalnego systemu plików.

5. WNIOSKI KOŃCOWE

Klastry typu LiveCD są dobrym rozwiązaniem dla użytkowników, którzy potrzebują znacznych mocy obliczeniowych i mają dostęp do komputerów w sieciach lokalnych, takich jak np. laboratoria komputerowe. Dzięki tego typu dystrybucjom systemu Linux można w szybki sposób zbudować klaster i przeprowadzić na nim obliczenia. Rozwiązania takie w przyszłości powinny dalej się dynamicznie rozwijać.

6. BIBLIOGRAFIA

1. Silberschatz A., Galvin P. B.: Podstawy systemów operacyjnych, WNT, 2002.
2. Tanenbaum A.: Rozproszone systemy operacyjne, PWN, Warszawa, 1997.
3. http://openmosix.sourceforge.net/
4. Kaczmarek J., Wróbel M.: Obszary zastosowań dystrybucji cdlinux.pl, Zeszyty Naukowe WETI PG, Technologie Informacyjne, nr 3, 2004, str. 221-226.
5. Wilkinson B., Allen M.: Parallel Programming, Prentice Hall, 1999.
6. Bourke T.: Wyrównywanie obciążenia serwerów. O'Reilly, Warszawa, 2002.

UTILIZATION OF LIVECD LINUX DISTRIBUTION IN COMPUTABLE CLUSTER BUILDING

Article describes a method of cluster building, which is a good way to utilize computer power accumulated in local networks. The method allows to use operating system stored on read-only storage. Multiprocessor computable cluster based on this method allows to make load balancing on every processor. This article describes mechanism of cluster's dynamic configuration. User interface to control and monitor cluster performance was also described. The method can be used in every local network.