ZESZYTY NAUKOWE WSEI
SERIA:
TRANSPORT I INFORMATYKA

4(1/2014)

Zeszyty naukowe
Wyższej Szkoły ekonomii i Innowacji w Lublinie
SERIA: TRANSPORT I INFORMATYKA

Printed in Poland
Innovatio Press Wydawnictwo Naukowe
Wyższej Szkoły Ekonomii i Innowacji
20-209 Lublin, ul. Projektowa 4
tel.: + 48 81 749 17 77, fax: + 48 81 749 32 13
www.wsei.lublin.pl

ISSN: 2084-8005

Spis treści

Zeszyty Naukowe WSEI seria: TRANSPORT I INFORMATYKA, 4(1/2014), s. 5-10

Mariusz NYCZ, Bartosz MICHNO, Rafał MLICKI

Politechnika Rzeszowska, 35-959 Rzeszów, e-mail:sasit@prz.edu.pl

ANALIZA PODATNOŚCI NA ATAKI SOCJOTECHNICZNE W JEDNOSTCE SAMORZĄDU TERYTORIALNEGO

ANALYSIS OF VULNERABILITY TO SOCIAL ENGINEERING ATTACKS IN A LOCAL GOVERNMENT UNIT

Streszczenie

Celem tej analizy było sprawdzenie podatności na ataki socjotechniczne w lokalnej siedzibie samorządu terytorialnego. W tym celu została opracowana ankieta, w których poruszono problemy bezpieczeństwa informacji na poziomie zarówno podstawowym jak i bardziej rozbudowanym. Ankietowani odpowiadali na około 20 pytań. Wyniki ankiety dały podstawy do zbudowania przybliżonych profili ankietowanych wraz z poziomem ich podatności na określone zagrożenia socjotechniczne, m. in. w Internecie. Ze względów bezpieczeństwa ankieta oraz jej bezpośrednie wyniki nie będą ujawniane.

Summary

The purpose of the analysis was to check the vulnerability to social engineering attacks in a local government unit in Poland. To that end, a survey was designed where security of information issues were raised on both basic and advanced levels. Interviewees were answering to about 20 questions. On the basis of the results, approximate profiles of interviewees were created and as well as vulnerability levels to certain social engineering attacks were determined. For security reasons, the questions and the results of the survey will not be published.

Słowa kluczowe: socjotechnika, ankieta, bezpieczeństwo, analiza, podatność

Keywords: sociotechnics, survey, security, analysis, vulnerability

1. Wprowadzenie

Socjotechnika jest opisywana jako kategoria ataków, w której atakujący manipuluje innych w celu uzyskania prywatnych i istotnych informacji. Definicja ta obejmuje jedno z najszerszych zagadnień socjotechniki, ale nie jest kompletna. Socjotechnika to nie tylko manipulacja ludźmi, ale także nieautoryzowany dostęp do przedmiotów fizycznych i używanie w sposób niedozwolony publicznie dostępnych informacji. Na przykład, zbieranie numerów telefonów i list mailingowych w celu szkodliwej działalności jest również uważane za socjotechnikę.

W socjotechnice występuje wiele różnych typów ataków. W zależności od potrzeb typy ataków mogą być wymyślane, ulepszane bądź też łączone ze sobą. Jednak większość z nich bazuje na znanych popularnych typach. Oto niektóre z nich:

- kradzież urządzeń mobilnych

Najstarszy typ ataku. W czasach, gdy urządzenia przenośne nabierają coraz większego znaczenia i stają się co raz bardziej popularne, ten typ ataku okazuje się być jednym z najbardziej skutecznych. Prawdopodobieństwo sukcesu rośnie w firmach / instytucjach, gdzie wdrożona jest polityka BYOD.

- shoulder-surfing

Jest to najprostszy typ ataku. Atakujący stara się monitorować fizyczną aktywność użytkownika i jego urządzenia. Atakujący może monitorować ekran, klawiaturę lub ruchy rąk w celu przechwycenia prywatnych informacji.

- monitorowanie sieci

Monitorowanie sieci może ukazać typy usług najczęściej używane przez użytkowników. Poprzez ich identyfikację atakujący może rozpocząć rekonesans zabezpieczeń danej usługi i przeprowadzić ewentualny atak.

- digital dumpster diving

Co raz krótszy czas życia urządzeń elektronicznych z racji postępującego rozwoju technologii sprawia, że przestarzałe urządzenia w składowiskach np. elektrośmieci mogą mieć pozostawione w swojej pamięci prywatne i poufne dane.

- phishing

Przeważnie związany z fałszywymi stronami oraz wiadomościami e-mail. W przypadku fałszywych wiadomości e-mail atakujący używa nieprawdziwej tożsamości internetowej w celu oszukania odbiorcy [1].

2. Ankieta

Ankieta została umieszczona na specjalnej witrynie przeznaczonej do ankietowania użytkowników. Znajdowała się ona pod prywatnym linkiem URL, niedostępnym z poziomu publicznego dla wyszukiwarek i użytkowników nieznających pełnego adresu.

Ankietowani odpowiadali na ponad 20 pytań. Niektóre z ich miały charakter pouczający, sugerujący prawidłową odpowiedź. Pytania te były nieliczne i znajdowały się na końcu ankiety. W zamyśle autorów ankieta miała oprócz weryfikacji wiedzy użytkowników uzupełnianie jej.

Pytania wstępne zawierały podstawowy zestaw typowej ankiety dotyczącej tematów informatycznych. Pytano w niej o system operacyjny oraz rodzaj używanej przeglądarki internetowej.

Ankieta była całkowicie anonimowa.

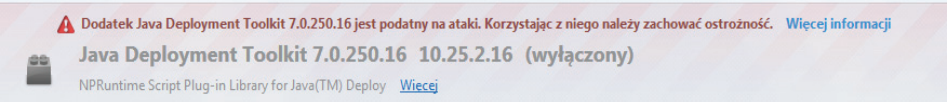
3. Analiza otrzymanych wyników

Analiza opisuje kolejno metody oszustw socjotechnicznych stosowanych w ankiecie. Na jej podstawie obliczono prawdopodobieństwo skuteczności danej metody. Są one posortowane malejąco według prawdopodobieństwa skuteczności. Nie wszystkie z nich są czysto socjotechniczne, ale przenoszone są z wykorzystaniem ataków socjotechnicznych. W takich przypadkach socjotechnika jest dopiero pierwszym etapem przeprowadzania ataku, a właściwe wykorzystywanie luk bezpieczeństwa występuje jako faza druga.

3.1. Java

Z przeprowadzonej analizy wynika, że najwięcej użytkowników może być narażonych na ataki związane ze środowiskiem programistycznym Java. Java jest bardzo popularnym oprogramowaniem wśród internautów – często wymagana jest do poprawnego wyświetlania stron internetowych. W przeglądarkach internetowych Java występuje pod postacią wtyczek włączanych na żądanie. W nowszych wersjach przeglądarek wtyczki te wyłączane są automatycznie. Jedna z nich, Mozilla Firefox, informuje użytkownika o niebezpieczeństwie związanym z użytkowaniem Javy.

Z przeprowadzonej analizy wynika, że najwięcej użytkowników może być narażonych na ataki związane ze środowiskiem programistycznym Java. Java jest bardzo popularnym oprogramowaniem wśród internautów – często wymagana jest do poprawnego wyświetlania stron internetowych. W przeglądarkach internetowych Java występuje pod postacią wtyczek włączanych na żądanie. W nowszych wersjach przeglądarek wtyczki te wyłączane są automatycznie. Jedna z nich, Mozilla Firefox, informuje użytkownika o niebezpieczeństwie związanym z użytkowaniem Javy.



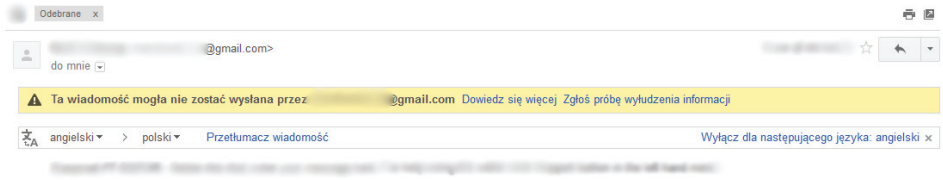
Rys. 1. Wbudowane zabezpieczenia w przeglądarce Firefox dot. Javy

Niektóre aplikacje, ze względu na konieczność obsługi wielu platform systemowych są napisane w Javie. Jest ona wobec tego środowiskiem bardzo podatnym na ataki, ponieważ większa uniwersalność ułatwia wyszukiwanie luk. Java jest również oprogramowaniem bardzo skomplikowanym. Wbrew pozorom również umożliwia to nadużywanie Javy w celu łamania jej systemów bezpieczeństwa (im bardziej „rozległe" oprogramowanie, tym większa szansa natrafienia na ewentualną lukę).

Atak na Javę może zostać przeniesiony np. w załącznikach wiadomości e-mail.

3.2. Fałszywe wiadomości e-mail

Kolejną metodą ataku, na którą podatni byliby użytkownicy rozwiązujący ankietę to wiadomości e-mail, gdzie nadawca używa fałszywej tożsamości. Wiadomości te wyglądają bardzo autentycznie. Istnieje wiele metod uwiarygodniania takich e-maili. Jedną z nich jest adres nadawcy – wykorzystuje się łatwe do przeoczenia literówki. Drugim sposobem jest wysyłanie wiadomości ze specjalnych generatorów wiadomości (najczęściej strony www). Cechą takich generatorów jest to, że oprócz podania adresata wiadomości podaje się również… adres nadawcy. Generator oszukuje serwery pocztowe i przesyła taką wiadomość dalej. Niektóre serwisy pocztowe, np. Gmail są w stanie wykryć większość takich oszustw i wyświetlają stosowne ostrzeżenie dla użytkownika.



Rys. 2. Ostrzeżenia w usłudze Gmail

Kiedy odbiorca wiadomości uzna fałszywą tożsamość nadawcy za prawidłową, atak w większości przypadków można uznać za udany.

3.3. Podstawione strony internetowe

Sfabrykowane strony internetowe wykorzystują identyczny lub do złudzenia podobny wygląd oryginalnej witryny. Z autentycznej strony zachowany został tylko układ elementów, a faktyczna funkcjonalność strony to przesyłanie wpisanych poświadczeń użytkownika (login, hasło, inne) do atakującego.

3.4. Oszustwa telefoniczne

Podobnie jak w przypadku wiadomości e-mail, kluczem jest tutaj podstawiona tożsamość. Oszustwo może ułatwić fakt samej rozmowy telefonicznej, która odbywa się „w locie". W przeciwieństwie do e-maili, gdzie treść wiadomości można

analizować wiele razy po odebraniu, w rozmowie telefonicznej czas na namysł jest ograniczony. Sprzyja to skuteczności ataku w przypadku rozmowy kreowanej na nagły i niespodziewany przypadek, wymagający szybkiej interwencji użytkownika.

3.5. Stare wersje przeglądarki Internet Explorer

Pewna część użytkowników ciągle używa wersji Internet Explorera, która jest przestarzała. Wersja wspomnianej przeglądarki o numerze 7 została wydana siedem lat temu. Z kolei wersja 8 swoją premierę miała pięć lat temu. Przeglądarka Microsoftu jest bardzo chętnie atakowanym oprogramowaniem – wśród internautów znana jest z ilości luk w bezpieczeństwie. Korporacja często wydaje aktualizacje typu security do IE i stara się niwelować braki. Aktualna wersja Internet Explorera to 11.

3.6. System operacyjny

Ten problem dotyczy użytkowników korzystających z systemu Microsoft Windows XP. System przestał być wspierany w kwietniu 2014 r. Microsoft przestał wydawać aktualizacje bezpieczeństwa dla tego systemu. Czyni to go podatnym na ataki typu „zero day" – poważne luki bezpieczeństwa, nieznane do czasu pierwszego ataku z nią w roli głównej. Jeśli taka luka zostanie odkryta, deweloperzy Microsoftu nie zapewnią już łatki bezpieczeństwa.

4. Szkolenie i powtórzenie ankiety

W okresie trzech miesięcy od przeprowadzonej ankiety, w lokalnej siedzibie samorządu terytorialnego odbyło się szkolenie, które miało charakter wykładowy. Prelegenci omawiali zagadnienia dotyczące szeroko pojętego bezpieczeństwa w sieci, w tym również dotyczące socjotechniki. Miesiąc po szkoleniu przeprowadzono ponownie ankietę wśród użytkowników.

Wyniki drugiej ankiety wskazywały na średnio o jedną czwartą lepsze i trafniejsze odpowiedzi. Oznacza to, że świadomość użytkowników na ataki socjotechniczne znacznie wzrosła.

5. Wnioski

Najczęstszymi obiektami ataków socjotechnicznych są pracownicy firm, instytucji. To pracownicy często nieświadomi, niedoszkoleni lub posiadający specjalne przywileje, „atrakcyjne" z punktu widzenia dla atakującego. Ofiarą ataków mogą paść też pracownicy kluczowych działów instytucji.

Co sprzyja atakom socjotechnicznym? Niska świadomość użytkowników to najpoważniejszy powód. Dlatego tak ważne jest wykonywanie okresowych szkoleń pracowników (zalecane jest przeprowadzanie ich 1–2 razy w roku).

6. Profilaktyka

Socjotechnika to najniebezpieczniejsza forma ataków na bezpieczeństwo z racji swojej natury. Ponieważ socjotechnika nadużywa cech ludzkich. np. zaufania, nie ma możliwości obrony przed nią tylko i wyłącznie za pomocą hardware'u i software'u.

Tak naprawdę nie istnieje technologia, która mogłaby zapobiec atakowi socjotechnicznemu. Można jedynie zmniejszyć prawdopodobieństwo wystąpienia udanego ataku poprzez:

- politykę klasyfikacji danych: z danych przepływających przez instytucję wybiera się te najbardziej istotne i stosuje się środki mające na celu chronienie ich,
- kształtowanie świadomości użytkowników poprzez nieustanne doszkalanie.

Bibliografia

[1] NYAMSUREN E., CHOI H., Preventing Social Engineering in Ubiquitous Environment. IEEE Xplore Digital Library, odwiedziny: 5.06.2014, http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=4426307&queryText%3DPreventing+Social+Engineering+in+Ubiquitous+Environmen.

Zeszyty Naukowe WSEI seria: TRANSPORT I INFORMATYKA, 4(1/2014), s. 11-17

**Paweł ZAJĄC, Arkadiusz SZYMAŃSKI,
Damian CHALIMONIK, Igor ŁAPA**

Wyższa Szkoła Ekonomii i Innowacji w Lublinie, e-mail: Igor.Lapa@wp.pl, paulhare93@o2.pl

ZASTOSOWANIE MICROSOFT KINECT W TWORZENIU ANIMACJI KOMPUTEROWYCH

APPLICATION OF MICROSOFT KINECT IN CREATION OF COMPUTER ANIMATION

Streszczenie

Typowym zastosowaniem interfejsu Microsoft Kinect jest rozpoznawanie ruchu. Możliwość tę wykorzystuje się głównie w grach. W tej pracy zamieszczono opis innego zastosowania interfejsu Kinect. Przedstawiono kolejne kroki, które prowadzą od przechwycenia ruchu obiektu, aż do stworzenia prostej animacji komputerowej.

Summary

A typical application of interface Microsoft Kinect is motion detection. This possibility is mainly used in games. In this work, a description of other application of Kinect interface is placed. The paper presents the consecutive steps that lead from the motion capture, to making a simple computer animation.

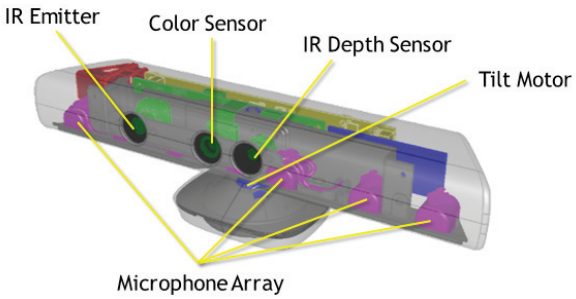
Słowa kluczowe: Motion Capture, Microsoft Kinect

Keywords: Motion Capture, Microsoft Kinect

1. Wstęp

Motion Capture (lub w skrócie Mocap) jest techniką produkcji animacji komputerowej pozwalającą na uzyskanie bardzo realistycznego odzwierciedlenia naturalnego ruchu. Pozwala na uzyskanie dużo dokładniejszego i łatwiejszego w obróbce materiału niż w przypadku konwencjonalnej animacji, jednak nie ogranicza się tylko i wyłącznie do przechwytywania ruchu.

W opisywanym przypadku do rejestracji ruchu wykorzystano czujnik ruchu Kinect firmy Microsoft. Istotnym dla nas elementem tego urządzenia jest kamera rejestrująca obraz o rozdzielczości 640×480 z częstotliwością 30 Hz, wyposażona w system określania odległości. Chmura promieni podczerwonych zostaje rzucona na nagrywanego aktora, a kamera z nałożonym filtrem podczerwieni odczytuje głębokościowe położenie poszczególnych punktów. Otrzymane nagranie daje wyraźny obraz położenia aktora, jego sylwetki i ruchu w przestrzeni z perspektywy kamery. Do rejestracji ruch zastosowano program iPi Recorder firmy iPi Soft [2].



Rys. 1. Widok elementów interfejsu Microsoft Kinect [4]

Surowy materiał, którym jest zwykły film video w formacie AVI będzie przetwarzany w programie iPi Mocap Studio. W narzędziu tym jest zdefiniowany domyślny obiekt referencyjny składający się z uniwersalnie rozmieszczonych kości. Obiekt jest umieszczony w przestrzeni trójwymiarowej, na którą nakłada się nagranie. Nagranie zawiera informacje z kamery głębokościowej, które iPi Mocap Studio wizualizuje obiekty na trójwymiarowej scenie [2]. Obiekt referencyjny umieszczamy na obrazie otrzymanym przez nagranie ruchu sensorem Kinect. Zaimportowanie filmu powoduje ustawienie się w jego pierwszej klatce i to właśnie w tym miejscu następuje dopasowanie sylwetki obiektu referencyjnego do sylwetki nagranego aktora. W tym celu skaluje się obiekt referencyjny do wymiarów nagranego aktora i stara jak najdokładniej nałożyć go na sylwetkę nagranego aktora.

Mając punkt odniesienia, jakim jest pierwsza klatka nagrania, przeprowadza się proces tzw. trackingu. Program śledzi poklatkowo zmiany w sylwetce aktora na nagraniu i dopasowuje położenie obiektu referencyjnego, aby pokrywał się z ru-

chem na filmie. Po dopasowaniu położenia w każdej klatce zapisywane są rotacje kości obiektu referencyjnego do późniejszego eksportu jako niezależną animację.

2. Przechwytywanie ruchu w praktyce

2.1. Aktor

Motoryka aktora jest istotna dla uzyskania efektu zgodnego z oczekiwaniami. Najlepszym wyborem na aktora jest osoba zajmująca się pantomimą lub pochodną tego gatunku.

Nagrywane sekwencje wymagać będą od aktora umiejętności odgrywania scen bez rekwizytów i dostosowanego otoczenia. Dla efektywniejszej pracy programu otoczenie powinno być jak najmniej różnorodne. Należy unikać przedmiotów, którymi miałby się posługiwać aktor i które mogłyby go zasłonić. Aktor powinien być wstępnie przygotowany do typu ujęcia, aby zaoszczędzić czas na powtarzanie nagrań. Ważnym elementem jest również strój aktora. Powinien on być przylegający do ciała.

Od strony programu jedynymi wymaganiami w stosunku do aktora jest wprowadzenie parametru jego wzrostu oraz minimalne dopasowanie od sylwetki aktora do interfejsu programu. Program będzie w stanie przechwycić i utworzyć animację, jednak przypadkowa osoba nie gwarantuje dobrej animacji, ponieważ takie osoby mogą wykonywać mimowolnie nienaturalne ruchy, co wiąże się z koniecznością wprowadzenia większej ilości poprawek w samym pliku animacji. Nie należy również nastawiać się na brak jakichkolwiek poprawek.

2.2. Konfiguracja środowiska

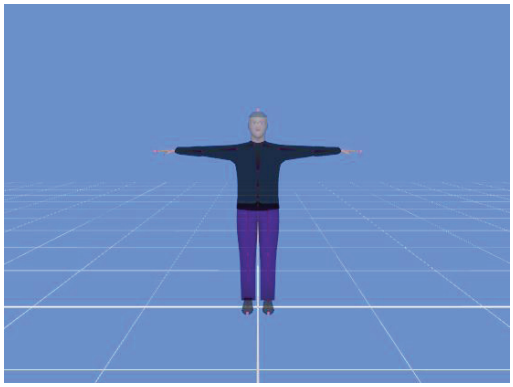
Zakłada się, że docelowym obszarem pozwalający na przeprowadzenie nagrania będzie minimalny obszar. Za minimalny obszar pozwalający na przechwyt uważamy obszar o rozmiarach około 3 m x 3 m, w takim wypadku ustawia się Microsoft Kinect na wysokości od 0,5m do 1 m i odległości od ściany(tła) około 5 m.

Rys. 2. Grupa animatorów w czasie pracy

Wyznaczenie obszaru roboczego rozpoczyna się od umieszczenia aktora w kadrze przy ścianie (tle). Pierwszy marker umieszcza się w miejscu, w którym występuje obraz aktora od stóp do głowy. Kolejny znacznik powinien znaleźć się w miejscu, gdzie występuje pełny widok aktora z podniesionymi rękoma. Kolejnymi markerami wyznacza się maksymalne wychylenie aktora w prawo i w lewo. Ostatnie zaznaczenie określa środek sceny i miejsce, w którym zaczynać będzie się każda animacja.

2.3. Proces przechwytywania

W zależności od wersji oprogramowania monity aplikacji iPi Recorder mogą być różne jednak zasada nagrywania przechwytu pozostaje ta sama. Pierwszym krokiem jest nagranie pustej sceny tj. sceny, w której nie pojawia się aktor. Potem następuje pojawienie się aktora, aktor ustawia się w wyznaczonym punkcie oraz staje w tzw. T-pose (pozycja ta przypomina literę T) (rys. 3) [2].

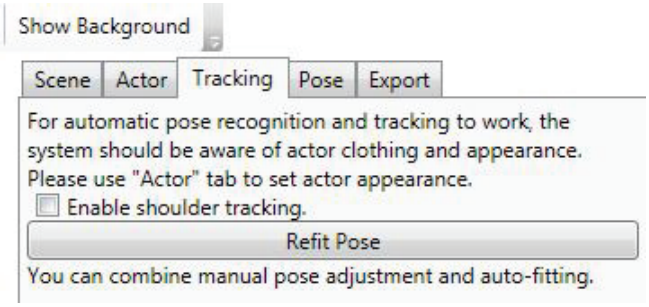


Rys. 3. Aktor w pozycji T-pose z perspektywy oprogramowania

Od wspomnianej pozycji zaczyna się każda animacja. Aktor może zakończyć animację pozycją T, lecz nie jest to wymagane od strony oprogramowania. Zaleca się stosowanie tej techniki. Efektem procesu przechwytywania jest, w zależności od wersji, plik .avi lub .iPiVideo.

Kolejnym naturalnym krokiem jest utworzenie nowego projektu na podstawie wcześniej nagranego klipu. Pierwszym krokiem jest określenie aktora jak wspomniano wyżej określamy jego rozmiar.

Po zaimportowaniu nagrania wyznacza się ramy pracy oprogramowania. Jako początek obszaru przetwarzanego wyznacza się miejsce, w którym to kończy się T-pose, a zaczyna właściwa animacja. W tym momencie należy użyć funkcji „Refit pose", aby dostosować aktora do postaci (rys. 4). Następnie wyznacza się koniec obszaru roboczego – zazwyczaj do końca zarejestrowanej animacji.

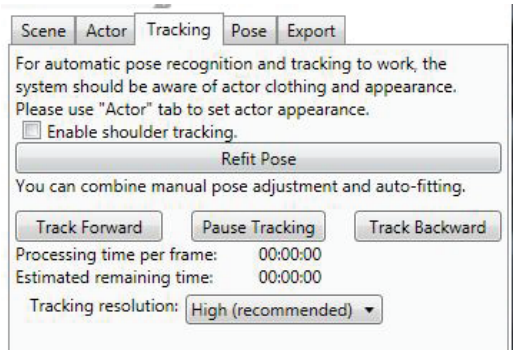


Rys. 4. Interfejs oprogramowania iPi Recorder - menu i przycisk „Refit Pose"

Po wyznaczeniu obszaru pracy uruchamia się algorytm przechwytujący poprzez funkcję „track forward". W tym momencie następuje proces przechwytywania, który w zależności od długości animacji może trwać bardzo długo. Program pozwala również na przechwytywanie od tyłu „track backward", co pozwala na wierniejsze oddanie ruchu. Praktykę taką stosuje się, aby zabezpieczyć się przed „zgubieniem kości" w nagraniu. Z tego samego względu zaleca się kończenie animacji pozycją T. Oba typy przechwytywania można stosować naprzemiennie, aby uzyskać jak najlepszy efekt. Zwiększona liczba markerów lub kamer ogranicza liczbę potencjalnych błędów [3, 5].

Projekt przygotowany w programie iPisoft należy wyeksportować do programu Blender. Blender udostępnia zestaw kości z gotowymi klatkami kluczowymi (ang. *keyframes*) rozpisanymi na linii czasu (ang. *timeline*). Do wykorzystania przechwyconej animacji potrzeba również modelu, który przypisuje się do szkieletu.

Opisywany model to humanoidalna postać bez twarzy, o proporcjach ciała odpowiadających proporcjom dorosłego mężczyzny. Model wykonany jest w Blenderze metodą „box modelling". Jest przystosowany do animacji (krawędzie w miejscach zgięć odpowiednio rozłożone tak, aby nie powodować nienaturalnych deformacji). Przygotowaną postać importuje się do Blendera razem z plikiem .bvh. Następnie dopasowuje się model postaci tak, aby krawędzie zgięć pokrywały się ze złączeniami kości. Jest to dużo trudniejsze niż przypisanie gotowego modelu tuż przed eksportem z programu iPisoft. Jednakże dzięki temu uzyskuje się w pełni kontrolowany efekt [1].

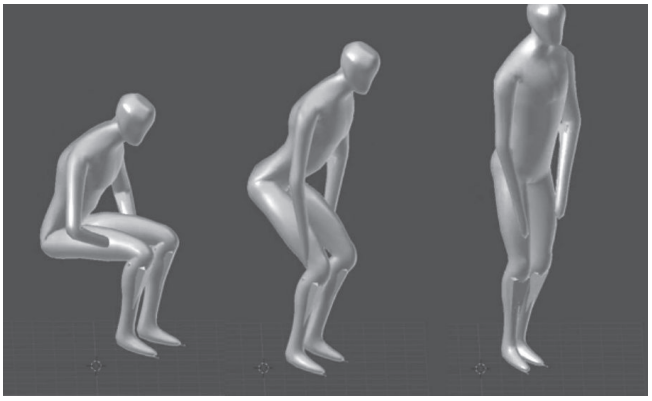


Rys. 5. Menu programu z najczęściej wykorzystywanymi opcjami „Track Forward" i „Track Backward"

Dla tak przygotowanej powierzchni deklaruje się relację ze szkieletem (ang. *armature*) „rodzic-dziecko"(ang. *parenting*), typu „odkształcanie szkieletowe" (ang. *armature deform*) z parametrem automatycznie wyliczającym przypisanie poszczególnych punktów do konkretnych kości (ang. *automatic weights*). W ten sposób model postaci jest już niemal gotowy do deformowania go za pomocą jego szkieletu. Nie zawsze ułożenie siatki względem kości jest zupełnie czytelne, co powoduje niekiedy niepożądane odkształcenia. W takiej sytuacji pozostaje ręczne przypisanie poszczególnych punktów konkretnym kościom za pomocą narzędzia malowania wpływów (ang. *weight painting*) [1]. Pozwala to wyeliminować opuszczone przez algorytm punkty oraz przypisać odpowiednią wartość wpływu kości na deformacje powierzchni w przypadku jej nieodpowiedniego odkształcania. Po tym model i jego szkielet są gotowe. Odtworzenie animacji wywołuje efekt zgodny z zarejestrowanym kilka etapów wcześniej.

3. Podsumowanie

Wykorzystanie Microsoft Kinect do przechwytywania animacji, jest stosunkowo prostą i niedrogą metodą, pozwalającą na uzyskanie animacji na zadowalającym poziomie realizmu. Na rys. 6 zamieszczono 3 klatki z opracowanej w ramach artykułu animacji.



Rys. 6. Sekwencja ruchu postaci z animacji utworzona na bazie zarejestrowanego ruchu

Otrzymaną w ten sposób animację można wykorzystać w grach komputerowych, filmach animowanych, czy też wizualizacjach. Metoda ta sprawdza się doskonale w projektach, które nie wymagają idealnego odwzorowania ruchu.

Bibliografia

[1] CGMASTERS: Animation in Blender, http://www.cgmasters.net/free-tutorials/animation-in-blender-2-5/, 27.04.2011.
[2] IPI SOFT: Motion Capture For The Masses, http://ipisoft.com/pr/iPi_Motion_Capture_Brochure.pdf, 09 V 2014.
[3] Liu G., McMillan L.: Estimation of missing markers in human motion capture. The Visual Computer, Volume 22, Issue 9-11, 2006.
[4] Microsoft Developer Network: Kinect for Windows Sensor Components and Specification, http://msdn.microsoft.com/en-us/library/jj131033.aspx, 10.06.2014.
[5] Scharacter D.S., Donnici M., Nuger E., Macay M., Benhabib B.: A Multi-Camera Active-Vision System for Deformable-Object-Motion Capture, Journal of Intelligent & Robotic Systems, 2013.

Zeszyty Naukowe WSEI seria: TRANSPORT I INFORMATYKA, 4(1/2014), s. 19-27

Katarzyna GĄZWA, Patryk GĄZWA, Arkadiusz SPRAWKA

Wyższa Szkoła Ekonomii i Innowacji w Lublinie

OVERCLOCKING A ZUŻYCIE ENERGII

OVERCLOCKING versus ENERGY CONSUMPTION

Streszczenie

Większość dostępnych na rynku podzespołów komputerowych dysponuje zapasem mocy. Coraz częstszym zjawiskiem staje się overclocking. Jego założeniem jest zmuszenie podzespołu do pracy z większą niż fabryczna wydajnością przy zachowaniu stabilności pracy komputera. Proces ten może dotyczyć procesora, karty graficznej, pamięci RAM i płyty głównej. Praktyczne badania przeprowadzone autorów niniejszego opracowania dowodzą, że dzięki overclockingowi można, korzystając z popularnych podzespołów komputerowych, uzyskać wydajność porównywalną, bądź lepszą od komponentów znacznie droższych. Technicznym celem testów jest przedstawienie stosunku wzrostu wydajności do odniesieniu do wzrostu zużycia energii.

Słowa kluczowe: overclocking, wydajność, zużycie energii

Summary

Most commercially available computer components offer power reserve. More frequent phenomenon becomes overclocking. Its aim is to force the computer component to work with more than factory performance while maintaining the stability of PC. This process may relate to the CPU, graphics card, RAM and motherboard. Practical research conducted the authors of this paper shows that by overclocking it is possible to achieve performance expensive components PC using the popular computer components. The technical objective of the tests is to present a growth of performance relative to the increase of electrical energy consumption..

Keywords: overclocking, performance, energy consumption

1. Wstęp

Overclocking czyli przetaktowanie polega na zmuszeniu podzespołu zestawu komputerowego do pracy z większą niż typowa częstotliwością. Nie wolno przy tym utracić stabilności pracy. Przetaktowanie może dotyczyć zarówno procesora, karty graficznej, pamięci RAM, jak i płyty głównej. Należy pamiętać, że każdy podzespół systemu ma swoje ograniczenia fizyczne wynikające z budowy fizycznej, których przekroczenie grozi uszkodzeniem danego podzespołu. Można wyróżnić dwa główne podejścia do overclockingu:

- softwareowy (programowy) – tzw. p2,0 odkręcanie spod platformy systemowej (np. Windows, Linux), zwiększenie częstotliwości magistrali FSB, odblokowanie mnożnika, modyfikacja ustawień BIOS;
- hardwareowy (sprzętowy) np. mechaniczne ingerowanie w układy elektroniczne poprzez przerywanie i łączenie ścieżek obwodów drukowanych.

1. 1. Płyta główna/obudowa

Mimo, że może wydawać się inaczej, płyta główna jest podzespołem dającym możliwość zmian ustawień standardowych. Podstawą jest rozbudowany BIOS, pozwalający na szeroki zakres zmian parametrów podzespołów komputera. Dodatkowe funkcje pozwalające na podkręcanie to: kontrola ustawień z poziomu systemu operacyjnego, automatyczny overclocking, kontrola temperatury za pomocą smartphona. Oczekuje się, że płyta główna będzie wyposażona w wydajny układ zasilania procesora z dodatkowym radiatorem.

Wydajne zasilanie wymaga zastosowania obudowy z dobrą cyrkulację powietrza. Taka cyrkulacja wspomagana dodatkowymi wentylatorami powoduje, że overclocking będzie bezpieczniejszy i efektywniejszy. Przy wydajnej wentylacji nawet komputer z ustawieniami standardowymi będzie działał dłużej, gdyż podzespoły zużywają się wolniej w niskich temperaturach (w szczególności kondensatory, niektóre połączenia lutowane).

1.2. Procesor

Do zwiększenia taktowania procesora można użyć jednego z programów dołączanych przez producentów płyt. Niektórzy producenci płyt głównych dodają do swoich produktów łatwy w obsłudze program umożliwiający zmianę taktowania.

Innym sposobem przetaktowania procesora jest zmiana ustawień w BIOS. Ustawienia te pozwalają nawet na minimalne zmiany parametrów. Najprostszym sposobem jest zwiększenie mnożnika procesora. W większości procesorów jest on zablokowany. W najnowszych seriach (AMD – seria Black Edition, Intel – seria Extreme Edition) producenci pozwalają użytkownikowi na jego swobodną zmianę. W serii procesorów K7 AMD istniała możliwość odblokowania dzielnika zegara za pomocą połączenia lub przerwania odpowiednich mostków na płycie procesora. W przy-

padku blokady mnożnika procesora jedyną opcją overclockingu jest zwiększenie taktowania magistrali FSB, która jest dostępna w większości płyt głównych z poziomu BIOS lub rzadziej programowo z poziomu systemu operacyjnego.

Istnieje również mniej spotykana i bardziej ryzykowna metoda, częściej stosowana w laptopach. Polega ona na połączeniu pinów procesora cienkim drutem. Operacja musi być wykonana tak precyzyjnie, aby zmodyfikowany procesor zmieścił się w gnieździe.

Wzrost częstotliwości zegara jest ograniczony architekturą rdzenia, przez co jeden procesor uda się znacznie podkręcić nie zwiększając napięcia, a drugi z innej serii osiągnie podobny wynik dopiero po zwiększeniu napięcia rdzenia. Wiąże się to z podwyższeniem ilości wydzielanego ciepła oraz zwiększeniem zużycia energii.

1.3. RAM

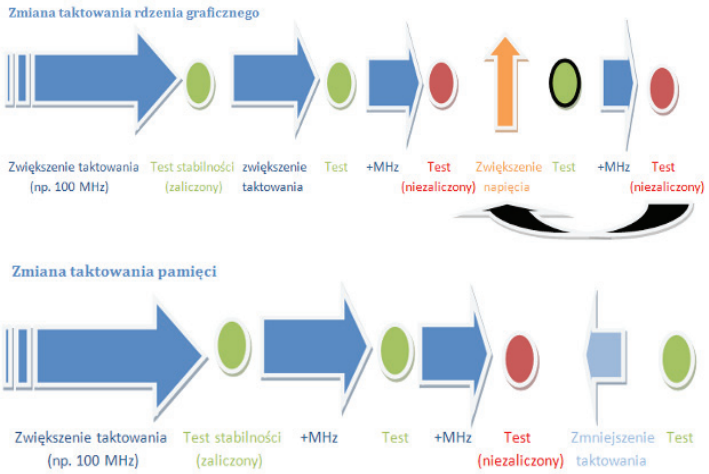
Możemy wyróżnić dwie metody „podkręcania pamięci RAM": zmniejszenie timingów (skracanie czasu dostępu procesora do poszczególnych komórek pamięci), zwiększenie częstotliwości taktowania kości RAM, co zwiększa przepustowość pamięci. Przy wyższych taktowaniach wymagana jest zwiększenie wartości opóźnień. Najlepiej jest zbalansować obie metody, choć zdarza się, że samo zmniejszenie opóźnień daje zadowalający efekt.

Pamięci DDR pierwszej, drugiej i trzeciej generacji reagują inaczej na ustawienia timingów i opóźnień, więc najlepiej przetestować dokładnie efektywność ich skonfigurowania. Dodatkowo można podnieść delikatnie napięcie, dla stabilności, większego taktowania i zmniejszenia timingów. Do większości kości RAM informacje na temat maksymalnych wartości napięcia można znaleźć na stronie producenta. Pamięci najlepiej dobierać w zestawy dwóch kości, chyba że płyta obsługuje triple lub quad channel.

1.4. Podsystem grafiki

Overclocking karty graficznej polega głównie na zmianie taktowań rdzenia graficznego oraz pamięci.

Istnieją dwa sposoby przetaktowywania kart graficznych. Pierwszy z nich zakłada, że najpierw podkręca się rdzeń graficzny, a następnie pamięć. Druga zaś polega na zmianie obu wartości na raz. W przypadku utraty stabilności obniża się raz jedną, raz drugą wartość.



Rys. 1. Przetaktowywanie karty graficznej przez: zmianę taktowania rdzenia graficznego (na górze), zmianę taktowania pamięci (na dole) [4]

2. Pomiar wydajności

2.1. Procesor

Do testowania wydajności procesora zostało wykorzystane oprogramowanie QwikMark. Jest to test pozwalający porównać szybkość działania komputerów poprzez zbadanie szybkości rozwiązywania gęstych układów liniowych. Wynik podawany jest w postaci liczby operacji zmiennoprzecinkowych na sekundę (FLOPS – floating point operations per second).

2.2. Ram

Do testowania pamięci RAM został użyty program Everest Ultimate Edition. Everest Ultimate Edition jest wszechstronnym narzędziem diagnostycznym. Umożliwia przeprowadzenie testów odczytu i zapisu do pamięci RAM oraz testów mierzących opóźnienie pamięci. Program posiada również podstawowe zestawy benchmarków, sprawdzających wydajność procesora, karty graficznej oraz dysków.

2.3. Podsystem grafiki

3DMark Vantage posłużył do testów wydajności kart graficznych. Program jest użyteczny zwłaszcza w zakresie kontroli wydajności wyświetlania grafiki trójwymiarowej. Wyniki testów prezentowane są w przejrzystej formie, który można porównać z innymi wartości zawartymi w dostępnej w programie bazie danych

użytkownikami komputerów z całego świata. Program stanowi również doskonały sprawdzian stabilności działania wszystkich podzespołów komputera, gdyż znacznie obciąża jego zasoby.

2.4. Stabilność zestawu

Do sprawdzenia stabilności zestawu podzespołów komputera posłużył program OCCT (OverClock Checking Tool). Oprogramowanie OCCT wymusza obciążenie wszystkich podzespołów, co pozwala na kontrolę maksymalnego pobór mocy przetaktowanego zestawu. Program daje jednocześnie możliwość monitorowania właściwości pracy podstawowych komponentów takich jak: poziom wykorzystania mocy procesora i pamięci RAM, napięcie, temperaturę poszczególnych rdzeni procesora i jednostki GPU.

3. Eksperyment

W trakcie wszystkich pomiarów wyłączono tryby oszczędzania energii, zarówno w systemie jak i w BIOS, tryb turbo procesora, w celu ujednolicenia wpływu ustawień zewnętrznych na wyniki pomiaru poboru mocy. Zużycie energii mierzono trzykrotnie przenośnym miernikiem wbudowanym we wtyczkę zasilania. Pomiary prowadzono w 3 konfiguracjach, gdy:

- nie są otwarte żadne aplikacje (pulpit);
- otwarta jest przeglądarka internetowa z kilkoma zakładkami, w tym z edytorem tekstowym online;
- przy włączonym teście power supply w programie OCCT z ustawieniami standardowymi.

3.1. Pomiar przy ustawieniach standardowych (fabrycznych)

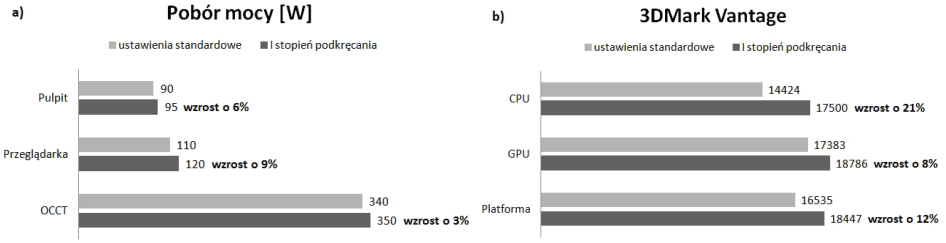
W trakcie testów zastosowano następującą konfigurację elementów zestawu: CPU – 2,8 GHz, 1,4 V; HT – 2 GHz, 1,2 V; NB – 2 GHz, 1,2 V; GPU – 880, 950, 1800, 1,062 V; RAM – 1600 MHz, 1,5 V; HTT – 200 MHz.

Modele testowanego procesora stosowano z niższym napięciem niż testowany egzemplarz, co wpływało pozytywnie na pobór mocy. Dzielnik pamięci RAM został ręcznie przestawiony na taktowanie o wartości 1600 MHz (ustawione typowo w BIOS 1333 MHz). Kości fabryczne są przystosowane do pracy na 1866MHz. Poniżej zamieszczono wyniki przeprowadzonych testów:

- pobór mocy: pulpit 90W; przeglądarka 110W; OCCT 340W;
- QwikMark: 62 GFLOPS;
- RAM: write 6471MB/s; read 8334MB/s; copy 9828MB/s; 48ns;
- 3DMark Vantage: CPU 14424; GPU 17383; platforma 16535.

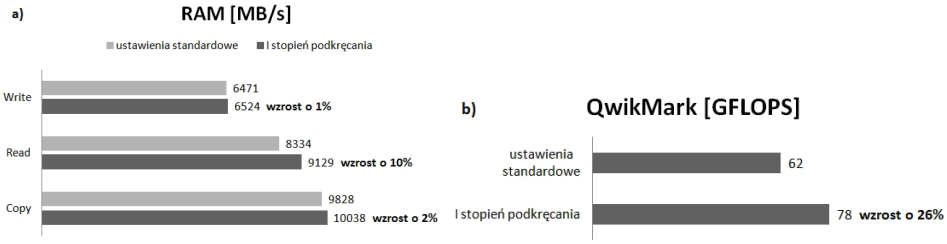
3.2. Pomiar po pierwszymstopniu przetaktowania

W trakcie testów zastosowano następującą konfigurację elementów zestawu: CPU – 3,43 GHz, 1,4 V; HT – 1,96 GHz, 1,2V; NB – 2,2 GHz, 1,2 V; GPU – 910, 975, 1820, 1,062 V; RAM – 1632 MHz, 1,5V; HTT – 245 MHz.



Rys. 2. Funkcjonowanie platformy po 1 fazie przetaktowywania: a) moc; b) wydajność

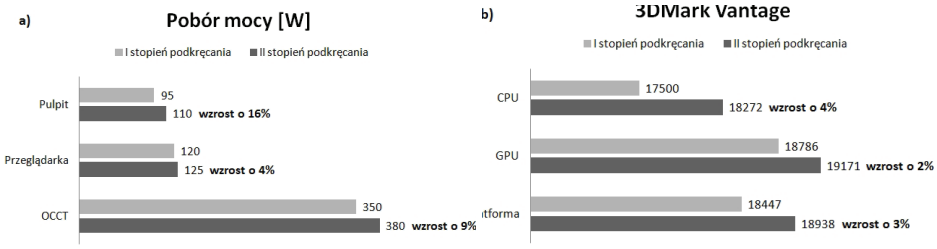
Założeniem testu nr 1 był overclocking bez zmian napięcia. Najwydajniejszy procesor z serii, z której pochodził testowy model, posiadał taktowanie 3,3 GHz. W pierwszej fazie testu został podkręcony do ponad 3,4 GHz. Wymagało to zwiększenia taktowania głównej magistrali do 245 MHz, a więc również zmniejszenia wartości mnożników pozostałych komponentów czyli dzielnika mostka NB, magistrali HT oraz pamięci RAM. W efekcie ustawienia te dały przyrost mocy obliczeniowej oraz wzrost pobieranej energii (rys. 2). Procesor działał o około 20% wydajniej (rys. 3b). Subiektywnie odbierane działanie komputera stało się „wyraźnie szybsze".



Rys. 3. Wydajność: a) pamięci RAM (Everest); b) procesora (QwikMark)

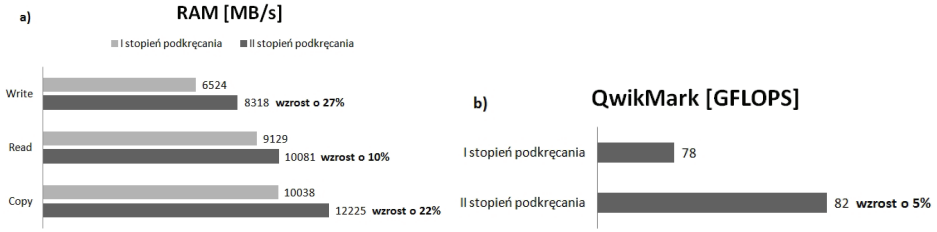
3.3. Pomiar po drugim stopniu przetaktowania

W trakcie testów zastosowano następującą konfigurację elementów zestawu: CPU – 3,61GHz, 1,472V; HT – 2,06GHz, 1,2V; NB – 2,32GHz, 1,22V; GPU – 920, 1000, 1840, 1,087V; RAM – 1720MHz, 1,5V; HTT – 258MHz.



Rys. 4. a) Wykres poboru mocy; b) Zmiana wydajności (3DMark Vantage)

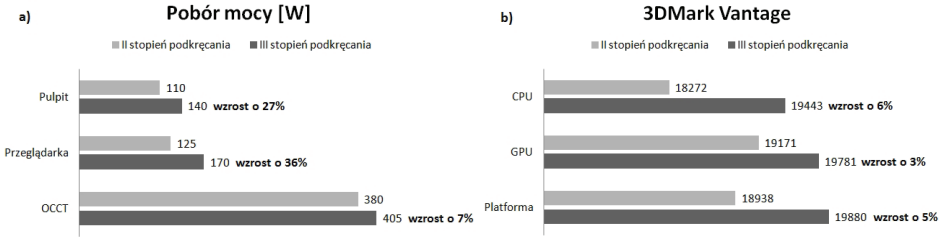
Celem drugiego testu było przetaktowanie z jednoczesną minimalizacją zmian napięcia. Procesor taktowany zegarem 3,6 GHz pracował stabilnie dopiero przy napięciu 1,472 V. To dość dużo, wobec podniesienia częstotliwości tylko o 180 MHz w stosunku do taktowania z poprzedniego testu. Podniesienie napięcia jest uzasadnieniem wzrostu poboru mocy w systemie Windows. Podczas przeglądania stron internetowych i maksymalnego obciążenia przyrost jest mniejszy (rys. 4a). Mała zmiana taktowań nie daje większych rezultatów we wzroście wydajności (rys. 4b, 5b).



Rys. 5. Wydajność: a) pamięci RAM (Everest); b) procesora (QwikMark)

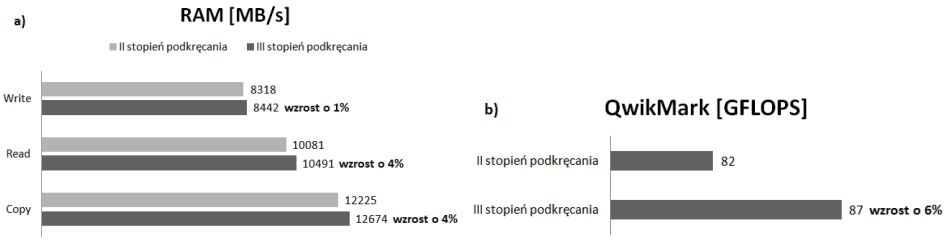
3.4. Pomiar po trzecim stopniu przetaktowania

W trakcie testów zastosowano następującą konfigurację elementów zestawu: CPU – 3,80GHz, 1,562V; HT – 2,44GHz, 1,2V; NB – 2,72GHz, 1,32V; GPU – 945, 1018, 1890, 1,087V; RAM – 1812MHz, 1,6V; HTT – 272MHz.



Rys. 6. a) Wykres poboru mocy; b) Zmiana wydajności (3DMark Vantage)

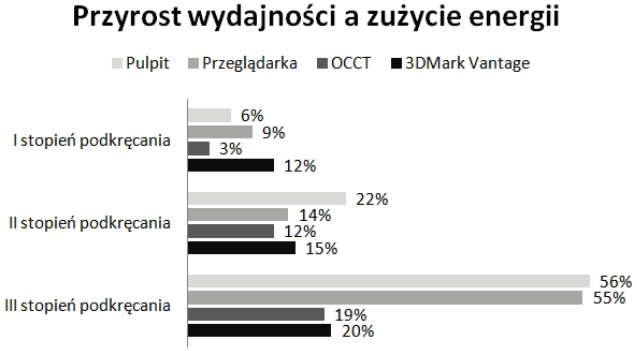
Zwiększenie taktowania procesora o kolejne 200 MHz podniosło napięcie aż do wartości 1,562 V. Wzrost mocy był największy spośród dotychczas zarejestrowanych (rys. 6a). Stabilność pracy wymagała zwiększenia napięcia pamięci RAM i kontrolera pamięci. W parze z kolejnym znaczącym zwiększeniem taktowania nie dało już tak znaczących wyników jak poprzednio (rys. 7a). Uzyskany czas oczekiwania pamięci wyniósł jedynie 41 ns. Wzrost wydajności „grafiki" był kilkuprocentowy w stosunku do poprzedniego testu (15% względem ustawień standardowych).



Rys. 7. Wydajność: a) pamięci RAM (Everest); b) procesora (QwikMark)

4. Podsumowanie

Na rysunku 8 zamieszczono wyniki testów. Czarny słupek odnosi się do względnego wzrostu zużycia energii. Podobnie jak słupki przyrostu wydajności prezentuje on zmiany wartości w stosunku do ustawień standardowych.



Rys. 8. Zależność przyrostu wydajności od zużycia energii

Przy zwiększaniu taktowania bez zmian napięcia pobór mocy we wszystkich trzech próbach wzrasta nieznacznie. Wzrost wydajności jest już odczuwalny. W każdej z trzech konfiguracji parametrów najmniejszy wzrost zużycia energii odnotowano w czasie największego obciążenia. W trakcie mniejszego i średniego obciążenia zużycie energii było znaczne.

Maksymalne stabilne podkręcanie wraz ze zwiększeniem napięć jest opłacalne tylko wtedy, gdy potrzebujemy bezwzględnej wydajności, np. edycja wideo. Podczas zwykłej pracy wzrost wydajności nie będzie odczuwalny dla użytkownika. Przy takich ustawieniach parametrów sprzętu i systemu, korzystając z komputera 4 godziny dziennie i cenie za kWh energii równej 0,60 zł, podniesie to koszt o około 50 zł rocznie. Korzystne jest przetaktowywanie bez zwiększania napięć. Wyżej taktowane modele tych samych serii różnią się najczęściej tylko mnożnikiem, więc zwiększając taktowanie i w niewielkim stopniu napięcie, otrzymuje się praktycznie za darmo wydajność większą od fabrycznej bez ponoszenia dodatkowego ryzyka.

Bibliografia

[1] Danowski B.: Tuning, wyciszenie i overclocking komputera PC. Gliwice, Helion, 2003.
[2] Danowski B: Zwiększ moc swojego komputera czyli 101 sposobów na poprawę wydajności. Gliwice, Helion, 2004.
[3] Gajewski P., Hałas S., Krężel S., Wyrzykowski A.: Przyspiesz swój komputer. PC Format. Krok Po Kroku, Warszawa, wyd. Bauer, 2010.
[4] Samołyk D.: Praktyka – podkręcamy!. http://overclock.pl/articles/show/id/326,4,6,praktyka-podkrecamy, 24.05.2014.
[5] Sokół R.: Podkręcanie procesorów. Gliwice, Helion, 2005.

Zeszyty Naukowe WSEI seria: TRANSPORT I INFORMATYKA, 4(1/2014), s. 29-37

Marcin JANOWSKI

Wyższa Szkoła Ekonomii i Innowacji w Lublinie, e-mail: janowski.m@gmail.com

PODSTAWY KONFIGURACJI SERWERA WWW W SYSTEMIE LINUX

CONFIGURATION BASICS OF LINUX WEB SERVER

Streszczenie

W obecnych czasach człowiek używający słowa Internet ma na myśli stronę WWW, dlatego tym referacie postaramy się opisać podstawową konfigurację serwera HTTP, opierając się o kilkuletnie doświadczenia z konfiguracją serwerów. Chcielibyśmy przybliżyć Państwu budowę nowoczesnego, bezpiecznego i wydajnego serwera WWW z wykorzystaniem systemu operacyjnego Linux, serwera Nginx, parsera PHP i serwera bazodanowego MySQL. Całe oprogramowanie wykorzystywane w tym referacie wydawane jest na licencjach typu OpenSource, co oznacza że kod źródłowy oprogramowania jest ogólnodostępny.

Słowa kluczowe: Linux, serwer www, php, mysql, http

Summary

Nowadays, almost everybody associates Internet with WWW services. Therefore this paper will present basic configuration of HTTP server. It results from server configuration carried out through several years. The presentation of modern, secure and efficient Linux WWW server will include Nginx server, PHP parser and MySQL database server. All software described and used in this paper is issued with OpenSource license. It follows that source code is generally available and may be freely used and modified.

Keywords: Linux, server www, php, mysql, http

1. Podstawowa konfiguracja systemu Linux

Przed rozpoczęciem każdej nowej instalacji serwera, należy zaplanować co i w jaki sposób chcemy osiągnąć. Tego punktu nie należy bagatelizować, gdyż w przypadku nie przemyślanych działań na początku lub źle zaplanowanych, prędzej czy później na produkcyjnym serwerze (czyli takim, który obsługuje już ruch) będzie trzeba wykonać niezwykle ciężkie i mozolne prace, które będą się wiązać z długim czasem niedziałania serwera.

Podstawową sprawą którą należy przemyśleć jest oczywiście rozkład partycji na dysku twardym. Najgorszą rzeczą która może nas spotkać to utrata danych związana z brakiem macierzy dyskowej. Z tego powodu bardzo ważne jest aby serwer który konfigurujemy miał dwa dyski twarde o takiej samej pojemności. Umożliwi to stworzenie macierzy dyskowej typu RAID1, dzięki której dane będą przetrzymywane na dwóch dyskach twardych. Oznacza to, że w przypadku padnięcia JEDNEGO dysku twardego, nasze dane zostaną w całości zachowane. Dopiero utrata dwóch dysków twardych spowoduje stratę danych [3].

Kolejną kwestią bezpieczeństwa danych jest wybór odpowiedniego systemu plików. W przypadku wyboru niestabilnego i nieprzetestowanego systemu plików, bardzo możliwe jest uszkodzenie partycji, co w drastycznych przypadkach może prowadzić do niemożliwości odzyskania danych! System plików który zalecam to ext4, jest to wiodący system plików w Linuksie, który jeszcze ani razu mnie nie zawiódł. System ten posiada „księgowanie" (ang. *journaling*), który zapisuje bieżące pliki w specjalnym dzienniku. Jeśli naszemu serwerowi nagle zabraknie prądu, przy odrobinie szczęścia nie utracimy żadnych danych, oprócz wspomnianego już dziennika. PrPodczas instalacji będziemy proszeni o podanie hasła do nowo utworzonego użytkownika root dla serwera MySQL. Domyśle ustawienia serwera pozwolą na dostęp do niego na adresie loopback (domyślnie 127.0.0.1) i porcie 3306.zy wyborze systemu plików musimy również używać najnowszej, stabilnej wersji jądra Linux, ponieważ to w jądrze są zaprogramowane systemy plików. W przypadku użycia niestabilnej wersji jądra, możemy się natknąć na błąd w systemie plików wynikający z nowej funkcjonalności, który nie został jeszcze poprawiony przez programistów.

Po wybraniu systemu plików który chcemy użyć, należy zaprojektować rozkład partycji. Główny system plików Linuksa dzieli się na wiele katalogów, najważniejszymi z naszego punktu widzenia są: / – główny system plików w Linux, każdy folder jest jego potomkiem, /boot – zawiera obrazy jądra, /home – zawiera wszystkie pliki użytkowników, stron, dane; z tego powodu powinna być to największa partycja, /tmp – partycja plików tymczasowych, /var – zawiera między innymi logi aplikacji. Oznacza to, że musimy stworzyć 5 partycji, pamiętając przy tym aby zaplanować je tak aby na żadnej nie zabrakło miejsca. Dla partycji na której pod montowany zostanie / należy przeznaczyć około 10GB, będą na niej przetrzy-

mywane pliki systemowe, ustawienia, pliki użytkownika root. Na partycję /boot należy przeznaczyć 500MB, będą na niej pliki obrazu init i ustawienia bootloader, służące do bootowania systemu. Dla /tmp należy przeznaczyć od 1GB do 10GB, w zależności od pojemności macierzy dyskowej. Dla punktu montowania /var wymagane jest miejsce od 5GB, jest to katalog w którym przetrzymywane są między innymi logi systemowe, i to od nas zależy ile logów chcemy przecPodczas instalacji będziemy proszeni o podanie hasła do nowo utworzonego użytkownika root dla serwera MySQL. Domyśle ustawienia serwera pozwolą na dostęp do niego na adresie loopback (domyślnie 127.0.0.1) i porcie 3306.howywać. Ostatnim katalogiem który wymaga osobnej partycji jest /home, powinniśmy przeznaczyć na niego pozostałe wolne miejsce, ponieważ jest to katalog przechowujący dane użytkowników i stron WWW, i z tego powodu powinien być największy [2].

Systemy z rodziny Linux rozprowadzane są w tak zwanych dystrybucjach, na które składa się jądro Linuksa, bootloader, podstawowe programy i ustawienia. Nie da się używać samego Linuxa (jako jądra), aby mieć możliwość naprawdę podstawowej pracy na Linuksie potrzebny jest zbiór narzędzi GNU Linux i bootloader.

Najpopularniejszymi dystrybucjami serwerowymi są Debian, CentOS wywodzący się z Red Hat, Ubuntu wywodzące się z Debiana i Red Hat. Ja na serwerze WWW zawsze używam Ubuntu, ponieważ w porównaniu do Debiana, Ubuntu ma o wiele nowsze oprogramowanie, a na serwerach Webowych użytkownicy wymagają nowych wersji oprogramowywania, np. PHP.

Ubuntu wydawane jest w dwóch liniach wydawniczych – zwykłej, co pół roku i LTS, czyli Wydłużony Czas Wsparcia (ang. *Long Time Support*). Wersja LTS wydawana jest co 2 lata, i jej czas wsparcia wynosi 5 lat, a w przypadku wersji zwykłej tylko 9 miesięcy. Aktualnie najnowszą wersją Ubuntu Server LTS jest wersja 14.04, i w oparciu o tą dystrybucję będę przedstawiał konfigurację serwera.

Po prawidłowym zainstalowaniu dystrybucji należy zacząć od instalacji pakietów które są nam potrzebne do uruchomienia serwera, oto one:

vim – edytor tekstu
aptitude – nakładka na manager pakietów apt-get
software-properties-common – narzędzie do prostego dodawania repozytoriów PPA
nginx-light – serwer http w wersji light
php5-fpm – demon php5
php5-mysqlnd – sterownik bazy danych MySQL dla php5

Przed instalacja powyższych pakietów należy dokonać aktualizacji repozytoriów pakietów, za pomocą polecenia:

```
apt-get update
```

Po wykonaniu aktualizacji repozytoriów można już zainstalować powyższe pakiety:

```
apt-get install vim aptitude software-properties-↪common
nginx-light php5-fpm php5-mysqlnd
```

2. Kompilacja kernela

Aby nasz serwer spełniał standardy bezpieczeństwa, musimy skompilować własną wersję kernela Linux z patchem GrSecurity. Patch ten nie zezwala na wykonanie większości exploitów na kernelu z błędami bezpieczeństwa. Ponadto, zapewnia podniesienie bezpieczeństwa lokalnego, za pomocą np. separacji /proc – dzięki temu, inni użytkownicy serwera nie będą widzieć włączonych procesów innych użytkowników.

Pierwszą rzeczą jaką należy zrobić to pobrać na serwer najnowszą wersję testową GrSecurity ze strony projektu http://grsecurity.net/, w moim przypadku jest to grsecurity-3.0-3.14.8-201406191347.patch, czyli patch GrSecurity dla kernela 3.14.8.

Następnie ze strony https://www.kernel.org/ trzeba pobrać na serwer źródła kernela o tej samej wersji dla której został pobrany patch, czyli w kernel 3.14.8. Oba pliki można pobrać przy pomocy polecenia: wget ADRES_URL_PLIKU, a pobraną paczkę z kernelem należy rozpakować używając polecenia: tar -xf linux-3.14.8.tar.xz.

Do skompilowania potrzebujemy pakietów kompilatora i bibliotek, instalujemy je w systemie za pomocą polecenia:

```
apt-get install build-essential kernel-package fakeroot ↪libncurses5-dev
```

Aby kod patchu GrSecurity pojawił się w źródłach kernela, powinniśmy go do nich wgrać. Służy do tego komenda patch, którą trzeba wywołać z poziomu katalogu z rozpakowanymi źródłami kernela:

```
patch -p1 < ../grsecurity-3.0-3.14.8-201406220132.patch
```

Ponieważ kernel ma wiele opcji konfiguracyjnych (w tej wersji 8093!) dla początkujących polecam pobrać konfigurację kernela stworzoną przez deweloperów Ubuntu Server, gdzie w adresie URL za XXX należy podstawić amd64 dla architektury procesora 64bit, lub i386 dla architektury 32bit:

```
wget http://kernel.ubuntu.com/~kernel-↪ppa/configs/trusty/
XXX-config.flavour.↪generic -O .config
```

Minusem takiej konfiguracji jest dość długi czas kompilacji i duży rozmiar kernela, spowodowane jest to tym że domyślna konfiguracja kernela ma włączone opcje dla każdego popularnego sprzętu. Przy tworzeniu własnej konfiguracji oczywiście możemy zaznaczyć tylko opcje dotyczące tylko urządzeń które są w naszym serwerze.

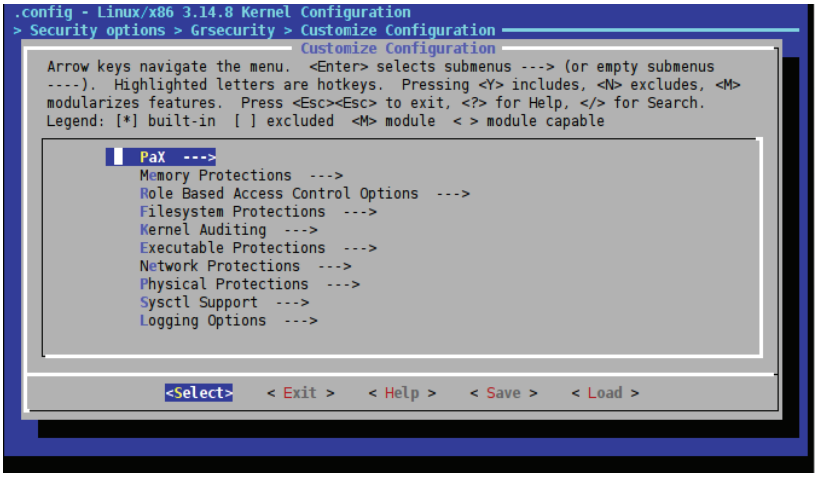
Minusem takiej konfiguracji jest dość długi czas kompilacji i duży rozmiar kernela, spowodowane jest to tym że domyślna konfiguracja kernela ma włączone opcje dla każdego popularnego sprzętu. Przy tworzeniu własnej konfiguracji oczywiście możemy zaznaczyć tylko opcje dotyczące tylko urządzeń które są w naszym serwerze.

Aby otworzyć menu konfiguracji kernela, trzeba użyć komendy: make menuconfig. W pobranej konfiguracji nie było wybranych ustawień dotyczących patcha GrSecurity, aby wejść w ich menu trzeba wybrać Security options > Grsecurity > zaznaczyć Grsecurity (NEW) za pomocą spacji, a w podmenu które się pokazało należy wejść do Customize Configuration. Aby nasz kernel został prawidłowo skonfigurowany dla serwera WWW zalecam wybrać poniższe opcje:

- Memory Protections: Disable privileged I/O, Deter exploit bruteforcing, Hide kernel symbols, Active kernel exploit response
- RBAC: Hide kernel process
- Filesystem Protections: Proc restrictions > Restrict /proc to user only, Additional restrictions, Linking restrictions, Sysfs/debugfs restriction
- Kernel Auditing: Fork failure logging
- Executable Protections: Dmesg(8) restriction, Trusted Path Execution (TPE) > Partially restrict all non-root users

Po wybraniu opcji podanych wyżej, za pomocą przycisku < Exit > trzeba wyjść z konfiguratora, a na pytanie o zapisanie zmian należy odpowiedzieć twierdząco. Gdy konfiguracja jest już gotowa, nie pozostało nam nic innego jak rozpocząć kompilację kernela:

```
make -j `getconf _NPROCESSORS_ONLN` deb-pkg ↪LOCALVERSION=-custom
```



Rys 1. Okno konfiguracji kernela, podmenu GrSecurity

Po zakończeniu się kompilacji, która może trwać nawet i parę godzin, w katalogu wyżej znajdą się pakiety .deb z skompilowanym kernelem. Aby je zainstalować trzeba użyć poolecenia dpkg:

```
dpkg -i *.deb
```

Nowy kernel zostanie uruchomiony po ponownym uruchomieniu serwera. [4]

3. Serwer MySQL

Serwer MySQL jest jednym z najpopularniejszych, otwarto źródłowych serwerów baz danych używanych w hostingu stron WWW.

W tym referacie opiszę użycie forka MySQL – MariaDB. Powstał on po wykupieniu przez Sun Microsystems firmy rozwijającej MySQL – MySQL AB.

Aby dodać repozytorium z pakietami MariaDB i podpisać je kluczem należy wykonać komendę:

```
apt-key adv --recv-keys --keyserver
↪hkp://keyserver.ubuntu.com:80 0xcbcb082a1bb943db
add-apt-repository 'deb
↪http://mirrors.coreix.net/mariadb/repo/10.0/ubuntu trusty
main'
apt-get update
```

Instalacja serwera następuje po wykonaniu polecenia:

```
apt-get install mariadb-server [5]
```

Podczas instalacji będziemy proszeni o podanie hasła do nowo utworzonego użytkownika root dla serwera MySQL. Domyśle ustawienia serwera pozwolą na dostęp do niego na adresie loopback (domyślnie 127.0.0.1) i porcie 3306.

4. Interpreter PHP

Język PHP jest obecnie najczęściej wykorzystywany przy tworzeniu stron WWW z dynamiczną zawartością. W instalacji którą przedstawiam użyję demona php5-fpm, który pozwala na oddzielne instancje procesów PHP dla każdego z użytkowników. Podnosi to bezpieczeństwo i możliwości konfiguracji per użytkownik. Polecenie do instalacji php5-fpm zostało podane w rozdziale 1.

Głównym plikiem konfiguracyjny php5-fpm jest /etc/php5/fpm/php-fpm.conf. Na samym jego końcu znajduje się linia:

```
include=/etc/php5/fpm/pool.d/*.conf
```

Informuje ona php5-fpm, że konfiguracji pooli (osobnych instancji php5-fpm) trzeba szukać w plikach o rozszerzeniu *.conf w katalogu /etc/php5/fpm/pool.d/.

Domyślnie włączony jest jeden tzw. pool o nazwie www. Jego konfiguracja znajduje się w pliku /etc/php5/fpm/pool.d/www.conf . Dla naszego zastosowania można go usunąć, lecz jest on dobrą dokumentacją konfiguracji php ponieważ zawiera komentarze do każdej z opcji.

Każdy użytkownik powinien mieć stworzony osobny pool php5-fpm, za pomocą osobnego pliku .conf w katalogu /etc/php5/fpm/pool.d/, dla zachowania przejrzystości konfiguracji nazwa pliku powinna odpowiadać nazwie użytkownika którego dotyczy.

Nowego użytkownika można stworzyć za pomocą polecenia adduser NAZWA_USERA, np.:

```
adduser janek
```

Następnie można stworzyć plik konfiguracyjny poola dla użytkownika janek / etc/php5/fpm/pool.d/janek.conf o zawartości:

```
[janek]
listen = /var/run/janek.php-fpm.socket
listen.allowed_clients = 127.0.0.1
listen.owner = www-data
listen.group = www-data
listen.mode = 0660
user = janek
group = janek
pm = dynamic
pm.max_children = 5
pm.start_servers = 1
pm.min_spare_servers = 1
pm.max_spare_servers = 1
pm.max_requests = 500
php_admin_flag[log_errors] = on
php_admin_flag[display_errors] = on
```

Pierwsza linia „[janek]" oznacza nazwę poola w nawiasach kwadratowych, musi być ona unikalna. W drugiej linii znajduje się ścieżka do socketu który ma utworzyć php5-fpm. To za jego pomocą odbywa się komunikacja pomiędzy php5 a nginxem. Tak jak nazwa poola, ścieżka musi być unikalna. Linie na które powinniśmy jeszcze zwrócić uwagę to linie 7 i 8. Jest to nazwa użytkownika i grupa, na którego prawach php powinno zostać uruchomione.

Po dodaniu powyższego pliku, trzeba przeładować demona php5-fpm za pomocą polecenia:

```
service php5-fpm reload [6]
```

Jeśli wszytko poszło poprawnie, polecenie ps aux | grep php powinno zwrócić procesy php, w tym proces janka.

5. Serwer Nginx

Oprócz serwera baz danych i interpretatora PHP potrzebny jest jeszcze serwer WWW który wyświetli wygenerowany kod. Dziś w modzie ze względu na bardzo wysoka wydajność jest serwer Nginx, którego instalacja została przedstawiona w rozdziale 1.

Przed przystąpieniem do konfiguracji należy utworzyć użytkownikowi janek katalog przeznaczony do przechowywania stron WWW, można go stworzyć za pomocą komendy:

```
mkdir -p /home/janek/www/ADRES.STRONY/htdocs;
chown -R janek:janek ↪/home/janek/www/ADRES.STRONY/htdocs;
```

W pliku /etc/nginx/sites-enabled/janek-ADRES.STRONY przechowywana jest konfiguracja strony która ma działać pod zdefiniowanym adresem:

```
server {
        server_name ADRES.STRONY;
         root /home/janek/www/ADRES.STRONY/htdocs;
         location ~ \.php$ {
               include /etc/nginx/fastcgi_params;
               fastcgi_index index.php;
               fastcgi_pass unix:/var/run/janek.php-↪fpm.socket;
               fastcgi_param SCRIPT_FILENAME
↪$document_root$fastcgi_script_name;
    }
}
```

W drugiej linii za ADRES.STRONY należy podać nazwę domenowa strony pod którą dana strona ma działać. Dzięki temu Nginx wie jaką zawartość ma hostować dla danej domeny. W następnej linii należy podać ścieżkę do katalogu, który zawiera pliki z naszą stroną WWW. W linii 4 rozpoczyna się blok konfiguracji, który definiuje ustawienia konfiguracji dla plików z rozszerzeniem.php. W linii 7 musi znajdować się ścieżka do socketu poola php5-fpm, który chwilę temu skonfigurowaliśmy.

Po utworzeniu pliku konfiguracyjnego, należy przeładować serwer Nginx poleceniem:

```
service nginx reload;
```

Po przeładowaniu serwera, pod adresem domeny powinna działać nasza strona internetowa [1].

Bibliografia

[1] Fjordvald M.: Instant Nginx Starter. Birmingham, Packt Publishing Ltd, 2013.

[2] Hill B. M., Rankin K.: Ubuntu Serwer. Oficjalny podręcznik. Gliwice, Helion, 2011, s. 54-58.

[3] Hill B. M., Rankin K.: Ubuntu Serwer. Oficjalny podręcznik. Gliwice, Helion, 2011, s. 331-355.

[4] https://wiki.ubuntu.com/KernelTeam/GitKernelBuild, GitKernelBuild, dostęp 30.06.2014

[5] https://downloads.mariadb.org/mariadb/repositories/#mirror=kisiek&distro_ release=precise&version=10.0, Setting up MariaDB Repositories, dostęp 30.06.2014.

[6] http://www.php.net/manual/en/install.fpm.configuration.php, PHP Manual Installation and Configuration FastCGI Process Manager (FPM), dostęp 30.06.2014.

Zeszyty Naukowe WSEI seria: TRANSPORT I INFORMATYKA, 4(1/2014), s. 39-47

Grzegorz TODRYK

Wyższa Szkoła Ekonomii i Innowacji w Lublinie, e-mail: grtod@po.pl

ARCHITEKTURA APLIKACJI INTERNETOWYCH

LINUX APPLICATIONS ARCHITECTURE

Streszczenie

Celem niniejszego artykułu jest przedstawienie architektury aplikacji internetowych (webowych), sposobu ich działania i zasad projektowania. Artykuł obejmuje zagadnienia związane z zasadą działania sieci internet, rodzajami technologii i języków programowania używanych przy tworzeniu aplikacji internetowych, a także używanych wzorców projektowych i szkieletów aplikacji. Jako przykład podano fragmenty kodu napisanego w języku PHP w Zend Framework.

Summary

This paper aims at presentation of internet applications architecture, their operation and design principles. Presentation starts with review of Internet operation. Next short presentation of useful technologies and programming languages follows. Also design templates and application frameworks are briefly described. Practical code examples using PHP language and Zend framework form original part of the paper.

Słowa kluczowe: aplikacje internetowe, wzorce projektowe, szkielety aplikacji

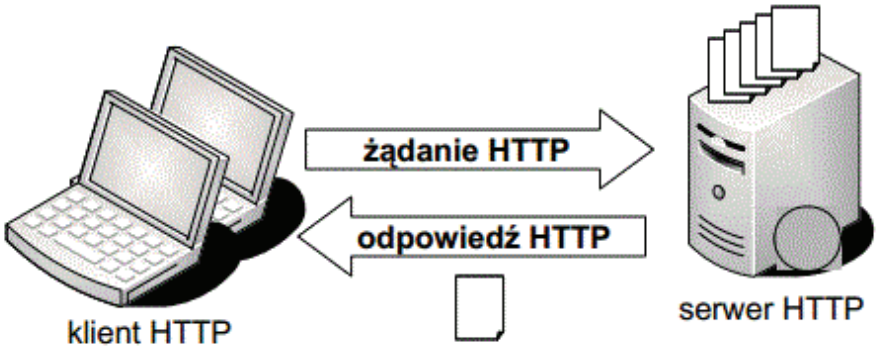
Keywords: Internet applications, design templates, applications frameworks

1. Wprowadzenie

Sieć WWW, nazywana także „światową pajęczyną", jest wszechobecna w naszym codziennym życiu, a większość komputerów i wszelkiego typu skomputeryzowanych urządzeń jest aktualnie podłączona do internetu. Z kolei przyszłość zapowiada się jeszcze bardziej „usieciowiona", przewiduje się istnienie „internetu rzeczy" (ang. *Internet of Things*).

Sieć web (WWW) posiada typową architekturę rozproszoną typu klient – serwer. Komunikacja pomiędzy tymi warstwami jest realizowana za pośrednictwem protokołu HTTP. Serwer jest programem nieprzerwanie pracującym, obsługującym repozytorium dokumentów (np. HTML), które udostępnia sieciowym klientom. Klient HTTP jest programem użytkowym, który odpowiada za wysyłanie żądań pobrania dokumentów, wizualizację pobieranych dokumentów oraz obsługę interakcji z użytkownikiem końcowym.

Protokół HTTP jest protokołem bezstanowym. Klient łączy się z portem (zazwyczaj 80) i przesyła żądanie przesłania informacji z serwera. Żądanie to jest analizowane i obsługiwane. Następnie serwer przesyła odpowiedź zawierającą żądane informacje, bądź komunikat o błędzie (w przypadku, gdy żądanie było nieprawidłowe lub niemożliwe do spełnienia). Po wykonaniu tych działań połączenie jest zamykane i serwer powraca do fazy nasłuchiwania kolejnych żądań od klientów.



Rys. 1. Architektura klient – serwer HTTP

Podstawową technologią służącą do tworzenia dokumentów WWW jest język HTML, który jest odpowiedzialny za poinformowanie przeglądarki w jaki sposób ma być prezentowany tekst oraz wszelkie inne elementy umieszczone na stronie. W celu separacji treści stron internetowych zapisanych w języku HTML od sposobu ich prezentacji stosowany jest mechanizm CSS (*Cascade Style Sheet*).

2. Dynamiczna zawartość stron internetowych

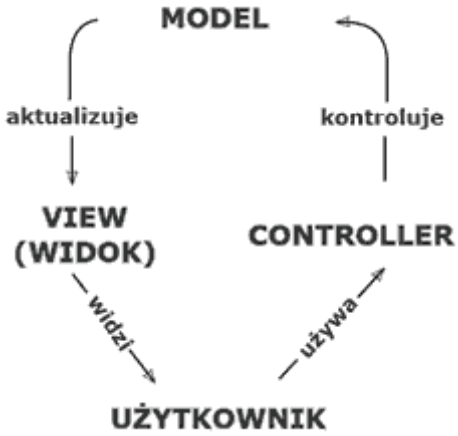
Na początku istnienia internetu, strony internetowe zawierały w większości statyczną zawartość. Można było zmieniać ich treść, natomiast bardzo mało było dynamicznych usług w sieci. Z biegiem czasu pojawiła się konieczność dynamicznego generowania zawartości stron internetowych. Taka dynamiczna zawartość stron może pochodzić z różnych źródeł. Wyszukiwarki oraz bazy danych mogą odpowiadać na zapytania zadane przez użytkownika i wyświetlać dane spełniające kryteria wyboru. Instrumenty pomiarowe mogą przesyłać dane o odczytach (np. temperatura, wilgotność). Kanały informacyjne czy usługi sieciowe mogą zapewniać dostęp do bieżących wiadomości, notowań giełdowych czy wyników sportowych. Dziś zamiast mówić o tworzeniu (budowaniu) stron internetowych, mówi się o tworzeniu aplikacji internetowych (Web application).

Istnieje wiele metod dostępu do danych dynamicznych. Początkowo korzystano z otwartego standardu Common Gateway Interface (CGI). Z czasem powstały alternatywy dla CGI. Microsoft stworzył swój własny system: Active Server Pages (ASP), Sun Microsystems stworzyło Java Server Pages (JSP) oraz serwlety Javy, z kolei w społeczności Apache narodził się język PHP. Język PHP jest aktualnie najpopularniejszym językiem wykorzystywanym do budowy dynamicznych stron internetowych. Ponad 82% stron internetowych (stan: lipiec 2014) jest generowanych z wykorzystaniem tego języka.

3. Architektura MVC (Model – View – Controller)

Podczas tworzenia jednorodnych, ale bardzo skomplikowanych aplikacji, ich kod rozrasta się do rozmiarów nad którymi trudno zapanować. W tym przypadku warto skorzystać z wzorców projektowych, które dzielą kod na fragmenty wykonujące określone funkcje.

Takim najczęściej wykorzystywanym w aplikacjach internetowych wzorcem jest model architektoniczny MVC (ang. *Model – View – Controller*). Pozwala odseparować funkcje systemowe komponentów aplikacji WWW i tym samym w znacznym stopniu upraszcza ich tworzenie.



Rys. 2. Schemat wzorca MVC

Wprowadza on ich podział na trzy niezależne składowe:
- warstwę M (model),
- warstwę V (widok),
- warstwę C (kontroler).

Sercem aplikacji jest warstwa C, która odpowiada między innym i za przetwarzanie żądań HTTP oraz sterowanie przebiegiem wykonania całej aplikacji. Kontroler uzyskuje dostęp do danych zapisanych w bazie danych za pośrednictwem warstwy M. Rolą warstwy M jest dostarczenie kontrolerowi wygodnego interfejsu do komunikacji z bazą danych. Dane pobrane przez kontroler za pośrednictwem warstwy M są formatowane przy użyciu szablonów nazywanych widokami. Przetworzone widoki generują kod HTML, który jest ostatecznie wysyłany do przeglądarki WWW.

Dużą zaletą zastosowania wzorca MVC jest to, że bardzo łatwo można zlokalizować kod wybranych funkcjonalności i go modyfikować, bez potrzeby przedzierania się przez nieczytelne i zagmatwane skrypty. Zastosowanie wzorca MVC ma również kilka wad. Po pierwsze, powstaje bardzo dużo plików. Po drugie, w modelu MVC trzeba przyjąć pewne założenia co do tego, co zdarzyło się przed danym momentem.

4. Framework czyli szkielet aplikacji

Framework to szkielet programu, w konkretnym projekcie wystarczy go tylko wypełnić kodem specyficznym dla danego zadania. Frameworki nie są pojedynczymi bibliotekami, lecz zbiorami odpowiednio dobranych i współpracujących ze sobą fragmentów kodu. Stanowią zręby aplikacji i zawierają występujące w wielu pro-

gramach podsystemy, np. odpowiedzialne za łączenie z bazą danych, sprawdzanie uprawnień, logowanie zdarzeń, obsługę błędów, itd. Frameworki z reguły korzystają z architektury MVC i wykorzystują paradygmat programowania obiektowego.

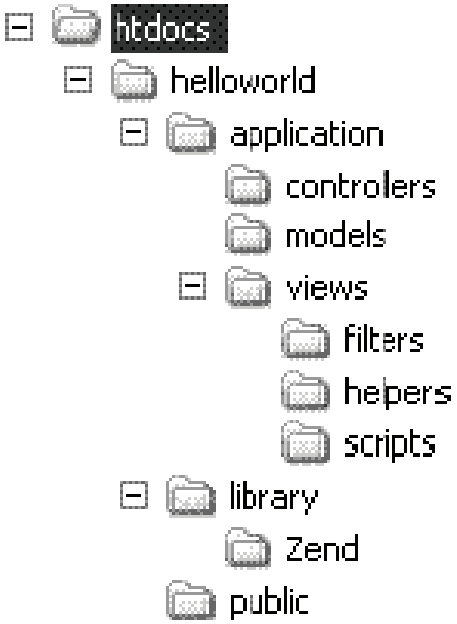
Do najpopularniejszych frameworków należą: Zend Framework, CakePHP, Symfony, CodeIgniter oraz Prado. Wciąż powstają nowe rozwiązania, np. aktualnie popularny staje się framework Laravel. Wybór konkretnego frameworku zależy od wielu czynników np. stopnia komplikacji projektu, preferowanej objętości bibliotek, jakości dokumentacji dla konkretnego rozwiązania, wsparcia społeczności, obsługi wielu baz danych, itd.

5. Przykładowy projekt Zend Framework

Poniżej przedstawiono pliki przykładowego projektu wykonanego w oparciu o Zend Framework. Zend Framework zapewnia pełną implementację wzorca MVC, co pozwala na separację logiki biznesowej od interfejsu użytkownika i modelu danych.

Przykładowy kod Kontrolera:

```
<?php
class IndexController extends Zend_Controller_Action
{
   public function searchAction()
   {
      $request = new Zend_Controller_Request_Http();
      //pobranie danych z widoku
      $text = $request->getParam(,text');
      //wywołanie modelu
      $data = new Application_Model_Data();
      //pobranie danych
      $result = $data->getData($text);
      //przekazanie danych do widoku
      $this->view->data = $result;
   }
}
```



Rys. 1. Struktura aplikacji Zend

Instrukcja: `$page = $request->getParam(‚param’);` pobiera parametr `‚param’` z widoku. Instrukcja `$this->view->var1 = $var2;` przekazuje zmienną `$var2` do widoku jako `var1`. Model jest wywoływany bezpośrednio przez Kontroler i jemu też zwraca odpowiedź na zapytanie do bazy danych.

Przykładowy kod Modelu:

```
<?php
class Application_Model_Data
{
   protected $_config = array ([konfiguracja DB]);

   public function getData($text)
   {
      try
      {
         //połączenie z bazą danych
         $db = Zend_Db::factory(‚pdo_mysql’,
             $this->_config);
         $db->getConnection();
         //konstrukcja zapytania SQL do DB
         $select = new Zend_Db_Select($db);
```

```
            $select = $db->select();
            $select->from (,data');
            $select->where(,Nazwa like \'%'.$text.'%\'');
            //wykonanie zapytania do DB
            $result = $db->fetchAll($select);
            //zwrócenie rezultatu zapytania
            return $result;
         }
         catch (Zend_Db_Exception $e) { //kod błędu; }
      }
   }
```

Model zwraca dane bezpośrednio kontrolerowi (tutaj w postaci tablicy). Przykładowy kod widoku (częściowego):

```
<h3>Wyszukiwanie rekordów</h3>
<form name=”search” method=”POST”>
   <p> Fraza: </p>
   <input name=”text” type=”text” size=”10”>
  <input name=”submit” type=”submit” value=”szukaj”> </form>
<?php $ile = count($this->data);?>
<?php if($ile != 0):?>
   <p>ilość znalezionych pozycji:<?=$ile?></p>
   <table>
      <tr>
         <th>Lp.</th>
         <th>Numer:</th>
         <th>Nazwa rekordu:</th>
      </tr>
      <?php $i=0?>
    <?php foreach ($this->data as $row):?>Linux applications architecture
         <tr>
            <td><?= $i+1?></td>
            <td><?= $row[,Numer']?></td>
            <td class=”title”><?= $row[,Nazwa']?></td>
         </tr>
      <?php endforeach;?>
   </table>
<?php else:?>
   <p>brak wyników wyszukiwania</p>
<?php endif;?>
```

Widok jest w zasadzie typowym szablonem w języku HTML, który zawiera tagi PHP (<?php ?> i <?= ?>). Otrzymuje on dane z kontrolera poprzez konstrukcję: `$this->zmienna`. Widok zawiera formularz, z którego dane są przesyłane do Kontrolera motodą POST. Formularze w Zend Framework można oczywiści two-

rzyć za pomocą odpowiednich klas, natomiast tutaj został przedstawiony typowy formularz w formacie HTML.

6. Podsumowanie

Użycie modelu MVC oraz frameworków (obojętne gotowych czy własnych) jest dzisiaj powszechne przy tworzeniu aplikacji webowych. Wzorzec MVC jest zalecany dla aplikacji o średniej i dużej skali komplikacji i jest powszechnie używany przy tworzeniu aplikacji webowych. Warto korzystać z doświadczeń innych programistów i całych ich społeczności, czego owocem są gotowe frameworki. Framework to nic innego jak doświadczenia wielu programistów zamknięte w konkretnym kodzie. Wykorzystanie we frameworkach obiektowego modelu programowania bardzo ułatwia tworzenie skomplikowanych aplikacji i czyni je bardziej przejrzystymi. WybórLinux applications architecture technologii (PHP, ASP lub JSP) zależy od tego w jakim języku programowania (PHP, C# lub Java) chcemy tworzyć aplikacje. Dynamiczna zawartość generowana przez strony WWW jest dzisiaj standardem. Coraz więcej serwisów jest dostępnych także na urządzenia mobilne – popularne są np. serwisy bankowe czy społecznościowe. Ten segment aplikacji (mobilne) rozwija się dzisiaj najszybciej, ze względu na najszybciej rozwijający się segment rynku urządzeń.

Bibliografia

[1] Gajda W.: PHP, MySQL i MVC. Tworzenie witryn WWW opartych na bazie danych. Gliwice, Helion, 2010.

[2] Kozłowski P.: Frameworki dla PHP, czyli wydajne tworzenie aplikacji WWW, PHP Solutions 2/2005. Warszawa, Software Wydawnictwo, 2005.

[3] Padilla A.: Beginning Zend Framework. New York, Apress, 2009.

[4] Shklar L., Rosen R.: Web Application Architecture. Chichester, John Wiley & Sons Ltd, 2003.

[5] Skaraczyński T., Zoła A.: PHP5. Programowanie z wykorzystaniem Symfony, CakePHP, Zend Framework. Gliwice, Helion, 2010.

[6] Vasvani V.: Zend Framework: A Beginner's Guide. New York, McGrawHill, 2010.

[7] Wandschneider M.: PHP i MySQL, Tworzenie aplikacji WWW. Gliwice, Helion, 2006.

[8] Welling L., Thomson L.: PHP i MySQL. Tworzenie stron WWW. Vademecum profesjonalisty. Gliwice, Helion, 2009.

[9] Ulmann L.: E-commerce. Genialnie proste tworzenie serwisów w PHP i MySQL. Gliwice, Helion, 2011.

[10] Strona internetowa: http://wazniak.mimuw.edu.pl/index.php?title=Aplikacje_WWW.

[11] Strona internetowa: http://webhosting.pl/Frameworki.PHP.przeglad.pieciu.najpopularniejszych.narzedzi.dla.programistow.WWW.

Zeszyty Naukowe WSEI seria: TRANSPORT I INFORMATYKA, 4(1/2014), s. 49-56

Tomasz SZYBORSKI

Wyższa Szkoła Ekonomii i Innowacji w Lublinie, e-mail: Tomasz. Szyborski@gmail.com

ROZWIĄZANIA TRANSMISJI DANYCH I MONITOROWANIA W SIECIACH SMART GRID PRZY UŻYCIU PRZEMYSŁOWYCH PRZEŁĄCZNIKÓW ETHERNET

DATA TRANSMISSION SOLUTIONS AND MONITORING IN SMART GRID NETWORKS USING INDUSTRIAL ETHERNET SWITCHES

Streszczenie

Rosnące zapotrzebowanie na energię elektryczną wymusiło konieczność rozbudowy stacji elektroenergetycznych, oraz stałej i niezawodnej komunikacji między nimi. Biorąc pod uwagę zróżnicowanie geograficzne odbiorców energii w postaci skupisk – tj. dużych miast, jest to problem nietrywialny. Zważywszy gwałtowny rozwój technologii światłowodowej a także powszechność Ethernetu postawiono na wykorzystanie wysokowydajnych przełączników Ethernet ze stykami RJ45 i slotami SFP umożliwiającymi przesyłanie danych po łączach światłowodowych. Dzięki spięciu przełączników w ring uzyskuje się wysokowydajną, łatwą w zarządzaniu i pokrywającą duży obszar sieć lokalną.

Słowa kluczowe: Smart Grid, Ethernet, switching, przełączniki, światłowody

Summary

The growing demand for electricity has forced the need for expansion of substations, and a constant and reliable communication between them. Given the geographical diversity of energy consumers in the form of clusters – like large cities, it is a nontrivial issue. Given the rapid development of fiber optic technology and the prevalence of Ethernet. The use of high-performance Ethernet switches with RJ45 and SFP slots allows data transfer by the fiber optic links. The result is high-performance, easy to manage local network covering vast area thanks to up to 100km ranges of SFP fiber optics modules.

Keywords: Smart Grid, Ethernet, switching, switchesi, light guides

1. Sieci Smart Grid

Termin "Smart Grid" jest w użyciu od 2003 roku, kiedy Michael T. Burr napisał artykuł „Reliability demands will drive automation investments". Mimo licznych definicji – posłużę się najbardziej powszechną:

Pod nazwą "Smart Grid" kryje się aplikacją przetwarzania sygnałów cyfrowych w sieciach elektroenergetycznych pozwalająca na dwukierunkową transmisję danych pomiędzy urządzeniami pomiarowymi w sieci a centralnym systemem zarządzającym zachowaniami Grida dotyczącymi przepływu energii do i z poszczególnych elementów.

Obecnie – w roku 2014 – sieci elektroenergetyczne podlegają trzem rodzajom przekształceń w inteligentne sieci:

- Strong Grid – jest kompletną zmianą infrastruktury.
- Smart Grid – dodanie warstwy komunikacyjnej do istniejącej sieci energetycznej.
- proces biznesowy – umożliwiający inwestycje i zwiększający świadomość dostawców i producentów energii w temacie inteligentnych systemów.

Co interesujące – ze względu na brak jednej organizacji nadzorującej prace, koncepcja Smart Grid trafiła pod skrzydła IEEE, w której niemal 400 000 członków ze środowisk akademickich i korporacyjnych ciągle rozwija i wdraża nowe pomysły do Smart Gridów. W Europie działa Smart Grid European Technology Platform http://www.smartgrids.eu/ .

Od początku XXI wieku poszukiwano możliwości wykorzystania rosnących możliwości komunikacji by minimalizować koszty i optymalizować przesyłanie energii w sieciach elektroenergetycznych, co przy rosnących ilościach odbiorników prądu zaczęło stanowić coraz większy problem. Kłopotliwym były nie tylko piki i dołki w zapotrzebowaniu energetycznym, ale również dbałość o środowisko naturalne – co spowodowało rozkwit elektrowni używających odnawialnych źródeł energii (przy użyciu m.in. ogniw fotowoltaicznych czy turbin wiatrowych).

Elektrownie słoneczne i wiatrowe są jednakże nieprzewidywalnym źródłem energii, ze względu na brak możliwości ciągłego dostarczania energii – co jest źródłem potrzeby odpowiedniego systemu kontroli by zapewnić stały dopływ prądu użytkownikom końcowym. Co istotne – spadek kosztów ogniw i turbin spowodował rozproszenie "dostawców" energii – od scentralizowanego w dużych elektrowniach do pojedynczych elementów znajdujących się np na dachach domów. Dzięki takiemu rozwiązaniu energia elektryczna jest produkowana i pożytkowana w różnych miejscach Gridu. Celem Smart Gridów jest zapewnienie nieprzerwanej dostawy energii i diagnozy uszkodzeń linii lub odcinania fragmentów Gridu, w których pojawia się zagrożenie mogące zaszkodzić całej sieci. Aby zapewnić łączność pomiędzy węzłami gridów używa się przełączników przemysłowych w topologii pierścienia (ang. *self–healing*) by zapewnić redundantną wymianę in-

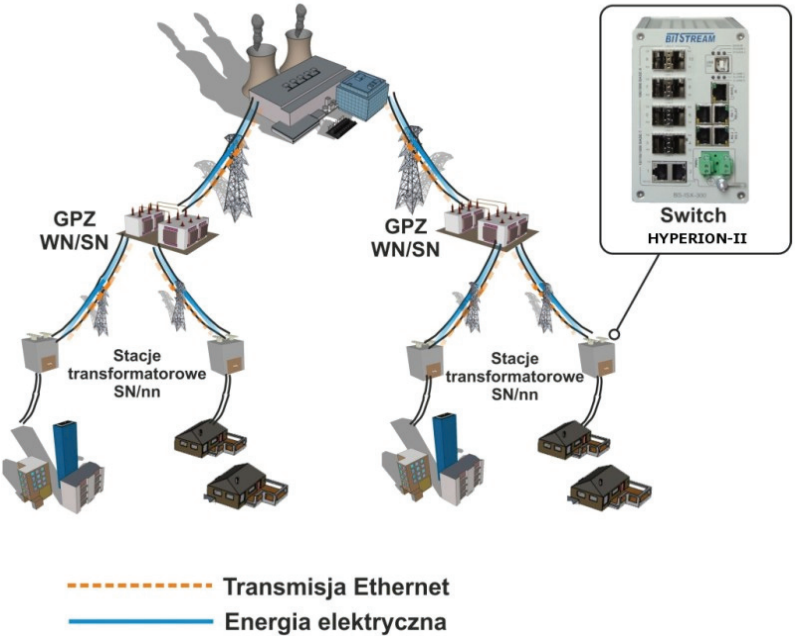
formacji w przypadku fizycznych uszkodzeń łącza takich jak katastrofy naturalne, ataki terrorystyczne lub zniszczenia spowodowane np. robotami budowlanymi. Inne rozwiązania sieci komunikacyjnej mogą powodować trudności w diagnostyce problemu bądź powodować efekt domina przy deaktywacji jednego z węzłów.

Sieci Smart Grid do sprawnego działania potrzebują licznych technologii – niestety nie wszystkie są współdziałające lub najnowocześniejsze, zatem dane mogą być przekazywane za pośrednictwem modemu aniżeli bezpośredniego połączenia sieciowego. Coraz częściej używa się połączeń światłowodowych zapewniających bezpieczeństwo transmisji od zakłóceń elektromagnetycznych a także nieporównywalnie wyższą prędkość i zasięg przesyłanych sygnałów w stosunku do połączeń miedzianych. Stabilność Gridu zależy jednak nie tylko od połączeń między węzłami, ale i od urządzeń pomiarowych zdolnych wykryć wahnięcia i zmiany w dostawie energii i monitorować stan niezbędnych urządzeń w czasie rzeczywistym. Rolę tę spełniają analogowe lub cyfrowe mierniki Smart Meter informujące Grid m. in. o temperaturze, natężeniu pola elektromagnetycznego czy poziomie wilgotności.

2. Technologie wykorzystywane w przełącznikach przemysłowych i konwerterach optycznych sygnałów zabezpieczeń

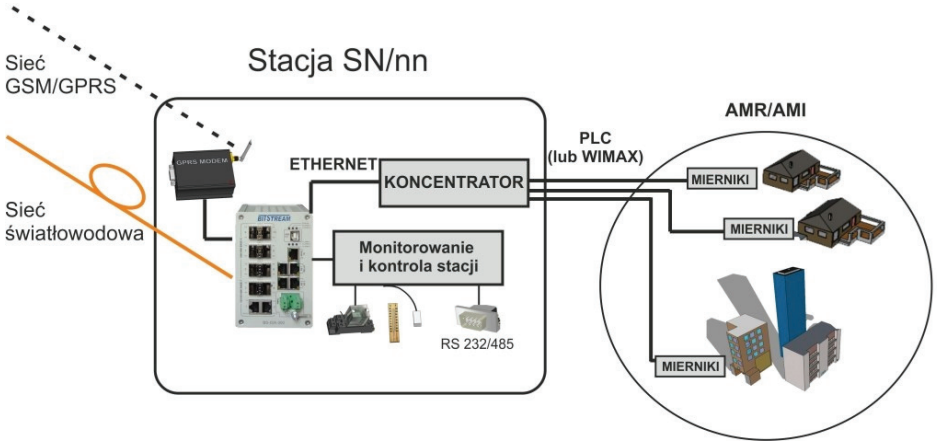
2. 1. Przełącznik przemysłowy i jego zastosowanie

W związku z wymaganiami stawianymi przez Smart Gridy urządzenia dedykowane do transmisji danych muszą spełniać określone warunki takie jak szybka i niezawodna, dwukierunkowa komunikacja między głównym systemem zarządzającym a urządzeniami pomiarowymi. Potrzeba szczegółowości, a co za tym idzie ilości przesyłanych danych w możliwie najmniejszym interwale czasowym w infrastrukturze Smart Grid skłania do użycia technologii Ethernet. Jest to o tyle proste, że jest to rozwiązane sprawdzone, skuteczne i nieustannie rozwijane – a w przypadku odpowiedniego wdrożenia – zapewnia redundantne, samo-naprawiające się połączenia (ang. *self-healing ring*) z możliwością szybkiej diagnozy błędów lub uszkodzeń w dowolnym fragmencie sieci.



Rys. 1. Zastosowanie przemysłowego przełącznika Ethernet w SmartGridach

Najistotniejszym etapem wdrożenia Smart Gridów jest zainstalowanie w stacjach transformatorowych systemów Advanced Metering Infrastructure. Zamontowane tam liczniki typu Smart Meter muszą być połączone z siecią komunikacyjną za pomocą punktów dostępowych w postaci przemysłowych przełączników Ethernet. Są to krytyczne urządzenia zapewniające sprawność działania Grida, ponieważ są odpowiedzialne za przesyłanie danych do i z systemu stosującego load balancing w sieci elektroenergetycznej. Dzięki zastosowaniu koncentratorów PLC lub modemów WiMax mogą agregować ruch nadbiegający ze Smart Meterów czym przyczyniają się do możliwie najszybszej wymianie informacji między rdzeniem decyzyjnym a fragmentami sieci.

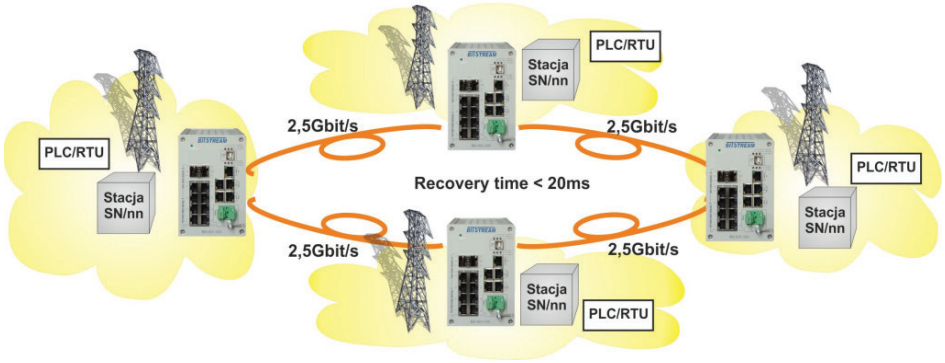


Rys. 2. Spektrum zastosowania przemysłowego przełącznika w w stacji SN/nn

Mając na uwadze kompatybilność wsteczną Smart Gridów z wcześnieszymi technologiami komunikacyjnymi przełączniki przemysłowe z rodziny Hyperion-300 – oprócz slotów SFP umożliwiających zastosowanie technologii światłowodowych są wyposażone w klasyczne porty elektryczne RJ-45 GigaBit Ethernet oraz mają opcję instalacji do dwóch modemów GSM/GPRS pozwalających na transmisję danych przez sieć komórkową, będących jednocześnie narzędziem zapewniającym protekcję połączenia w przypadku zerwania łączności z siecią optyczną bądź miedzianą. Należy również zwrócić uwagę na komunikację szeregową realizowaną przez porty RS232/485 – szeroko stosowane protokoły komunikacyjne w systemach kontrolno-pomiarowowych Modbus lub DNS3, dotychczas obsługiwane przez komunikację po RS mogą być utrzymane bez potrzeby zmiany infrastruktury, ponieważ przełącznik zapewnia tunelowanie sygnałów dzięki technologii RS over Ethernet. Jednocześnie – oprócz wsparcia wstecznego – w switchu Hyperion-300 są wykorzystywane interfejsy 1-wire do wykorzystania przy pomiarach temperatury, digital-input do monitorowania zalania bądź naruszenia przestrzeni fizycznej danego pomieszczenia oraz złącza relay-contact dające możliwość sterowania elementami zasilanymi z napięcia 230V.

Dzięki hardwarowemu wsparciu 1588-2008 IEEE Precision Time Protocol pozwalają na synchronizację czasu w sieci LAN z dokładnością do 1 mikrosekundy co jest nieocenione w przypadku systemów sterujących i pomiarowych, w których większy uchyb powodowałby błędy lub zaciemnienie obrazu sytuacji, w której znajduje się część sieci – w skrajnej sytuacji mogłoby to nie tylko zniwelować działanie Smart Grida, ale także doprowadzić do poważnych uszkodzeń pozostałych komponentów w sieci.

Podstawową zasadą inteligentnych sieci nadmiarowość połączeń zapewniających możliwie



Podstawową zasadą inteligentnych sieci jest nadmiarowość połączeń zapewniających możliwie bezawaryjne działanie w przypadku uszkodzenia jednego z torów wymiany danych. Naturalnie, najbardziej zaawansowaną topologią sieci zapewniającą redundancję połączeń jest mesh i partial-mesh, jednakże w sytuacjach w których liczy się czas przywracania ut2,0raconego połączenia a także szybka diagnoza jest ITU-T G.8032 Ethernet Ring. Dzięki switchom przemysłowym można monitorować stan włókien w światłowodzie. Oczywistym jest możliwość uszkodzenia – czy to w przypadku ingerencji w strukturę światłowodu czy złącz przy urządzeniach. Obecnie w przypadku uszkodzenia toru światłowodowego w celu lokalizacji usterki wysyłana jest ekipa serwisowa mająca za zadanie wpiąć się do traktu optycznego i za pomocą reflektometru zlokalizować uszkodzenie. Działania takie są kosztowne oraz przede wszystkim czasochłonne, co w przypadku wrażliwych aplikacji energetyki może stanowić poważny problem. Rozwiązaniem takiej sytuacji są systemy monitorujące warstwę łącza optycznego z możliwością automatycznej lokalizacji potencjalnego uszkodzenia. Zamiast wykonywania ręcznych pomiarów Hyperion-300 przy zastosowaniu modułów reflektometrycznych przeznaczonych do montażu w slotach SFP zapewnia stały monitoring warstwy pierwszej sieci. Dzięki transparentnej i opcjonalnie parametrycznie regenerowanej transmisji danych po dowolnej, będącej w zakresie modulacji prędkości system pozwala na lokalizację uszkodzenia w przypadku utraty sygnału na danym porcie optycznym poprzez automatyczne przejście do trybu pomiaru reflektometrycznego po czym wysyła komunikat SNMP trap. Takie rozwiązanie ujawnia miejsce uszkodzenia z dokładnością do 20m na odległości 10km – pozwala to na natychmiastowe podjęcie działań i usunięcie awarii. Automatyzacja testów reflektometrycznych może być użyta w przypadku monitorowanie używanych i ciemnych włókien – dzięki modułom SFP CWDM można monitorować stan włókien wykorzystujących różną długość fali optycznej.

2.2. Konwersja sygnałów optycznych zabezpieczeń

W sieciach elektroenergetycznych niezbędne jest zastosowanie zabezpieczeń odległościowych i różnicowych linii oraz nadprądowych, nadnapięciowych i różnicowych transformatorów – używających do celów komunikacyjnych interfejsów światłowodowych wykorzystujących światłowody wielomodowe. Ponieważ charakterystyka multimodów pozwala na ograniczoną odległość transmisji danych – co w przypadku rozległych sieci może stanowić problem. Następną barierę fizyczną może stanowić niedostateczna ilość wolnych włókien. Odpowiedzi na powyższe zagadnienia przynoszą konwertery optyczne sygnałów wraz z multiplekserami optycznymi. Ze względu na specyfikę sygnału transmitowanego przez zabezpieczenia (niska szybkość transmisji i nietypowa modulacja sygnału) do konwersji niezbędne są dedykowane rozwiązania. Urządzenie BS-MC-50 umożliwia konwersję sygnałów zabezpieczeń z możliwością zapewnienia dodatkowej protekcji połączenia światłowodowego.

Kolejnym urządzeniem możliwym do zastosowania w aplikacjach energetycznych typu Smart Grid jest BS-MC-90 będący konwerterem styku pojedynczego RS-232/485/422 na sygnał światłowodowy. Dzięki możliwości zastosowania interfejsu światłowodowego o długości fali 850nm, 1310 lub 1550nm, technologii WDM i CWDM a także, w zależności od potrzeb światłowodu wielo- lub jednomodowego jest to urządzenie doskonałe do instalacji w istniejącej już sieci elektroenergetycznej z możliwością komunikacji między urządzeniami celem automatyzacji procesów.

Urządzenie jest przezroczyste dla danych RS umożliwiając zastosowanie w dowolnym miejscu Grida, niezależenie od pozostałych urządzeń, których strumień danych ma być poddany konwersji.

Stan portu RS próbkowany jest z częstotliwością 10MHz, dzięki czemu dane dla styku RS-232 mogą być przesyłane z szybkością 460,8 kbit/s. Natomiast przepływność RS-485 wynosi 1M2,0bit/s. Dzięki temu BS-MC-90 wprowadzone opóźnienie transmisyjne wynosi odpowiednio nie więcej niż 400ns i 220ns.

Dla portów RS485 możliwy jest wybór RS-485(2W) lub RS-485(4W), czyli odpowiednio transmisji dwu lub czteroprzewodowej, a w przypadku połączenia ze sobą dwóch urządzeń za pomocą światłowodu każde może konwertować inny rodzaj strumienia RS.

Powyższe urządzenia obsługują protokoły komunikacyjne takie jak HTTPS, SSH i SNMPv3 zapewniające bezpieczną, szyfrowaną komunikację w celu zarządzania urządzeniem.

2.3. Parametry przełączników przemysłowych i konwerterów optycznych sygnałów zabezpieczeń

Istotnymi, a nie wymienionymi wcześniej cechami urządzeń jest zakres temperatury pracy wynoszący od -40 C do 70 stopni Celsjusza. Hyperion-300 korzysta z technologii Ethernet Ring Protection Switching ITU-T G.8032 z czasem rekonfiguracji połączeń poniżej 20ms oraz wspiera protokoły nadmiarowości: STP, RSTP i MSTP. Jako przełącznik warstwy drugiej obsługuje QoS, ramki jumbo oraz VLANy pojedynczo i podwójnie tagowane 802.1q i 802.QinQ jak również moduły optyczne o szybkości 2,5Gbit/s w przypadku liniowego połączenia przełączników. BS-MC-50 przetwarza interfejsy optyczne o długościach fal 820nm lub 850nm od strony lokalnej na dowolną falę obsługiwaną przez moduły optyczne o szybkości transmisyjnej 155Mbit/s poprzez port liniowy. Zapewnia konwersję sygnałów światłowodowych MM do SM a także sygnału lokalnego na falę CWDM oraz regenerację sygnału co znacznie zwiększa zasięg transmisji pomiędzy parą zabezpieczeń.

3. Podsumowanie

Temat inteligentnych sieci elektroenergetycznych jest nie tylko aktualny, ale w związku z dynamicznym rozwojem sieci komunikacyjnych i rosnących potrzeb dostawców i odbiorców energii – nieustannie zmieniający się w czasie. O ile technologie przesyłu informacji za pomocą RS czy Ethernet over Fiber są powszechnie znane o tyle zagnieżdżenie ich w przemyśle posiadającym stabilne rozwiązania (niejednokrotnie nieaktualne w stosunku do obecnie funkcjonujących) nie należy do zadań trywialnych. Koniecznie jest zastosowanie dedykowanych rozwiązań pozwalających na bezinwazyjne dołożenie ich do istniejącej infrastruktury. Dzięki urządzeniom Hyperion-300, BS-MC-50 i BS-MC-90 jest to możliwe.

Bibliografia

[1] http://www.bitstream.com.pl/ ; grafiki i nazwy urządzeń są wyłączną własnością Bitstream Sp. z o.o., autor niniejszego otrzymał zezwolenie na ich użycie.

[2] CieślaA A., Hanzelka Z.: Smart Grid. W: Platforma technologiczna Smart Grid [on-line]. Akademia Górniczo-Hutnicza. [dostęp 2010-08-16].

[3] Smart Grids Task force (ang.). Komisja Europejska. [dostęp 2010-08-16].

[4] Smart Grids European Technology Platform, www.smartgrids.eu". smartgrids.eu. 2011.

[5] The History of Electrification: The Birth of our Power Grid. Edison Tech Center. Retrieved November 6, 2013.

Zeszyty Naukowe WSEI seria: TRANSPORT I INFORMATYKA, 4(1/2014), s. 57-64

Aleksandra PIEREPIENKO

Wyższa Szkoła Ekonomii i Innowacji w Lublinie, e-mail: a.pierepienko@gmail.com

SYSTEM INTELIGENTNEGO DOMU

SMART BUILDING MANAGEMENT SYSTEM

Streszczenie

Szybko rozwijające się technologie, rozwój komputerów osobistych, bezprzewodowej komunikacji, oraz nowinki typu druk 3D spowodowały wysyp majsterkowiczów. Rośnie zainteresowanie domową automatyką. Ze względu na coraz niższe ceny elektroniki wiele osób decyduje się na samodzielny montaż tego typu systemów w swoich domach. W tym referacie przedstawię koncepcję budowy inteligentnego domu, podstawy sterowania elementami wykonawczymi, oraz zasady działania tanich i łatwo dostępnych urządzeń mogących pomóc w samodzielnej budowie zautomatyzowanego domu.

Summary

Rapidly developing technologies, the development of personal computers, wireless communications, and 3D printing led to a rash of DIY. There is a growing interest in domestic automation. Due to the treatment and lower prices of electronics, many people are choosing to self-assembly of such systems in their homes. In this paper introduce the concept of the smart home, the base of the control elements and principles of cheap and readily available devices that can help to build self-automated home. These instructions provide the authors with requirements concerning the layout and style which should be adopted during preparation of a paper.

Słowa kluczowe: inteligenty dom, Arduino, Raspberry Pi, mikrokontrolery, automatyka domowa

Keywords: smart building, Arduino, Raspberry Pi, microcontrollers, building automatics

1. Koncepcja inteligentnego domu

1.1. Wstęp

Coraz tańsza, oraz coraz bardziej dostępna elektronika sprawiła nagły wzrost popularności automatyki domowej. Kto z nas nie marzył o tym, by móc wrócić do domu, który już w drodze z pracy nastawił termostat, wstawił wodę na herbatę, a także zapisuje dla nas film, który telewizja puściła kiedy byliśmy zajęci? Kilka lat temu wydawało się, że to pomysł rodem z „Jetsonów", ale jesteśmy świadkami błyskawicznego rozwoju. Jak mówi stare porzekadło, potrzeba jest matką wynalazków. Można pójść krok dalej, i powiedzieć, że lenistwo jest matką potrzeby.

1.2. Historia

Inteligentny dom nie jest wcale nową koncepcją. Wywodzi się od automatyki przemysłowej. Taśmy fabryczne, ramiona, obcinarki, roboty – to wszystko powstało po to, by przyspieszyć i zoptymalizować produkcję. Wtedy na tego typu rozwiązania mogły pozwolić sobie tylko największe korporacje. Później systemy czujników zaczęli wykorzystywać hodowcy roślin, aby zapewnić im odpowiednie warunki wzrostu.

Skoro wszystko jest możliwe w przemyśle, dlaczego nie przenieść tego do domu? Tak samo jak komputerom osobistym, domowej automatyce nie wróżono dobrej przyszłości. Tylko najbogatsi mogli myśleć o zintegrowanych alarmach, elektronicznych zamkach, automatycznych bramach. Jednak elektronika tanieje z dnia na dzień. W tym momencie możemy za kilkanaście złotych (a kilka dolarów) kupić czujnik ruchu, temperatury, wilgotności i wiele, wiele innych.

1.3. Koncepcja inteligentnego domu

Koncepcja inteligentnego budynku jest prosta: jest to zestaw czujników, przełączników i elementów wykonawczych połączonych z centralną jednostką sterującą. Oczywiście, to nie my mamy być sprawcami zdarzeń – jak sama nazwa wskazuje, to dom ma myśleć za nas, a więc przystosowywać się do naszych potrzeb i upodobań. Oczywiście, nie chodzi tylko o lenistwo, ale także o bezpieczeństwo naszego domu. Owszem, zabawy ze zdalnym wyłączaniem światła, kiedy już się położyło do łóżka, albo automatyczne podlewanie kwiatków jest sporym ułatwieniem, ale prawdziwy inteligentny dom to coś o wiele więcej.

Co może wchodzić w skład takiego zestawu? Ogranicza nas tylko wyobraźnia i nasze możliwości konstruktorskie, chyba, że fundusze nie są dla nas aż tak istotne – wtedy, możemy kupić gotowe elementy. Zaczynając od czujników temperatury i sterowania termostatem, przez czujniki pożaru z wyłączaniem zasilania i awaryjnym oświetleniem, monitoring i ochronę przeciwwłamaniową aż po

spełnianie naszych zachcianek – bo dlaczego ekspres do kawy nie mógłby być zsynchronizowany z naszym budzikiem?

2. Budowa oparta na Raspberry Pi i Arduino

2.1. Wstęp

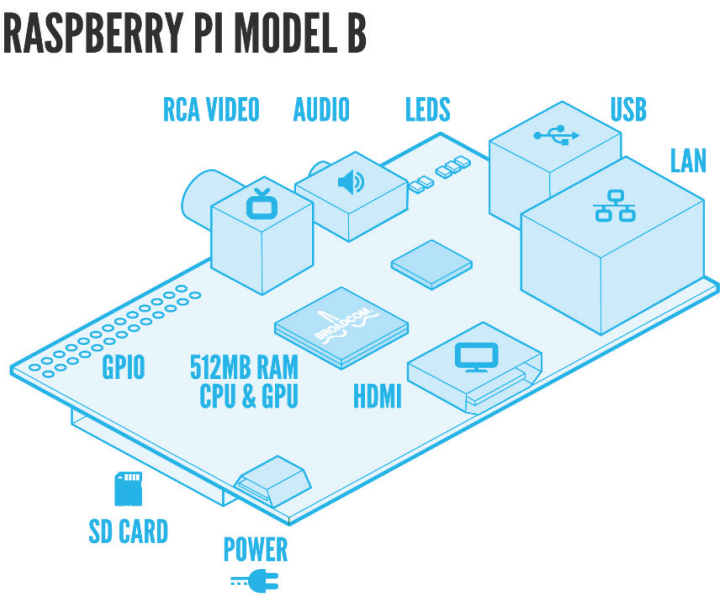
Wzrost zainteresowania tą technologią sprawił, że jak grzyby po deszczu powstają nowe projekty, które pozwalają na samodzielne stworzenie takiego systemu. Fakt, że ma to wiele wad – trzeba poświęcić czas na montaż, a przede wszystkim na naukę zapanowania nad każdą z części naszego układu. Niesie to ze sobą ogromną satysfakcję, oraz niemałe oszczędności – gotowe układy często kosztują ogromne pieniądze, nie oferując nam w zamian często poszukiwanych przez nas, bardzo indywidualnych funkcjonalności. W dodatku brakuje ciągle zunifikowanego protokołu komunikacji tych urządzeń. Każdy producent ma własny pomysł na rozwiązanie tego problemu, co często kończy się na tym, że do sterowania lodówką, światłem, termostatem i zamkiem w drzwiach mamy osobne aplikacje. Ich przełączanie przysparza wiele problemów, a napisanie oprogramowania do określania związków przyczynowo skutkowych (kiedy zgaszę światło przy łóżku, zwiń roletę) wydaje się często niemożliwe poprzez zamknięcie oprogramowania.

Przyjrzyjmy się więc możliwości stworzenia własnego systemu do sterowania domem. Oczywiście, nie mamy miejsca na szczegółową instrukcję, a elementów i czujników może być naprawdę ogromna ilość, dlatego też skupimy się na podstawowych elementach oraz podstawie montażu.

Należy pamiętać, że większość urządzeń w naszym mieszkaniu działa pod niebezpiecznym napięciem powyżej 200V. Każda ingerencja bez odpowiednich umiejętności może skutkować porażeniem prądem, poważnymi uszkodzeniami ciała, a nawet śmiercią. Do pracy z urządzeniami zasilanymi wysokim napięciem najlepiej będzie poprosić o pomoc doświadczonego elektryka.

2.2. Raspberry Pi – wprowadzenie

Należy zacząć od serca naszego układu. W tym momencie najciekawszym i najlepiej wspieranym przez społeczność projektem jest miniaturowy komputer, wielkości karty kredytowej – Raspberry Pi. Mimo niewielkich rozmiarów ma on ogromne możliwości – procesor ARM o częstotliwości 700Mhz, 512MB RAMU (wersja B 2.0), dwa porty USB, oraz gniazdo Ethernet. Do tego złącze HDMI, RCA, minijack – czyli wszystko, czego potrzebujemy na start. Jego najciekawszą i niezbędną do realizacji tego projektu częścią są piny GPIO umożliwiające sterowanie urządzeniami poprzez podawanie wysokiego, lub niskiego stanu. Wszystko to za relatywnie niską cenę – komplet, razem z obudową, zasilaczem i kartą SD z systemem można kupić za około 200 złotych.



Rys 1. Raspberry Pi z zaznaczonymi portami [2]

2.3. Raspberry Pi – system i programowanie

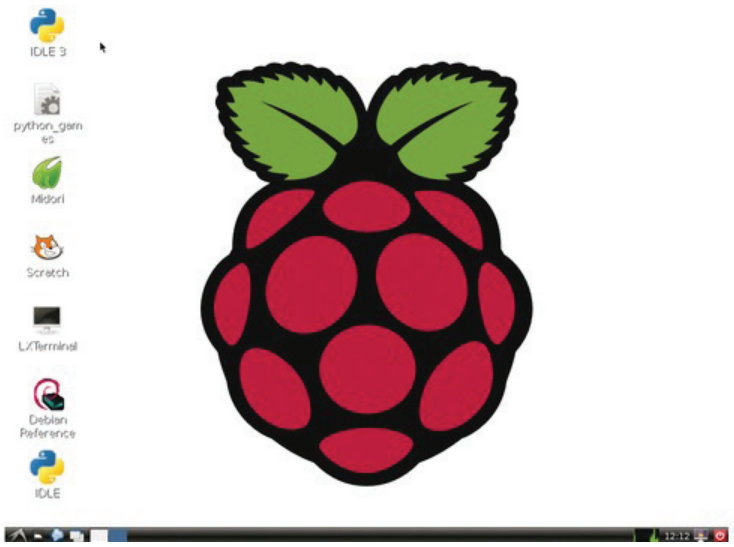
Najpopularniejszym systemem operacyjnym na Raspberry jest Raspbian. Jak łatwo się domyślić, jest to odmiana Debiana, odpowiednio uszczuplona tak, by nie przeciążać procesora. Na niej też będziemy działać.

Instalacja systemu odbywa się za pomocą karty SD. Warto zauważyć, że poza tą kartą Raspberry nie jest zaopatrzone w żaden magazyn danych.

Po zainstalowaniu systemu i wpisaniu w terminal komendy „startx" – czyli po uruchomieniu trybu graficznego – naszym oczom ukaże się pulpit, bardzo podobny do tego, jaki mamy w naszych domowych komputerach. Nie będziemy jednak zbyt często z niego korzystać – większość działań możemy wykonywać z poziomu terminala. Często „mózg" naszego domu będziemy musieli zamknąć na przykład w skrzynce z bezpiecznikami, więc aby nie było trudności z późniejszymi konfiguracjami, możemy porozumiewać się z nim za pomocą SSH. Jeśli będzie zależało nam na trybie graficznym, powinniśmy zainstalować serwer VNC, czyli wirtualny pulpit.

Do porozumiewania się z czujnikami i elementami wykonawczymi przy użyciu portów GPIO musimy wykorzystać jakiś język programowania. Twórca właściwie nie jest zbytnio ograniczony – z obsługą GPIO za pomocą odpowiednich bibliotek poradzi sobie python, C/C++ oraz wiele innych popularnych języków progra-

mowania. Możemy nawet sterować pinami poprzez skrypty w bashu. Na potrzeby tej pracy za podstawowy język uznamy jednak Pythona, ze względu na wsparcie innych użytkowników „Malinki", mnogość poradników dostępnych w internecie, oraz ciągle wzbogacane pozycje ze specjalistycznej literatury.



Rys. 2. Pulpit systemu Raspbian [2]

2.4. Arduino – wprowadzenie, wersje i programowanie

Samo Raspberry nie wystarczy, zwłaszcza, jeśli projekty będą skomplikowane i rozbudowane. Do bezpośredniego sterowania niektórymi urządzeniami wykorzystywać będziemy inną płytkę, bardzo lubianą przez majsterkowiczów – Arduino. Jest to płytka służąca do prototypowania rozwiązań, oparta na mikrokontrolerze ATMega. Możemy przebierać w jej wersjach – od rozbudowanej, przydatnej przy dużych projektach wersji Mega, przez coraz to mniejsze – UNO, MICRO, PRO NANO. Jest też specjalna wersja LILYPAD, mała i okrągła, której przeznaczeniem na przykład są inteligentne tekstylia. Najbardziej popularną płytką, ale też jedną z droższych, jest Arduino UNO. Mimo niewielkich rozmiarów (jest troszkę mniejsza od Raspberry) często ciężko będzie nam ją zmieścić w niektórych układach, zresztą będzie to zupełnie niepotrzebne i bardzo kosztowne. Arduino UNO ze względu na łatwość obsługi będzie nam służyło do prototypowania naszych rozwiązań. Ich końcowe działanie powierzymy Arduino PRO MINI lub – przy bardzo małych projektach i gdy nabierzemy trochę wprawy – mikrokontrolerom z rodziny ATMega lub ATTiny. Są to te same układy, jakie znajdziemy

w ARDUINO, jednak ich programowanie wygląda nieco inaczej – zainteresowanych odsyłam do bibliografii.

Z Arduino komunikujemy się podobnie, jak z Raspberry – za pomocą pinów. Jedyną różnicą jest to, że w Raspberry występują piny męskie, a w Arduino żeńskie, więc przy kupnie odpowiednich kabelków i zworek trzeba zwrócić na to uwagę. Arduino programujemy za pomocą języka C w specjalnie przygotowanym środowisku Arduino IDE.

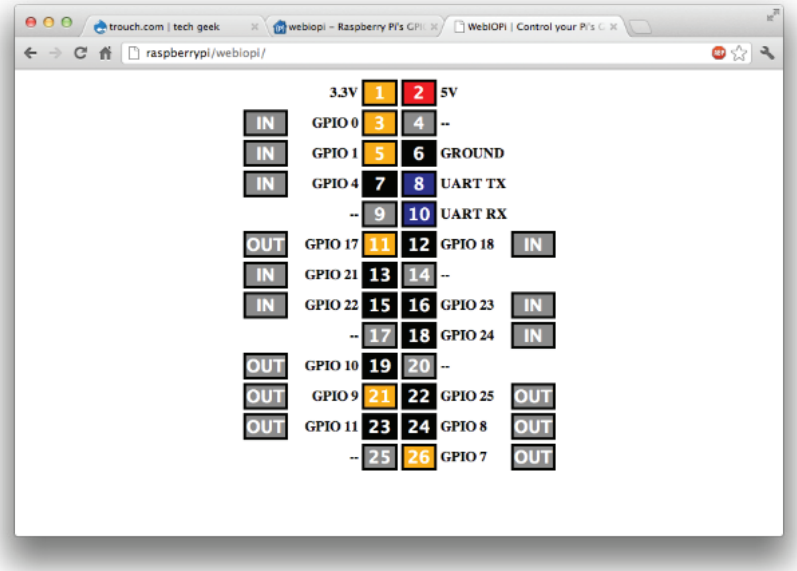
3. Sterowanie inteligentnym domem, rozbudowa, zabezpieczanie

3.1. Założenia

W zależności od tego, jakiego działania oczekujemy, potrzebujemy różnych czujników i elementów wykonawczych. Opisanie ich wszystkich jest oczywiście niemożliwe, ale interesujące nas układy z pewnością mają noty katalogowe, a może też również gotowe schematy połączeń. Jako, że sterowanie odbywać się powinno przez jednostkę centralną, a więc Raspberry Pi, nie wystarczy jedynie napisanie odpowiedniego programu. Nie przewidzimy wszystkich sytuacji, nie możemy „na sztywno" określić, że o godzinie osiemnastej, w każdy czwartek będziemy oglądać film. Mimo tego, że pewne zachowania można określić, to pewne będziemy musieli sterować ręcznie.

3.2. Sterowanie

Sterować ręcznie możemy na wiele sposobów, od oczywistego, ale najmniej wygodnego pisania komend. Możemy robić to też na przykład za pomocą pilota, jednak przy zakładanym przypadku, że Raspberry będzie znajdowało się w pobliżu skrzynki z bezpiecznikami sygnał z pilota może nie docierać do urządzenia. Na szczęście dostępnych jest wiele programów i bibliotek które nam w tym pomogą – jedną z nich jest WebIOPi. Jest to aplikacja do sterowania wejściami GPIO poprzez HTML. Dzięki niej za pomocą komputera lub telefonu będziemy mogli sterować naszym komputerem, a co za tym idzie – domem.



Rys. 4. WebIOPi – interfejs [1]

3.2. Nadzieje i zagrożenia

Projekty związane z domową automatyką są najróżniejsze – od czujnika wilgotności w doniczce i sygnalizowania, że czas na podlanie, po kamerki w lodówkach z możliwością sprawdzenia na zakupach, czy na pewno mamy jeszcze mleko. Możemy również założyć system alarmowy, a być może doczekamy chwili, kiedy będziemy mogli sterować różnymi urządzeniami za pomocą myśli.

Z powodu rosnącej popularności inteligentnych domów pojawia się jednak bardzo wiele zagrożeń. Coraz więcej urządzeń podłączonych do internetu nie jest komputerami – zjawisko to określa się „internetem rzeczy". Bardzo często w takich urządzeniach brakuje odpowiednich zabezpieczeń – przecież nie przechowujemy ważnych danych w pralce czy ekspresie do kawy. Jednak stanowią one często bramę dostępu na przykład do naszego routera – a stąd droga do kradzieży haseł do banku jest już prostsza. Nie życzylibyśmy sobie również aby ktoś niepowołany dostał się do naszego systemu monitoringu – należy więc pamiętać, o bardzo dokładnym zabezpieczeniu sieci.

Bibliografia

[1] Boxall J.: Arduino, 65 praktycznych projektów. Gliwice, Helion, 2014, s. 35-50.

[2] Halfacree G., Upton E.: Raspberry Pi, Przewodnik użytkownika. Gliwice, Helion, 2013, s. 26-61.

[3] Simon M.: Raspberry Pi, Przewodnik dla programistów Pythona. Gliwice, Helion, 2014, s. 13-49.

[4] http://www.raspberrypi.org/learning/python-intro/ 23.06.2014.

[5] http://malinowepi.pl/post/78555878338/co-mozna-zrobic-z-raspberry-pi-wraz-z-koncem 21.06.2014.

[6] https://www.youtube.com/channel/UCRAvo5cQWyfog8nRzlf_jWg.

Zeszyty Naukowe WSEI seria: TRANSPORT I INFORMATYKA, 4(1/2014), s. 65-81

Bartosz KOWALEWSKI

Wyższa Szkoła Ekonomii i Innowacji w Lublinie, e-mail: bartoszkowalewski@gmail.com

PRAWDZIWE PUZZLE

TRUE JIGSAW PUZZLE

Streszczenie

Poniższy tekst przedstawia podstawowe mechanizmy interakcji z elementami, które są częścią gry puzzle. W pierwszej kolejności przedstawiany jest materiał dzięki, któremu rozgrywka zawiera element realizmu. Mowa tu o podnoszeniu elementów układanki, które są widoczne w całości lub chwytane za ich widoczne części. W drugiej i najbardziej obszernej części zapoznamy się z koncepcją, która ułatwi graczowi dopasowywanie do siebie elementów puzzli tak, aby nie tworzyły się mało estetyczne przerwy między nimi.

Słowa kluczowe: programowanie, algorytmy, gry, puzzle

Summary

Article presented below covers issues of basic human-puzzle piece's interaction mechanisms used in jig-saw puzzle game. First of all, we will go over material concerning realistic site of above mentioned game, which is the phenomenon of picking up wholly visible puzzle pieces and their partly covered versions when hold onto them. Second and most absorbing part is easy placing mode, which turns out to be really convenient when playing this kind of game. Thanks to that users will not see undesirable space between puzzle pieces.

Keywords: programming, algorithms, games, jigsaw puzzle

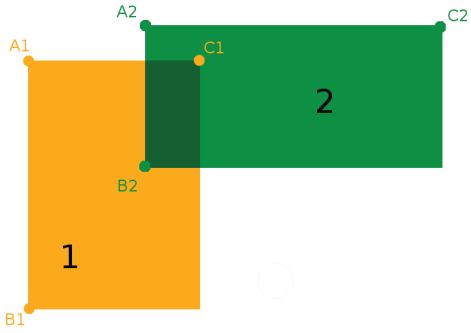
1. Wstęp

Któż jako dziecko nie spędził paru chwil nad tak trywialną grą jaką są puzzle. Obecnie wszystkim znana łamigłówka, która powstała dawno przed elektroniczną formą rozrywki, dorobiła się wiele różnych wersji. Jedną z nich jest, odwzorowany w świecie wirtualnym, jej pierwowzór.

2. Poziomowanie puzzli

Pierwszym, i w sumie najważniejszym elementem, który zapewni nam „naturalność" puzzli jest tzw. poziomowanie. Mówiąc bardziej ludzkim językiem, jest to mechanizm, który będzie obsługiwał nakładanie się elementów układanki na siebie oraz umożliwi nam podnoszenie puzzla, którego odsłonięty fragment, właśnie chwyciliśmy. Aby zbudować taki mechanizm, oprócz rozpatrzenia przypadków w jaki puzzle mogą się na siebie nakładać, musimy jeszcze wprowadzić pojęcie poziomu, czyli sposobu numerowania puzzli tak, aby opisać, które kawałki znajdują się na warstwie niższej, tej samej, bądź wyższej od pozostałych.

Na potrzeby tworzenia algorytmu, przyjmiemy model składający się z dwóch puzzli oznaczonych w odpowiedni sposób.



Rys. 1. Model puzzli z oznaczeniami punktów kluczowych

Na tym etapie prac natkniemy się na dwa problemy, które zostaną omówione jeden po drugim. W pierwszej kolejności zajmiemy się zagadnieniem wykrywania powierzchni wspólnej puzzli, która występuje w momencie kiedy elementy układanki nakładają się na siebie. Należy zauważyć, że algorytm, który zostanie za chwilę opisany, działa wyłącznie dla puzzli o kształcie prostokąta(kwadrat to też prostokąt).

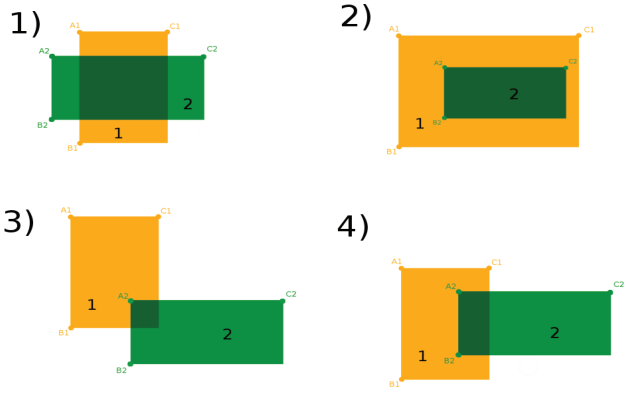
Przejdźmy zatem do samego algorytmu. W trakcie analizy pod uwagę weźmiemy trzy punkty z każdego kawałka. Są to punkty oznaczone na rys. 1, czyli A1, B1, C1 oraz A2, B2, C2. Sama analiza będzie polegała na prostym porównaniu

punktów i sprawdzaniu czy zostały zachowane odpowiednie zależności. Jednak jak wyznaczyć owe zależności? Najprostszym i najbardziej skutecznym sposobem na to, jest narysowanie na kartce papieru wszystkich, możliwych sytuacji. Będzie to solidną podstawą w naszych rozważaniach.

Poniżej zostały pokazane najważniejsze przypadki.

Na podstawie rys. 2 możemy określić następujące warunki :

$$1.1.1^{o} \quad A1_x \geq A2_x \quad \&\& \quad A1_x \leq C2_x$$
$$1.1.2^{o} \quad C1_x \geq A2_x \quad \&\& \quad C1_x \leq C2_x$$
$$1.2.1^{o} \quad A1_y \geq A2_y \quad \&\& \quad A1_y \leq B2_y$$
$$1.2.2^{o} \quad B1_y \geq A2_y \quad \&\& \quad B1_y \leq B2_y$$



Rys. 2. Najważniejsze przypadki nakładania się puzzli na siebie

W następnym kroku, gdy mamy już ustalone warunki, możemy zabrać się za analizę, która będzie polegała na wykonaniu paru operacji logicznych:

$$(1.1.1^{o} \vee 1.1.2^{o}) \wedge (1.2.1^{o} \vee 1.2.2^{o})$$

Algorytm wygląda dobrze i faktycznie działa, ale w praktyce od razu zauważymy, że obsługuje on tylko przypadek 3). Jeżeli spojrzymy na ułożenie puzzli nr 2) rozwiązanie nasuwa się samo. Wystarczy wykonać powyższy algorytm po raz drugi tylko, że z zamienionymi parametrami. Wtedy warunki ustalone poprzednio będą wyglądać następująco:

$$2.1.1^{o}\ A2_x \geq A1_x \ \ \&\& \ \ A2_x \leq C1_x$$
$$2.1.2^{o}\ C2_x \geq A1_x \ \ \&\& \ \ C2_x \leq C1_x$$
$$2.2.1^{o}\ A2_y \geq A1_y \ \ \&\& \ \ A2_y \leq B1_y$$
$$2.2.2^{o}\ B2_y \geq A1_y \ \ \&\& \ \ B2_y \leq B1_y$$

Podsumowując, jeżeli prototyp metody wyglądałby w ten sposób PowWsp(Puzzel1, Puzzel2), to wywołanie jej wyglądałoby tak:

$$1^{o}\ PowWsp(P1, P2);$$
$$2^{o}\ PowWsp(P2, P1);$$

gdzie P1 to puzzel o nr 1 na Rys. 1.2, a P2 to puzzel nr 2. Ostatnie działanie jakie zostanie wykonane to alternatywa wartości zdań logicznych 1o i 2o .

$$1^{o} \vee 2^{o}$$

Na tym etapie warto wspomnieć o oczywistości, która jednak może umknąć. Powyższa metoda zwraca wartość True w momencie wykrycia powierzchni wspólnej puzzli, w innym False.

Powyższy algorytm został ukończony. Używając go dostajemy, w trzech na cztery przypadki, jednoznaczną i poprawną odpowiedz. Jednak co w przypadku gdy napotkamy sytuację 1) zilustrowaną na rys. 2 ? Tutaj algorytm stwierdzi, iż puzzle nie mają powierzchni wspólnej, co nie jest zgodne z prawdą, ponieważ warunki ustalone wcześniej nie są spełniane. Aby taki porządek rzeczy nie sprawiał nam problemu, należy zauważyć specyficzną sekwencję wartości zdań logicznych i opisać ją za pomocą bloku warunkowego:

$$1.1^{o}\ True$$
$$1.2^{o}\ False$$
$$2.1^{o}\ False$$
$$2.2^{o}\ True$$

lub

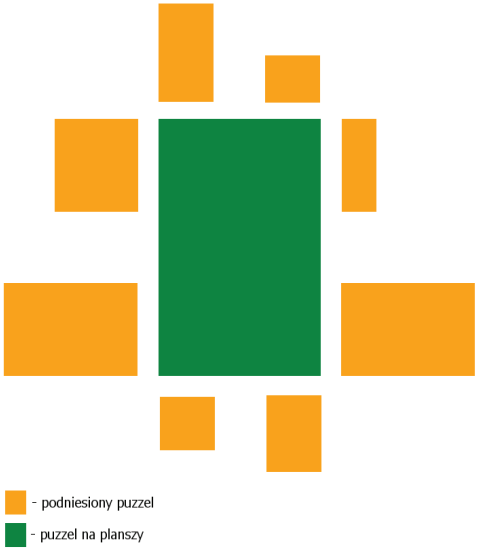
$$1.1^{o}\ False$$
$$1.2^{o}\ True$$
$$2.1^{o}\ True$$
$$2.2^{o}\ False$$

W tej kwestii to już wszystko. Teraz jesteśmy w pełni stwierdzić czy puzzle nakładają się na siebie czy nie. Teraz już śmiało możemy zacząć pracę nad pozio-

mowaniem(to jest drugi problem). Poziomowanie będzie polegało na detekcji powierzchni wspólnej puzzla podniesionego przez gracza, z każdym puzzlem pojedynczo. Jeżeli puzzle nakładają się na siebie to porównujemy ich poziomy (podniesionemu puzzlowi automatycznie nadawany jest poziom 1). Jeżeli poziom podniesionego puzzla jest mniejszy, bądź równy poziomowi puzzla, z którym jest porównywany, to puzzlowi podniesionemu nadajemy poziom o jeden większy niż poziom tego drugiego puzzla. Powtarzamy to (ilość puzzli – 1) razy.

3. Wspomaganie układania puzzli

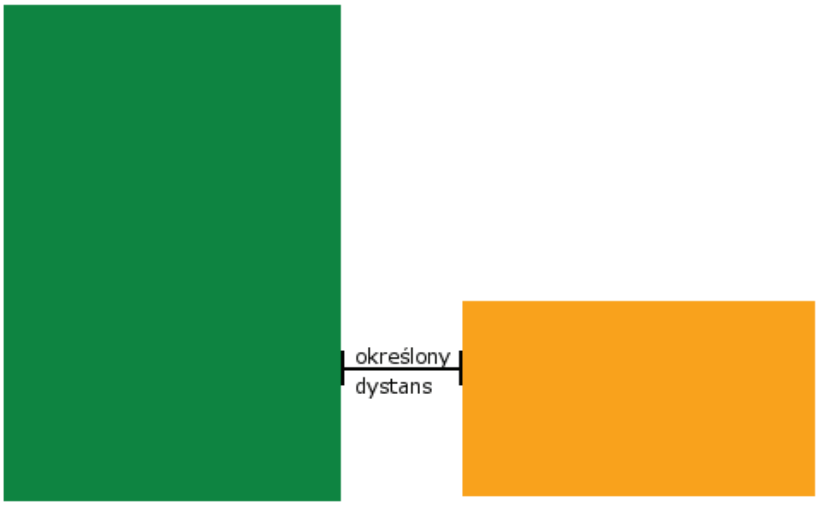
Aby układanka została ukończona, gracz musi ustawić wszystkie jej elementy na właściwym dla nich miejscu. W rzeczywistości jest to bardzo łatwe, oczywiście mam na myśli zestawienie dwóch lub więcej puzzli obok siebie, a nie ukończenie układanki, jednak w rzeczywistości komputerowej jest to o tyle trudniejsze, iż musimy układać je co do piksela tak aby nie nakładały się na siebie. Tylko w ten sposób stworzymy ukończoną układankę, która odpowiada tej ze świata rzeczywistego. Taki efekt można uzyskać ręcznie, ale będzie on wymagał bardzo precyzyjnych ruchów, które są mało naturalne oraz bardzo męczące, a przecież komfort jest bardzo ważny, zwłaszcza kiedy chodzi o gry komputerowe.



Rys. 3. Wszystkie przypadki przyłożenia podniesionego puzzla do puzzla na planszy

Kod, który napiszemy, będzie wymagał od nas rozpatrzenia 8 przypadków.

Na powyższym rysunku zostały zilustrowane wszystkie 8 sytuacje, jakie możemy spowodować beztrosko układając cyfrowe puzzle. W momencie, gdy program napotka jedną z nich, np.

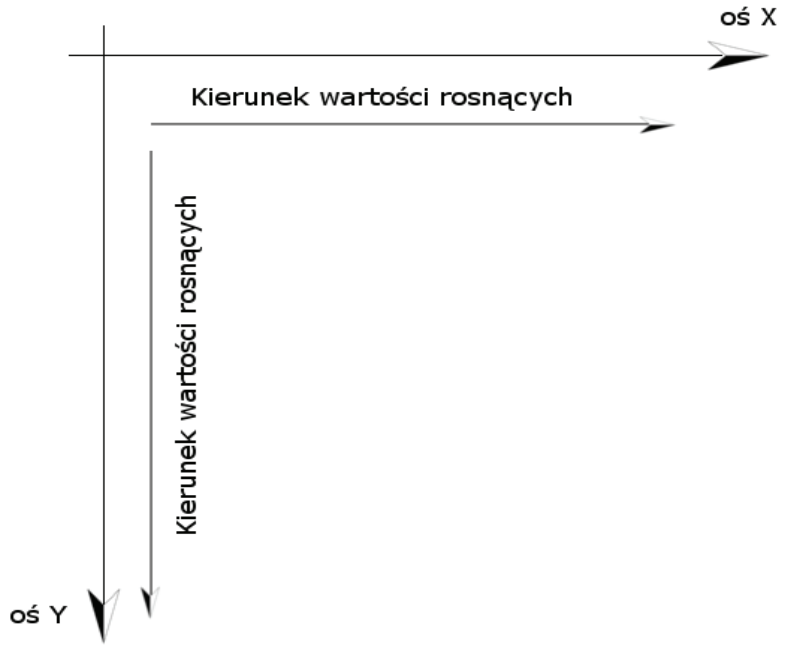


Rys. 4. Jeden z przypadków przyłożenia podniesionego puzzla

dobrze napisane oprogramowanie powinno podjąć działanie, którego wynikiem będzie oto taki stan

Rys. 5. Wynik działania algorytmu

Warto na tym etapie wspomnieć o jednym szczególe. Rysowanie grafik na ekranie zaczyna się od lewego górnego rogu, czyli pierwszego piksela w obrazie. Modyfikując koordynaty x i y obrazu, tak naprawdę zmieniamy współrzędne pierwszego piksela, a pozycja pozostałych pikseli w ekranowym układzie współrzędnych jest określana na podstawie tego pierwszego, dlatego mamy dostęp tylko do punktów oznaczonych na schematach literą A np. A1, A2 itp. W celu wyczerpania tematu zamieszczono jeszcze ilustrację przedstawiającą osie układu współrzędnych ekranu komputera.



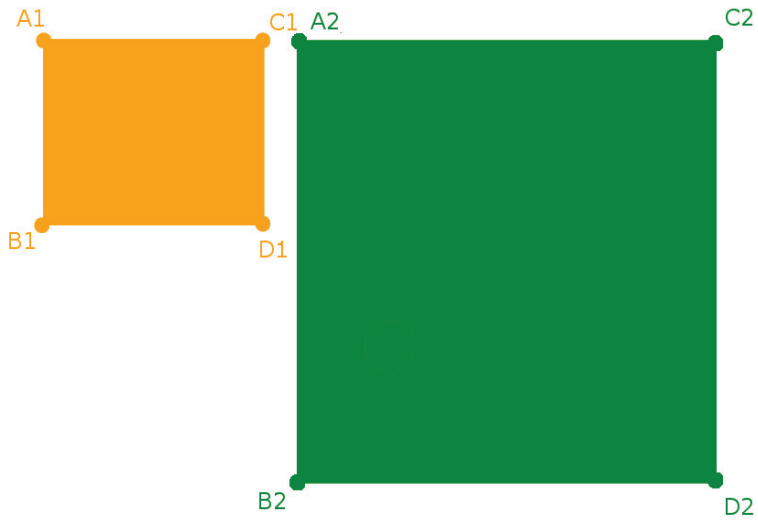
Rys. 6. Osie ekranowego układu współrzędnych

4. Przypadek nr 1

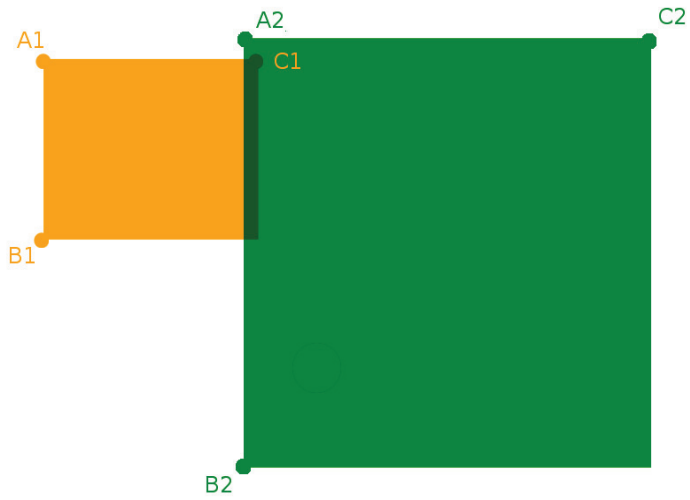
W pierwszej kolejności badamy czy podniesiony puzzel, w momencie upuszczenia go na planszę gry, jest dostatecznie blisko jakiegokolwiek, innego puzzla (rys. 7). W tym przypadku badamy odległość pomiędzy dwoma punktami, C1 oraz A2. Badanie polega na sprawdzeniu czy koordynaty tych dwóch punktów spełniają następujący warunek:

$$(C1_x - A2_x \leq dys) \&\& (C1_y - A2_y \leq dys)$$

Przy tego typu rozwiązaniu pojawia się niewielka, lecz istotna kwestia. Dotyczy ona znaku liczby otrzymanej drogą odejmowania. Ten problem wynika z faktu, że oprócz sytuacji ukazanych na ilustracjach, mogą wyniknąć sytuacje podobne, ale z tą różnicą, iż puzzel upuszczany może nakładać się z puzzlem na planszy (rys. 8).



Rys. 7. Przypadek nr 1



Rys. 8. Nakładające się na siebie puzzle

Żeby nie sprawdzać, które wartości koordynat są większe, a które mniejsze, dokonamy małej modyfikacji już istniejącego warunku bloku warunkowego:

$$(C1_x - A2_x \geq (-1 * dys)) \&\& (C1_x - A2_x \leq dys)$$
$$\&\& (C1_y - A2_y \geq (-1 * dys)) \&\& (C1_y - A2_y \leq dys)$$

Dzięki takiemu zabiegowi ograniczamy ilość bloków warunkowych do minimum. Jedyną niewiadomą jak nam pozostaje to część opisana skrótem „dys". Zmienna „dys" jest to komórka w pamięci, która przechowuje, z góry ustaloną przez programistę, maksymalną wartość odległości jaka może dzielić dwa elementy układanki, żeby algorytm zadziałał. Jeżeli warunki zostały spełnione i algorytm zareagował, jedyne co musimy zrobić to ustawić współrzędne x i y podniesionego, a raczej upuszczanego przez gracza, puzzla. Celem naszego działania będzie uzyskanie podobnego efektu jak na rys. 5. Konkretnie w tym przypadku będziemy próbować nadać punktowi C1 współrzędnych punktu A2. Najprościej jest to zrobić nadając punktowi A1 współrzędnych punktu A2. Po tym jak to zrobimy, puzzel upuszczany i puzzel na planszy nakładają się na siebie, a nie do końca o to nam chodziło. W takim razie przesuńmy ten pierwszy fragment układanki w lewo. Modyfikujemy współrzędną x punktu A1 o szerokość przesuwanej grafiki. W postaci instrukcji operacja ta miała by postać:

$$A1_x = A2_x - szerokość_{A1\,B1\,C1\,D1}$$
$$A1_y = A2_y$$

Jeszcze parę słów wyjaśnienia na koniec tej części artykułu. Pomimo konieczności samodzielnego wyliczania koordynat punktów innych niż punkty oznaczonymi literą A, zdecydowano się używać ich oznaczeń w warunkach bloków warunkowych dla uproszczenia zapisu. Zanim przejdziemy dalej, poniżej zamieszczono wzory na obliczanie koordynat punktów BN, CN i DN (N jest dowolną liczbą) :

$$BN_x = AN_x$$
$$BN_y = AN_y + wysokość_{AN\,BN\,CN\,DN}$$
$$CN_x = AN_x + szerokość_{AN\,BN\,CN\,DN}$$
$$CN_y = AN_y$$
$$DN_x = AN_x + szerokość_{AN\,BN\,CN\,DN}$$
$$DN_y = AN_y + wysokość_{AN\,BN\,CN\,DN}$$

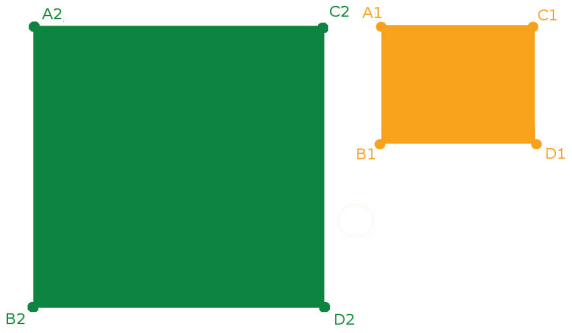
5. Przypadek nr 2

W tym przypadku na warsztat bierzemy punkty C2 i A1. Warunek wygląda następująco:

$$(C2_x - A1_x \geq (-1 * dys)) \&\& (C2_x - A1_x \leq dys)$$
$$\&\& (C2_y - A1_y \geq (-1 * dys)) \&\& (C2_y - A1_y \leq dys)$$

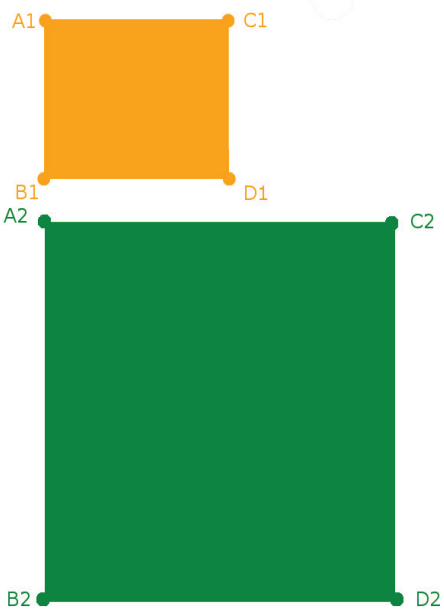
Zwalniany puzzel ustawiamy na pozycję punktu C2, czyli tak jak poprzednio nakładamy puzzle na siebie i następnie upuszczany puzzel przesuwamy w prawo o szerokość puzzla na planszy :

$$A1_x = A2_x + szerokość_{A2\,B2\,C2\,D2}$$
$$A1_y = A2_y$$



Rys. 9. Przypadek nr 2

6. Przypadek nr 3



Rys. 10. Przypadek nr 3

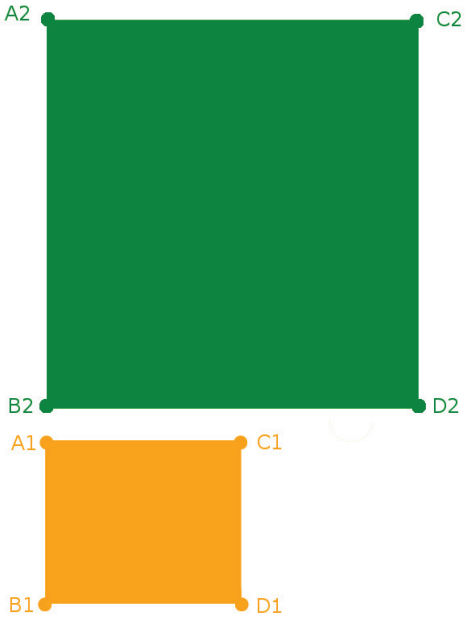
Para punktów, którą zajmiemy się teraz jest to para złożona z punktu A2 i B1. Warunek potrzebny do określenia czy mamy do czynienia z omawianą sytuacją wygląda następująco:

$$(B1_y - A2_y \geq (-1 * dys)) \&\& (B1_y - A2_y \leq dys)$$
$$\&\& (B1_x - A2_x \geq (-1 * dys)) \&\& (B1_x - A2_x \leq dys)$$

Aby ustawić punkt B1 na pozycji punktu A2 najpierw nakładamy puzzle na siebie, a potem przesuwamy upuszczany puzzel w górę o jego wysokość :

$$A1_x = A2_x$$
$$A1_y = A2_y - wysokość_{A1B1C1D1}$$

7. Przypadek nr 4



Rys. 11. Przypadek nr 4

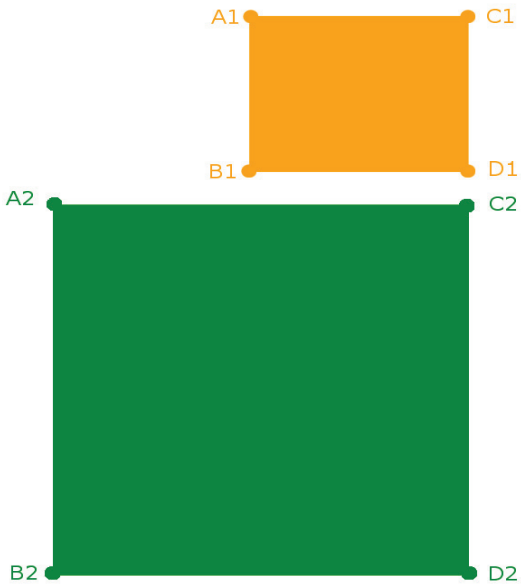
Przy tej odsłonie główną rolę odegrają punkty B2 i A1. Niezmiennie, także i tutaj określamy warunek niezbędny do zidentyfikowania przypadku :

$$(B2_y - A1_y \geq (-1 * dys)) \&\& (B2_y - A1_y \leq dys)$$
$$\&\& (B2_x - A1_x \geq (-1 * dys)) \&\& (B2_x - A1_x \leq dys)$$

Ta sytuacja praktycznie nie różni się niczym od poprzedniej, z wyjątkiem kierunku, w którym przesuwamy upuszczany puzzel, jednak dystans jaki musi przebyć jest wyznaczany przez wysokość puzzla na planszy:

$$A1_x = A2_x$$
$$A1_y = A2_y + wysokość_{A2\,B2\,C2\,D2}$$

8. Przypadek nr 5



Rys. 12. Przypadek nr 5

Punkty D1 i C2 będą punktami, które weźmiemy pod uwagę, jeżeli chodzi o zbudowanie odpowiedniego wyrażenia logicznego :

$$(C2_x - D1_x \geq (-1 * dys)) \&\& (C2_x - D1_x \leq dys)$$
$$\&\& ((C2_y - D1_y \geq (-1 * dys)) \&\& (C2_y - D1_y \leq dys))$$

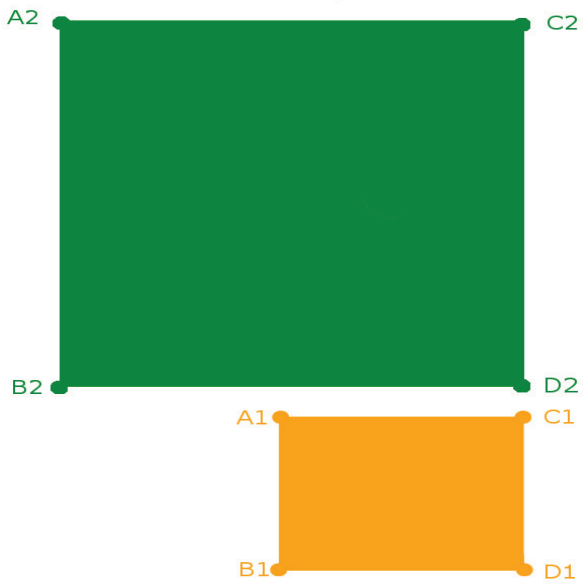
Pierwszym krokiem w stronę właściwego ułożenia, oprócz nałożenia puzzli na siebie, jest przesunięcie, upuszczanego puzzla, w górę ekranu o jego wysokość. Najprostsza cześć za nami. Ta trudniejsza to przesunięcie „w prawo" puzzla. Zrobimy to poprzez przesunięcie puzzla A1 B1 C1 D1 wzdłuż osi X względem punktu A2 o różnicę szerokości tych dwóch elementów. Całość wygląda w ten oto sposób:

$$roz = szerokość_{A2\,B2\,C2\,D2} - szerokość_{A1\,B1\,C1\,D1}$$
$$A1_x = A2_x + roz$$
$$A1_y = A2_y - wysokość_{A1\,B1\,C1\,D1}$$

Trudność drugiej części nie polega na napisaniu kodu, ponieważ nie jest on jakoś specjalnie skomplikowany. Jednak gdy przypomnimy sobie o zmienności znaku liczby otrzymanej drogą odejmowania (chodzi o zmienną „roz"), to cała sprawa nieco się komplikuje. Można odnieść wrażenie, że czegoś tu brakuje, ale

na szczęście tak nie jest. Jeżeli chodzi o sytuację pokazaną na ilustracji powyżej, to nie ma najmniejszego problemu. Wynik różnicy jest dodatni, a dzięki temu podniesiony puzzel jest przesuwany w prawo i tu nie ma problemu. Druga wersja tego samego wydarzenia to sytuacja odwrotna, czyli gdy podniesiony puzzel jest szerszy niż puzzel na planszy. W rezultacie mamy do czynienia z ujemną zawartością zmiennej „roz", a sam puzzel jest przesuwany w lewo, a to wszystko dzięki cudownej właściwości dodawania liczby ujemnej do dodatniej.

9. Przypadek nr 6



Rys. 13. Przypadek nr 6

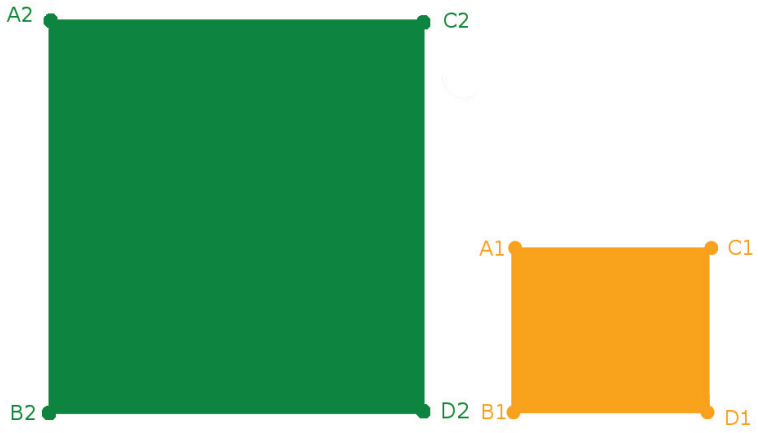
W punkcie szóstym punkty najbliżej siebie to punkty D2 i C1. Oprócz warunku i przesunięcia upuszczanego puzzla w dół, o wysokość puzzla na planszy, to w zasadzie nic nowego tu nie znajdziemy. Warunek:

$$(D2_x - C1_x \geq (-1 * dys)) \&\& (D2_x - C1_x \leq dys)$$
$$\&\& (D2_y - C1_y \geq (-1 * dys)) \&\& (D2_y - C1_y \leq dys)$$

Oraz czynności wykonywane po spełnieniu wcześniej wspomnianego wymagania:

$$roz = szerokość_{A2\,B2\,C2\,D2} - szerokość_{A1\,B1\,C1\,D1}$$
$$A1_x = A2_x + roz$$
$$A1_y = A2_y + wysokość_{A2\,B2\,C2\,D2}$$

10. Przypadek nr 7



Rys. 14. Przypadek nr 7

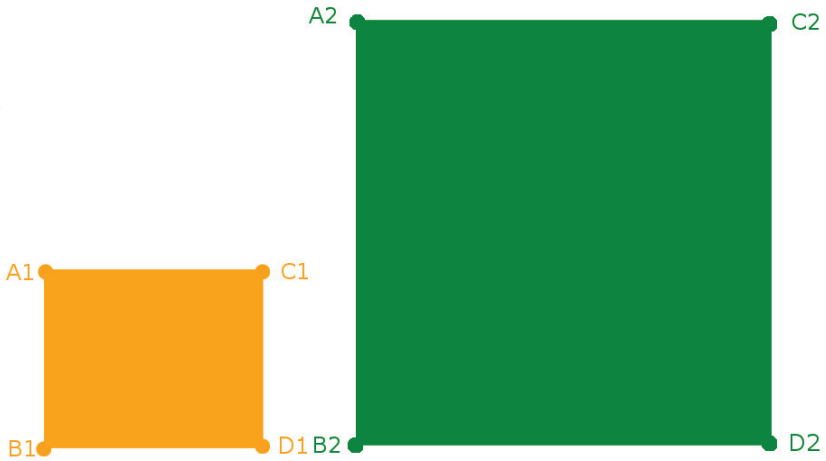
Warunek dla siódmego przypadku będzie zawierał w sobie punkty D2 i B1. Wyrażenie będzie miało postać:

$$(D2_x - B1_x \geq (-1 * dys)) \&\& (D2_x - B1_x \leq dys)$$
$$\&\& (D2_y - B1_y \geq (-1 * dys)) \&\& (D2_y - B1_y \leq dys)$$

Ustawienie podniesionego puzzla na właściwym miejscu ograniczy się do paru zabiegów. Pierwszy to przesunięcie go w prawo o szerokość puzzla na planszy, drugi natomiast to przesunięcie o różnicę wysokości jednego i drugiego puzzla. Nie zostało sprecyzowane czy odbędzie się ono w górę czy w dół, ponieważ mamy tu sytuację o podobnej charakterystyce jak w przypadku 5 i 6.

$$roz = wysokość_{A2\,B2\,C2\,D2} - wysokość_{A1\,B1\,C1\,D1}$$
$$A1_x = A2_x + szerokość_{A2\,B2\,C2\,D2}$$
$$A1_y = A2_y + roz$$

10. Przypadek nr 8



Rys. 15. rzypadek nr 8

Ósmy i ostatni przypadek. Tak samo jak para złożona z punktu D1 i B2.

$$(B2_x - D1_x \geq (-1 * dys)) \&\& (B2_x - D1_x \leq dys)$$
$$\&\& (B2_y - D1_y \geq (-1 * dys)) \&\& (B2_y - D1_y \leq dys)$$

Na końcu warto również wspomnieć o czynnościach wykonywanych w kierunku ustawienia puzzla na właściwym miejscu.

$$roz = wysokość_{A2\,B2\,C2\,D2} - wysokość_{A1\,B1\,C1\,D1}$$
$$A1_x = A2_x + szerokość_{A1\,B1\,C1\,D1}$$
$$A1_y = A2_y + roz$$

Tak oto doszliśmy do końca... artykułu, a nie samego tematu. Jeżeli chodzi o realizację tak prostej gry jak puzzle to można by stworzyć jeszcze wiele interesujących koncepcji oraz wprowadzić w życie, wiele wspaniałych pomysłów. Ten skromny tekst jest tylko wprowadzeniem oraz można na niego spojrzeć jak na jedną z wielu propozycji, którą można by było ulepszyć, bądź zrealizować w jeszcze lepszy sposób. Celem tego artykułu jest zachęcenie początkujących jak i zaawansowanych programistów do rozwijania tego typu projektów oraz do doskonalenia rzeczy, które czysto teoretycznie już powstały.

Bibliografia

[1] Kendall G.; Parkes A., and Spoerer K.: A Survey of NP-Complete Puzzles. International Computer Games Association Journal, 31(1), (2008), pp. 13–34

[2] Rodney P., Carlisle . Encyclopedia of Play. SAGE, (2 April 2009) p. 181. ISBN 978-1-4129-6670-2. Retrieved 5 October 2012.

[3] Kelley, J. A., Lugo M. The Little Giant Book of Dominoes. Sterling, 2003. ISBN 1-4027-0290-6.

Zeszyty Naukowe WSEI seria: TRANSPORT I INFORMATYKA, 4(1/2014), s. 83-96

Aleksander WÓJCIK

Wyższa Szkoła Ekonomii i Innowacji w Lublinie, e-mail: aleksander.w1992@wp.pl

NIERELACYJNE BAZY DANYCH

OBJECT DATABASES

Streszczenie

Prześledzenie historycznych sposobów zapisu danych na nośniki trwałe. Przedstawienie podstawowych założeń relacyjnego modelu danych. Omówienie jego wad i zalet, a także zwrócenie uwagi na długoletnią dominującą pozycję na rynku. Zapoznanie słuchaczy z konwencją obiektowych baz danych. Omówienie dwóch modeli (relacyjnego i nierelacyjnego) na podstawie przykładu. Omówienie terminu „NoSQL database" jako zbioru technologii wykorzystujących rozwiązania nie bazujące na relacyjnym modelu danych. Podział technologii ze względu na rodzaj modelu danych. Dyskusja zalet i wad technologii w porównaniu do relacyjnego modelu.

Summary

Presenting the historical ways of recording data on durable media. Presentation of the basic assumptions of the relational data model. Discussion of the pros and cons and the explanation of the long-term dominant position on the market. Familiarization with the convention object databases. Discussion of the two models (relational and non-relational) on the basis of an example.Discussion of the term „NoSQL database" as a collection of technologies using solutions, which do not rely on the relational data model. Division of technology due to the type of data model. Discussion about NoSQL technology advantages and disadvantages compared to the relational model.

Słowa kluczowe: NoSQL, bazy danych, relacyjny model danych, aplikacje bazodanowe, obiektowy paradygmat programowania

Keywords: NoSQL, databases, relational data model, database applications, object oriented programming paradigm

1. Model relacyjny

1.1. Przyczyny przejścia na system baz danych

Systemy informatyczne w miarę rozwoju wymagały zarządzania coraz większą ilością informacji. Początkowo dane były zapisywane w systemie plików. Jednak z wielu przyczyn nastąpiło przejście na systemy baz danych, na przykład:

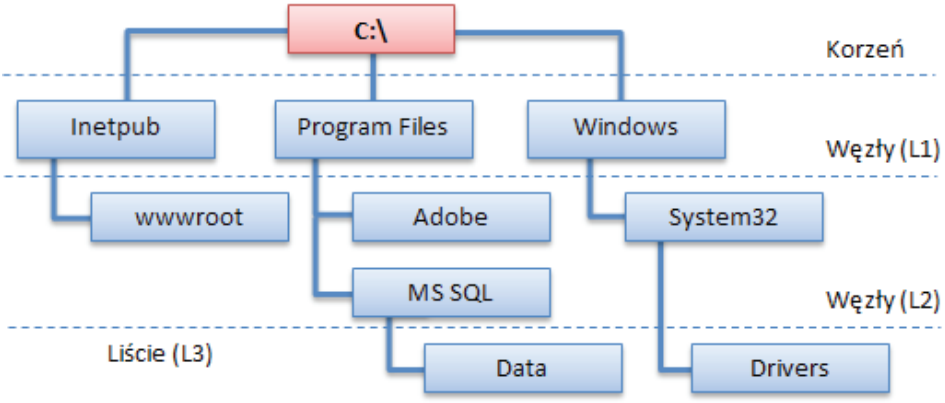
- fizyczna i logiczna niezależność danych
- przechowywanie bez redundancji
- centralna kontrola integralności
- języki na wysokim poziomie abstrakcji m.in. ujednolicony mechanizm odczytu i zapisu danych

Baza danych – to zbiór informacji wraz z możliwością łatwego dostępu oraz ich zmiany (tj. modyfikacją, dodawaniem nowych i usuwaniem starych) z poziomu aplikacji z niej korzystającej.

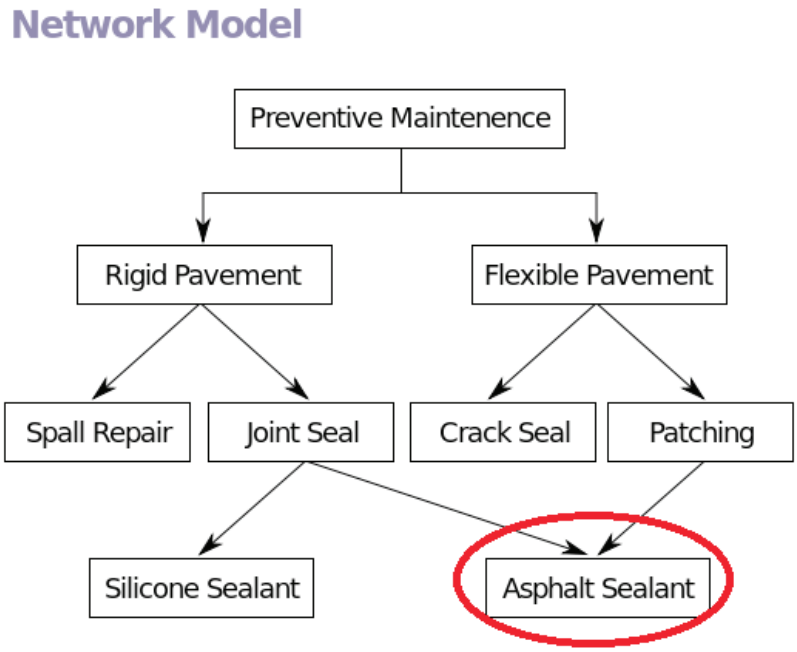
1.2. Model hierarchiczny i sieciowy

- Pierwsza generacja

Pod koniec lat 60. po raz pierwszy do opisania struktur fizycznych z logicznego punktu widzenia zastosowano matematyczne modele danych. Na podstawie pojęcia grafu opracowano modele hierarchiczny i sieciowy, które dziś mają niewielkie znaczenie.



Rys. 1. W hierarchicznym modelu danych każdy węzeł ma dokładnie jednego rodzica, oprócz węzła stojącego na szczycie hierarchii. Drzewo katalogów jest przykładem struktury hierarchicznej



Rys. 2. Sieciowy model danych jest mniej restrykcyjny od hierarchicznego. W czerwonej pętli został zaznaczonywęzeł, który jest synem dwóch węzłów

- Druga generacja

Drugą generacją nazywamy systemy oparte na relacyjnym modelu danych i relacyjnej algebrze zaproponowanych przez Edgar F. Codd'a w 1970r., które szybko zdobyły popularność. Model ten ma bardzo solidne i precyzyjne podstawy matematyczne, co jest jego ogromną zaletą.

1.3. Podstawowe założenia modelu relacyjnego

- Wszystkie wartości danych oparte są na prostych typach danych.
- Wszystkie dane w bazie relacyjnej przedstawiane są w formie dwuwymiarowych tabel (w matematycznym żargonie noszących nazwę „relacji"). Każda tabela zawiera zero lub więcej wierszy (w tymże żargonie – „krotki") i jedną lub więcej kolumn („atrybutów"). Na każdy wiersz składają się jednakowo ułożone kolumny wypełnione wartościami, które z kolei w każdym wierszu mogą być inne.
- Po wprowadzeniu danych do bazy, możliwe jest porównywanie wartości z różnych kolumn, zazwyczaj również z różnych tabel, i scalanie wierszy, gdy pochodzące z nich wartości są zgodne. Umożliwia to wiązanie danych i wykonywanie stosunkowo złożonych operacji w granicach całej bazy danych.

- Wszystkie operacje wykonywane są w oparciu o algebrę relacji, bez względu na położenie wiersza tabeli. Nie można więc zapytać o wiersze, gdzie (x=3) bez wiersza pierwszego, trzeciego i piątego. Wiersze w relacyjnej bazie danych przechowywane są w porządku zupełnie dowolnym – nie musi on odzwierciedlać ani kolejności ich wprowadzania, ani kolejności ich przechowywania.
- Z braku możliwości identyfikacji wiersza przez jego pozycję pojawia się potrzeba obecności jednej lub więcej kolumn niepowtarzalnych w granicach całej tabeli, pozwalających odnaleźć konkretny wiersz. Kolumny te określa się jako „klucz podstawowy" (ang. *primary key*) tabeli.

1.4. Trwałość modelu

Model relacyjny od 40-tu lat ma najważniejsze znaczenie i nie chce ustąpić nowym technologiom. W informatyce 40 lat to jest bardzo długi okres czasu, dlatego model relacyjny jest fenomenem jeśli chodzi o trwałość technologii.

Przyczyny:

- system relacyjny nadawał się do budowy dużych systemów informatycznych w latach 80. i bardzo dużo kodu zostało w nim napisane;
- został rozwijany i wdrażany przez dwie bardzo duże firmy informatyczne: IBM i Oracle;
- ma wyjątkowo prosty interfejs komunikacji: select, insert, update i delete, natomiast w systemach NoSQL lub w XML-owych bazach danych wymagana jest znajomość programowania do operacji na danych.

1.5. MINUSY MODELU RELACYJNEGO

Już kilka lat po spopularyzowaniu się modelu relacyjnego zaczęła się szeroka krytyka modelu. Okazał się on nieelastyczny i trudny w modelowaniu rzeczywistości, z następujących powodów:

1. Brak typów złożonych
2. Atrybuty złożone (np. adres) nie mogą być reprezentowane bezpośrednio – składowe muszą być deklarowane indywidualnie jako atrybuty
3. Powiązania pomiędzy tabelami tylko poprzez klucze obce
4. Nieelastyczność modelu, brak możliwości rozszerzenia

Szczególnie uciążliwe w modelowaniu rzeczywistości jest:

- Potrzeba definiowania klucza sztucznego, gdy atrybuty nie są wystarczające do uzyskania unikatowej identyfikacji (np.nazwa firmy może się powtarzać)
- Atrybuty zbiorowe (np. pracownicy) muszą być rozróżnialne od atrybutów jednowartościowych i reprezentowane w innym schemacie relacyjnym
- Agregacje i specjalizacje nie są w łatwy sposób obsługiwane i wymagają specjalnych więzów integralności

2. Obiektowe bazy danych

2.1. Problemy z modelem relacyjnym

a) Niekompatybilność

Główną przyczyną szukania innych rozwiązań niż relacyjny model danych jest konflikt występujący w aplikacjach bazodanowych tj. niezgodność pomiędzy modelem umożliwiającym przechowywanie danych (relacyjnym) a modelem, w którym implementowany jest program. Większość dzisiejszych aplikacji jest implementowanych w językach Java, C# lub C++, które bazują na paradygmacie obiektowym.

b) Brak jednoznacznego paradygmatu

Zarówno standard SQL2 jak i założenia paradygmatu obiektowego są jednoznaczne. Jednak przy próbie ”połączenia” paradygmatów i wypracowania jednolitego standardu pojawiają się duże rozbieżności. Nie jest znany żaden ogólnie, szeroko przyjęty paradygmat obiektowych baz danych, który by dobrze łączył dwa oddzielne paradygmaty. Różnice pomiędzy językami obiektowymi (np. C++ a Java) są dużo mniejsze niż pomiędzy poszczególnymi systemami obiektowych baz danych.

2.2. Rozwiązania

Istnieją dwa główne podejścia do projektowania języka, który by rozwiązał problem niekompatybilności pomiędzy modelem danych (relacyjnym), a językiem implementacji programu (obiektowym):

1. Zbliżenie języka proceduralnego SQL do języków obiektowych poprzez uwzględnienie zasad funkcjonowania obiektów. Dokładniej: stworzenie języka baz danych rozszerzonego o funkcje obiektowe. Przykładem takiego rozwiązania jest język SQL 3.0.

2. Zbliżenie możliwości języków obiektowych do SQL poprzez włączenie funkcjonalności baz danych. Dokładniej: rozszerzenie obiektowych języków programowania o funkcje typowe dla baz danych poprzez dołączenie klas bibliotek.

3. Kolejnym pomysłem jest użycie mapowania obiektowo-relacyjnego (tzn. system ORM lub bazy relacyjno-obiektowe), które polega na manipulowaniu danymi jako zestawem obiektów, ale użycie bazy relacyjnej jako wewnętrzny mechanizm przechowywania danych.

2.3. Charakterystyka obiektowej bazy danych

W obiektowej bazie danych przechowywane są obiekty zamiast wierszy tabeli jak w modelu relacyjnym. Umożliwia to łatwiejszą integracje z obiektowymi językami programowania. Na obiekt składa się identyfikator obiektu (adres pamięci), który jest unikalny oraz wartości jego wszystkich pól. W szczególności obiekty

mające te same wartości pól mogą być różnymi obiektami (wskazywać na inne adresy pamięci) jak i tymi samymi obiektami (wówczas obiekty są referencjami do tego samego adresu pamięci). W języku Java obiekty (Object o1, Object o2) porównuje się za pomocą dwóch narzędzi:

1. Porównanie: wyrażenie

   Zwraca wartość prawda, kiedy o1 i o2 to ten sam obiekt, tzn. obiekty wskazują na ten sam adres pamięci

2. Metoda equals: wyrażenie

   (o1.equals(o2));

Dla większości standardowych klas zwraca wartość prawda, jeśli wartości wszystkich pól są identyczne, jednak programista może tą metodą dowolnie nadpisać w klasach niefinalnych.

2.4. Przykłady obiektowych systemów baz danych [1]

- **Illustra** Rozwinięcie Postgresu. Rozszerza pojęcia relacyjne przy pomocy pojęć charakterystycznych dla obiektowości
- **ObjectStore**. Rozszerza obiektowy język C++, dodając trwałość
- **GemStone** (rozszerzenie Smalltalk-u)
- **Oracle**

Zauważmy, że najpopularniejsze języki na rok 1997, czyli C++ oraz Smalltalk posiadały biblioteki obsługujące obiektowe bazy danych. Również znane firmy z rynku relacyjnych baz danych: Postgres oraz Oracle miały swoje obiektowe wersje. Po roku 2000. zainteresowanie obiektowymi bazami danych spadło na rzecz systemów mapowania obiektowo-relacyjnego.

2.5. Standard SQL 3.0.

Standard języka SQL 2.0. z r. był standardem długo obowiązującym, do którego większość firm się dostosowała wprowadzając swoje własne dialekty języka SQL. Nowy standard SQL 3.0. miał w założeniu wprowadzać do języka SQL cechy obiektowe. W prezentacji [9] z 1998 r. autor przekonuje, że standard nie zostanie przyjęty w realnych zastosowaniach, mimo potencjalnych dużych możliwości i elastyczności standardu. Standard przewiduje:

- Obiektowość: SQL3 reprezentuje podejście "hybrydowe", dodając niektóre cechy obiektowości (takie jak ADT) do tablic znanych z systemów relacyjnych.
- Rozszerzalność: umożliwienie użytkownikom deklarowania własnych typów.
- Niekonwencjonalne typy danych: multimedialne, przestrzenne, temporalne.
- Pełne możliwości uniwersalnego języka programowania dla definiowania i zarządzania trwałymi, złożonymi obiektami.
- Rozszerzenia w zakresie aktywnych reguł, interfejsów do innych języków programowania, autoryzacji, procedur bazy danych, ewolucji schematu, i inne.

- Deklarowana jest zgodność 'w dół" ze standardem SQL-92.
- W nowym standardzie SQL3 twórcy próbują wyeliminować wszelkie wady jego poprzedników. Jednocześnie, próbują wszystko z nich pozostawić.
- Standard jest ogromny, według różnych szacunków 1200-1600 stron, zaś jego poszczególne części są niezbyt spójne.

3. Nierelacyjne bazy danych

3.1. Termin NoSQL

Termin NoSQL czyli Not Only SQL – określa systemy zarządzania, które nie bazują na modelu relacyjnym. Konsekwencją odrzucenia modelu relacyjnego jest to, że dane przechowywane w bazie danych nie wymagają ściśle określonych schematów (np. tabel), w wielu przypadkach nie ma w nich złączeń dzięki czemu umożliwiają łatwe skalowanie w poziomie, a co za tym idzie realizacja zapytań jest efektywniejsza. Termin NoSQL często jest mylony z jedną konkretną technologią. W rzeczywistości można go określić jako zbiór rozwiązań służących do przechowywania danych, które stoją w opozycji do modelu relacyjnego.

Termin „Not Only SQL" podkreśla, że systemy mogą dopuszczać elementy SQL-a, w szczególności języki zapytań podobne do SQL-owych.

3.2. Założenia ruchu NoSQL

- Standaryzacja interfejsu dostępu do baz NoSQL5
- Eliminacja najsłabszych rozwiązań. Na rynku muszą pozostać tylko najlepsze rozwiązania. Obecnie można wybierać spośród ponad 50 technologii. Nie sposób ich wszystkich przetestować.
- Problem z budową zapytań. Aktualnie tworzenie takich zapytań wymaga znajomości programowania
- Zagwarantowanie wsparcia dla swoich rozwiązań. Na razie są to rozwiązania typu open source, tworzone najczęściej przez małe firmy lub pojedyncze osoby. Istnieje potrzeba zaangażowania dużych firm związanych z bazami danych.

3.3. Założenia technologii NoSQL

Rezygnacja z wielu elementów baz relacyjnych. Zauważono, ze duża liczba złączeń tabel powoduje zdecydowany spadek wydajności, a ścisły schemat bazy danych nie zawsze bywa zaletą, gdyż wiele danych nie ma określonej struktury.

Zmniejszenie znaczenia schematów danych. Wg założeń NoSQL uwaga powinna być skupiona najpierw na danych, a dopiero potem na schematach.

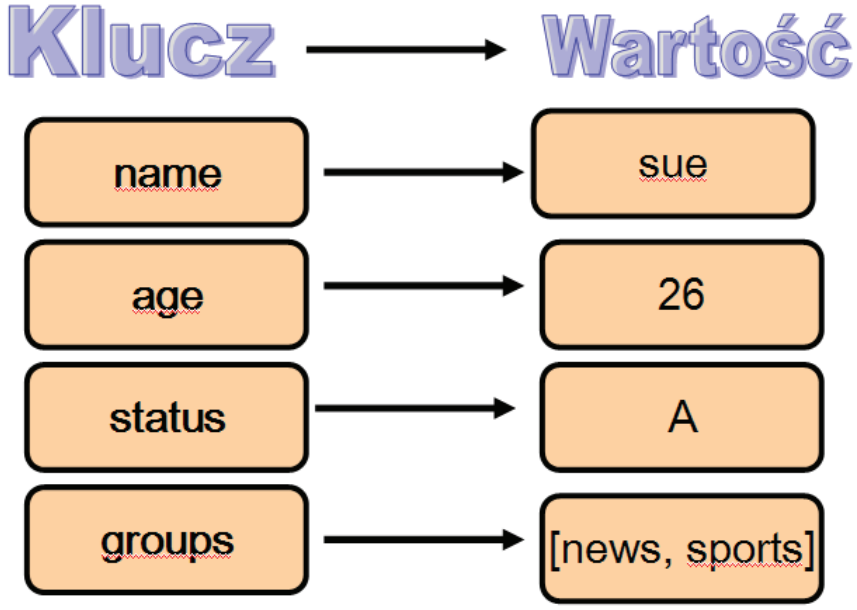
Odejście od postulatów ACID. Stwierdzono, że postulaty są zbyt restrykcyjne.

- Atomicity – atomowość

- Consistency – spójność
- Isolation – izolacja
- Durability – trwałość

3.4. Niektóre modele baz NoSQL

a) Bazy klucz-wartość (ang. *key-value*). W dużym uproszczeniu są to tabele, zawierające dwie kolumny tekstowe: klucz oraz wartość. Główną zaletą tego modelu jest to, że jest niezwykle szybki (zarówno jeśli chodzi o zapis jak i o odczyt danych). Natomiast wadą jest mała możliwość zastosowania takich baz w codziennym użyciu, ze względu na małe możliwości segregacji danych. Przykładem danych, które mogą być przechowywane w takich tabelach jest usługa DNS oferująca translację nazwy mnemonicznej serwera (np. pl.wikipedia.org) na adres ip (91.198.174.232). Bazy dokumentowe są uogólnieniem baz typu klucz-wartość i mają szersze zastosowanie.



b) Bazy kolumnowe. Ich główną ideą jest zmiana sposobu postrzegania danych. Dane zamiast zapisywać dane w wierszach, zapisuje się je w kolumnach.

c) Bazy obiektowe.

d) Bazy dokumentowe. W ostatnim czasie są szczegółowo rozwijane. W bazach tego typu zamiast tradycyjnych wierszy używa się pojęcia dokumentu, zawierającego pary klucz-wartość. Rozwiązanie to jest bardzo elastyczne, a co za tym idzie dzięki nim możliwe jest bardzo wierne odtwarzanie rzeczywistych danych w systemach informatycznych. Przykładem takiej bazy jest coraz bar-

dziej popularne ostatnio MongoDB z którego korzystają już New York Times, Disney, MTV Networks, IGN Entertainment i The Guardian. Lista ta coraz szybciej się wydłuża. Co miesiąc baza pobierana jest około 100 tys. razy. Projekt szybko się rozwija dzięki wielu dotacjom. W niedawno opublikowanej wersji 2.0 znacząco podniosła wydajność, co doceniło już wiele firm. Wg John A De Goes [9] projekt jest dziś wart 1.2 mld $.

e) Grafowe bazy danych.
f) XML-owe bazy danych.

3.5. Porównanie prezentacji danych na podstawie przykładu:

Załóżmy, że chcielibyśmy zapisać w bazie następujące dane [A]:
Sue ma 26 lat, status A oraz jest zapisana do grup „news" oraz „sports"
John ma 24 lat, status B oraz jest zapisany do grup „news" oraz „sports"
Al ma 18 lat, status D oraz jest zapisany do grupy „politics"

Zastosujmy do tego różne modele danych:

1. Model relacyjny

Będziemy mieli na pewno tabelę People, w której będziemy przechowywać dane o osobach. Model relacyjny wymusza na nas stosowanie klucza głównego. Mimo, że imiona się nie powtarzają, klucz główny musi być liczbą naturalną, zatem ponumerujemy osoby: 1,2,3 za pomocą dodatkowego pola-klucza sztucznego *person_id*.

W tabeli znajdą się poza tym: informacja o wieku, statusie oraz grupach. Ponieważ ostatnia wartość jest atrybutem zbiorowym, musimy wprowadzić tabelę Group, w której trzymamy wszystkie nazwy grup, do której osoby mogą należeć. Każdą taką nazwę numerujemy za pomocą klucza sztucznego. W tabeli People zapisujemy wszystkie występujące kombinacje grup, do której należy każda osoba, tzn. ciąg kluczy obcych do tabeli Group.

Jeśli byłaby potrzeba przechowywać informacje, że osoba może mieć dwa statusy, to również trzeba by było wprowadzić osobną tabelę Status, a w tabeli People przechowywać tylko identyfikatory do odpowiednich wierszy tabeli Status.

Dane [A] można prezentować w tabeli w programie Microsoft Excel:

	A	B	C	D	E
1	person_id	name	age	status	groups
2	1	Sue	26	A	1,2
3	2	John	24	B	1,2
4	3	Al.	18	D	3

Rys. 4. Tabela People z wypełnionymi danymi

Zwróćmy uwagę, że informacja kto należy do jakiej grupy jest zupełnie nieczytelna bez tabeli Group.

	A	B
1	group_id	name
2	1	"news"
3	2	"sports"
4	3	"politics"

Rys. 5. Tabela Group z wypełnionymi danymi

2. Bazy obiektowe

W bazie obiektowej będziemy mieli definicję klasy Person. Wystarczy utworzyć obiekt klasy Person, przypisać mu odpowiednie atrybuty oraz zapisać stan osoby do bazy za pomocą interfejsu, który dostarcza konkretna technologia.

```
package others;

import java.util.List;
@Data
public class Person {
    private String name;
    private Integer age;
    private String status;
    private List<String> groups;
}
```

Przykładowa klasa Person napisana w języku Java. Dzięki adnotacji @Data z biblioteki projektu lombok [10] nie trzeba generować getterów, setterów, ani nadpisywać metody toString(). Kod wygląda zwięźlej. W przypadku nie korzystania z tej biblioteki należy za pomocą skrótu alt+Insert wygenerować odpowiedni kod w środowisku IDE NetBeans, Eclipse lub Intellij IDEA.

3. Bazy dokumentowe

W bazach dokumentowych dokumentem nazywamy zbiór par klucz-wartość. W naszym przypadku jeden dokument będzie odpowiadał jednej osobie.

```
{
    name: Sue
    age: 26
    status: A
    groups: ["news","sports"]
},{
    name: John
    age: 24
    status: B
    groups: ["news","sports"]
},{
    name: Al
    age: 18
    status: D
    groups: ["politics"]
}
```

W bazach dokumentowych sposób prezentacji i zapisu danych jest najbardziej czytelny. Nawiasy klamrowe wyznaczają granicę jednego dokumentu – w tym przypadku jednej osoby. Poza tym bazy dokumentowe należą do najszybszych.

4. XML-owe bazy danych

W XML-owych bazach danych podstawowym formatem zapisu jest format XML. Zaczyna się on deklaracją wersji formatu XML oraz kodowania. Następnie dane są przechowywane w jednym drzewie znaczników.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<people>
    <person>
        <name>Sue</name>
        <age>26</age>
        <status>A</status>
        <groups>
            <group>news</group>
            <group>sports</group>
        </groups>
    </person>
    <person>
        <name>John</name>
        <age>24</age>
        <status>B</status>
        <groups>
            <group>news</group>
            <group>sports</group>
        </groups>
    </person>
    <person>
        <name>Al</name>
        <age>18</age>
        <status>D</status>
        <groups>
            <group>politics</group>
        </groups>
    </person>
</people>
```

Format XML jest prawie tak samo czytelny jak w bazach dokumentowych. Minusem jest jego mała zwięzłość – do zapisu tych samych danych jest przeznaczana duża ilość tekstu. Format jest na tyle popularny, że istnieją programy, które automatycznie formatują kod XML oraz podświetlają znacznik zamykający i otwierający.

3.6. Wady i zalety nierelacyjnych baz danych

Biorąc pod uwagę powyższy przykład można powiedzieć, że sposób zapisu i prezentacji w nierelacyjnych bazach danych jest prostszy, elastyczniejszy, przejrzystszy i łatwiej czytelny dla człowieka niż w modelu relacyjnym.

Warto też zwrócić uwagę na jeden drobny, ale ważny fakt, który porusza autor postu na portalu javadzone [8]. Model relacyjny był w ogromnej większości przedsiębiorstw i uczelni używany jako jedyny, a nierelacyjne bazy danych dopiero niedawno zaczęły pojawiać się w nowszych zastosowaniach. Wobec tego

jest bardzo dużo dobrych, przetestowanych i popularnych narzędzi ułatwiających pracę z relacyjnym światem. Natomiast o ile silniki do obsługi nierelacyjnych baz (tzn. przetwarzanie, odczytywanie i zapisywanie danych) są dopracowane i działają niezawodnie i bardzo szybko, to programiści nie mają odpowiednich narzędzi do wygodnej i szybkiej pracy z kodem aplikacji. Powoduje to, że programistom ciągle wygodniej się pracuje z relacyjnymi bazami, pomimo, że chętnie zdecydowaliby się na inny model danych. Autor artykułu nazywa tą niedogodność „piętą achillesową" (ang. *Achilles heel*) technologii NoSql.

Przedstawione zostanie jeszcze jedno porównanie różnych modeli danych pod względem szybkości i elastyczności.

Data Model	Performance	Scalability	Flexibility	Complexity	Functionality
Key–Value Store	high	high	high	none	variable (none)
Column-Oriented Store	high	high	moderate	low	minimal
Document-Oriented Store	high	variable (high)	high	low	variable (low)
Graph Database	variable	variable	high	high	graph theory
Relational Database	variable	variable	low	moderate	relational algebra

Wg Ben'a Scofield'a [7] relacyjny model danych jest najmniej elastyczny i razem z grafowymi bazami danych osiąga najmniejszą wydajność. Funkcjonalność baz key-value jest nieokreślona lub żadna, ale ma najwyższą wydajność i najmniejszą złożoność.

4. Podsumowanie

Relacyjny model danych nie był pierwszym modelem danych, ale pierwszym, który przyjął się na bardzo szeroką skalę i od 40-stu lat jest stosowany w większości zastosowań. Do głównych jego wad należy trudność w modelowaniu rzeczywistości, nieelastyczność oraz brak typów złożonych. Największą zaletą jest stabilna pozycja na rynku i dostosowanie się szeregu programistycznych narzędzi do modelu.

Wraz ze wzrostem popularności języków obiektowych w latach 90. wzrosło zainteresowanie i nadzieje w związku z obiektowymi bazami danych. Dziś stanowią one bardzo mały odsetek używanych technologii, nie przyjęły się komercyjnie. Firmy zdecydowały się stosować rozwiązanie mapowania obiektowo-relacyjnego, które w ostatnim czasie jest intensywnie wdrażane w wielu zastosowaniach.

Obecnie istnieje ponad 50 różnych technologii bazujących na nierelacyjnym modelu danych oraz osiągają one pierwsze komercyjne sukcesy. W technologii ukryty jest bardzo duży potencjał, a każdy sukces jest dostrzegany w informatycznych blogach i czasopismach. Brak wielu profesjonalnych narzędzi do obsługi technologii oraz niejednolity i skomplikowany interfejs dostępu do danych są trudnościami, z powodu których technologia nie jest aż tak chętnie wybierana jakby mogła być.

Bibliografia

[1] Lausen G., Vossen G.: Obiektowe bazy danych. WNT, Warszawa, 2000, ISBN 83-204-2487-9.

[2] http://wazniak.mimuw.edu.pl/index.php?title=Bazy_danych.

[3] http://wazniak.mimuw.edu.pl/index.php?title=Zaawansowane_systemy_baz_danych

[4] http://edu.pjwstk.edu.pl/wyklady/.

[5] http://cts.com.pl/Aktualnosci/Co-to-jest-NoSQL.

[6] http://webhosting.pl/Nie.skreslajcie.NoSQL.3.przyklady.zastosowania.z.sukcesem.nierelacyjnych.baz.

[7] http://en.wikipedia.org/wiki/NoSQL.

[8] http://java.dzone.com/articles/achilles-heel-nosql.

[9] Subieta K.: SQL3: Nowy standard języka SQL. Instytut podstaw informatyki PAN, Warszawa.

[10] http://projectlombok.org/features/Data.html.

Zeszyty Naukowe WSEI seria: TRANSPORT I INFORMATYKA, 4(1/2014), s. 97-102

Łukasz WITKOWSKI

Wyższa Szkoła Ekonomii i Innowacji w Lublinie, e-mail:luka78@gmail.com

PODCASTING I VIDEOCASTING JAKO POMOC W NAUCZANIU Z WYKORZYSTANIEM TECHNIK KOMPUTEROWYCH

PODCASTING AND VIDEOCASTING IN TECHNOLOGY-ENHANCED LEARNING COURSES

Streszczenie

Praca poświęcona jest przedstawieniu technologii podcasting i videocasting oraz omówieniu zalet i przeszkód w zastosowaniu tych technologii w procesie dydaktycznym. Praca analizuje tematy i typy wykładów, które mogą być realizowane z użyciem podcastingu lub videocastingu.

Summary

The paper discusses using podcasting and videocasting in Technology-Enhanced Learning courses. Background information regarding podcasting and videocasting is provided. Paper discusses what topics and type of lectures could be covered, also presents podcasting and videocasting as an easy to implement methods of involving students in the learning process.

Słowa kluczowe: podcasting, videocasting, e-learning

Keywords: podcasting, videocasting, e-learning

1. Podcasting i videocasting

1.1. Wstęp

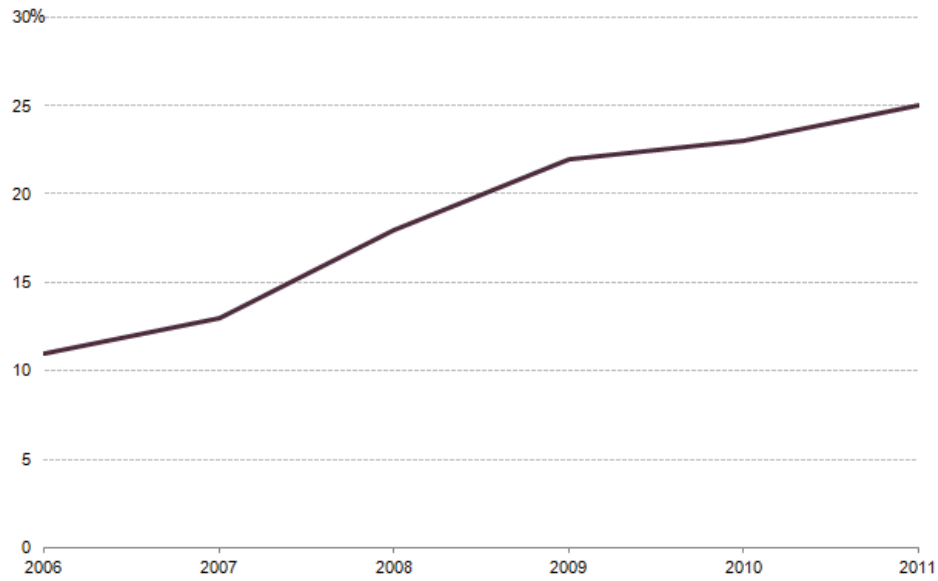
Podcasting to technologia nadawania sygnału audio, w regularnie publikowanych w Internecie częściach (epizodach).

Termin „podcast" powstał po połączeniu dwóch słów: „broadcasting" czyli „przekaz", „nadawanie" i „iPod" – jeden z najpopularniejszych odtwarzaczy plików audio. Mimo tego, że w słowie „podcasting" występuje część słowa „iPod" nie jest koniecznym posiadanie tego odtwarzacza by móc słuchać podcastów. Użytkownik ma możliwość odsłuchu audycji na każdym odtwarzaczu plików mp3, smartfonie, tablecie lub komputerze.

1.2. Dystrybucja

Pliki z poszczególnymi odcinkami podcastu zapisywane są w formacie mp3. Do powiadamiania o nowych odcinkach podcastu używane są kanały RSS. Plik z feed'em podcastu zapisywany jest w formacie xml.

Wraz z rosnącą popularnością podcastów (Rys. 1) zwiększa się też zainteresowanie tym zjawiskiem producentów sprzętu, twórców rozwiązań ułatwiających publikowanie materiału w Internecie.



Rys. 1. Odsetek Amerykanów słuchających podcastów przynajmniej raz w miesiącu [2]

Największy wybór darmowych podcastów znajduje się w sklepie iTunes firmy Apple. Dlatego też twórcy chcący dotrzeć do jak największej grupy słuchaczy na-

mawiani są do stosowania specyficznych dla sklepu iTunes znaczników RSS [6]. Przykładowe znaczniki podano poniżej:

```
<itunes:image href="http://wsei.lublin.pl/pod/log.jpg" />
<itunes:category text="Technology">
<itunes:category text="Gadgets"/>
```

Główna podstrona programu iTunes poświęcona podcastom w bardzo przyjazny i przejrzysty sposób prezentuje poszczególne odcinki danego podcastu (Rys. 2). Słuchacz widzi wszystkie opublikowane epizody, tytuły, datę publikacji, oraz dane dotyczące popularności danego odcinka.

9	PD317-2012-05-07	7 May, 2012	Lech Walesa w Detroit, GM, Kwa…	*i*	
10	PD316-2012-04-22	22 Apr, 2012	PD is back!	*i*	
11	PD315-2012-01-09	9 Jan, 2012	NAIAS 2012: Ford, Chrysler, Chevr…	*i*	
12	PD314-2011-12-31	31 Dec, 2011	2011/2012	*i*	
13	PD313-2011-12-12	12 Dec, 2011	Detroit Lives!, Ross Capicchioni, N…	*i*	
14	PD312-2011-11-25	25 Nov, 2011	Kindle Fire	*i*	
15	PD311-2011-11-13	13 Nov, 2011	Polska Szkola w Ann Arbor, MI	*i*	
16	PD310-2011-10-31	31 Oct, 2011	Sport w Ameryce	*i*	
17	PD309-2011-10-16	16 Oct, 2011	Procter and Gamble Polska	*i*	
18	PD308-2011-10-02	2 Oct, 2011	Wolontariat Euro 2012, DUI Ben'a …	*i*	
19	PD307-2011-09-18	18 Sep, 2011	Sluzace, Adamek - Kliczko	*i*	
20	PD306-2011-09-04	4 Sep, 2011	DOMA, Amerpol, Taurus	*i*	
21	PD305-2011-08-28	28 Aug, 2011	Dick Purtan, woda, Believe in Detroit	*i*	
22	PD304-2011-08-16	16 Aug, 2011	wZlot Polskich Podcasterow 2011	*i*	
23	PD303-2011-08-07	7 Aug, 2011	Woodstock, Meskie Granie, Plus i …	*i*	
24	PD302-2011-07-29	29 Jul, 2011	Polska na minus	*i*	
25	PD301-2011-07-24	24 Jul, 2011	Polska na plus	*i*	
26	PD300-2011-07-11	11 Jul, 2011	300xPD, 4 Lipca w Ambasadzie U…	*i*	
27	PD299-2011-07-03	3 Jul, 2011	Robert Gamble - Media Rodzina: …	*i*	

Rys. 2. Przykładowy ekran podcastu w serwisie iTunes firmy Apple [3]

Dystrybucja videocastów odbywa się podobnie – jedyna różnicą jest treść, rozmiar i wymagania sprzętowe. Videocasty to pliki video – ich objętość jest w większości przypadków większa niż plików audio, do obejrzenia potrzebny jest sprzęt posiadający odpowiedni ekran (iPad, tablet, komputer) [5].

2. Podcasting w procesie dydaktycznym

2.1. Wstęp

Nauczyciele dysponują obecnie wieloma możliwościami wyboru właściwych pomocy naukowych, które w doskonały sposób będą uzupełniać ich własne metody nauczania. Pomoce, które będą atrakcyjne dla studentów przy jednoczesnym spełnieniu podstawowego wymogu – uzyskania jak najlepszych wyników nauczania.

2.2. Zalety i wady

Zastosowanie technologii podcastingu lub videocastingu w usprawnieniu procesu nauczania posiada wiele zalet i kilka wad. Najważniejszą zaletą jest połączenie tradycyjnych wykładów z nauką na odległość z wykorzystaniem dostarczania treści za pomocą omawianych technologii. Ważnym jest, by w prawidłowy sposób zestawić ilość materiałów prezentowanych w podcaście bądź videocaście z tradycyjnymi materiałami – wykładami, seminariami, pracą grupową [1]. Podcasting nadal musi być uważany za narzędzie pomocnicze. Większość zagadnień poruszanych na wykładach może być przedstawionych w formie audio lub video. Ograniczenia prawie nie istnieją. Wyobraźnia, umiejętność prezentacji materiału i czas to bariery w przygotowaniu niesztampowego i wciągającego materiału, jednakże te przeciwności mogą być przezwyciężone przy niewielkim zaangażowaniu twórcy.

Podcasting to perfekcyjne rozwiązanie dla ustnych opracowań. Wykłady multimedialne mogą być zapisywane w formie videocastów. Jednocześnie zajęcia wymagające aktywnego zaangażowania studentów, takie jak dyskusje, laboratoria, czy debaty, nie mogą być prezentowane w formie podcastów bądź videocastów ze względu na ograniczenia interakcji uczestników. Burze mózgów czy studia przypadku są niezmiernie interesujące, ale nadal interakcja pomiędzy nauczycielem a studentami jest wymagana. Omawiane technologie nie mogą być wykorzystywane we wszystkich rodzajach zajęć, nie mają zastosowania w nauczaniu każdego zagadnienia. Sposób przekazania wiedzy i stosunek użycia tradycyjnych materiałów wraz z nowymi technologiami musi być dobierany w każdym przypadku indywidualnie.

2.3. Przykładowe zastosowania

Podcasting i videocasting oferuje łatwe do implementacji metody zaangażowania studentów w proces kształcenia. We współczesnym świecie łatwy dostęp jest kluczem do sukcesu. Studenci to zazwyczaj młodzi ludzie, którzy podążają za najnowszymi trendami w technologii. Książki i inne tradycyjne materiały naukowe są zawsze mile widziane, jednakże studenci domagają się także dostępu do wiedzy przekazywanej w ciekawy, atrakcyjny sposób. Istnieje wiele sposobów by sprawić, że zajęcia staną się atrakcyjniejsze, bardziej dostępne – dzięki podcastingowi i videocastingowi.

Jednym z najbardziej popularnych sposobów jest dostarczenie obszernego podsumowania wykładu w formie podcastu; student biorący udział w wykładzie i odsłuchujący podsumowania w formie audio zrozumie poruszany temat o wiele lepiej.

Dostarczenie instruktażowego filmu (w formie videocastu), pokazującego krok po kroku działania podejmowane podczas zajęć w laboratorium jest kolejnym sposobem ułatwienia przyswajania materiału.

Precast to kolejny rodzaj podcastu edukacyjnego. Precast nagrywany jest przez nauczyciela w okresie poprzedzającym rozpoczęciem zajęć z danego przedmiotu. Ma za zadanie motywować do czynnego uczestnictwa w zajęciach. Może też zawierać informacje na temat danego przedmiotu [4].

Bardzo popularnym sposobem na wyjaśnienie zagadnień związanych z programowaniem jest użycie screencastu czyli zapisu obrazu ekranu komputera w pliku wideo. Student widzi na dzielonym ekranie w tym samym czasie kod źródłowy i działający program. Ta technika pozwala na lepsze zrozumienie procesu programowania.

Popularnym staje się zlecanie wykonania podcastu osobom trzecim. Istnieją zewnętrzne firmy czy osoby prywatne, świadczące usługi przygotowania, nagrania, opublikowania i promowania podcastu. Rolą podagoga jest tylko praca nad przygotowaniem wartościowego materiału dydaktycznego.

2.4. Praktyka

Od 2010 roku studenci Wyższej Szkoły Ekonomii i Innowacji w Lublinie mogą oglądać w Intranecie wiele materiałów wideo poświęconych wydarzeniom związanych z Uczelnią. Obecnie trwają pracę nad transferem materiału wideo do formatu przyjaznego videocastingowi. W drugim kwartale 2015 roku planowane jest rozpoczęcie nadawania videocastu WSEI. Jednocześnie trwa proces angażowania wykładowców w tworzenie własnych podcastów i videocastów. Poprawi to komunikację na linii student – wykładowca, a także uatrakcyjni proces nauki danego przedmiotu.

Kandydaci na studia będą mogli zapoznać się z serią podcastów omawiających m.in. historie Uczelni, ofertę dydaktyczną. Wysłuchają wywiadów z wykładowcami, studentami, absolwentami, jak również pracodawcami obecnych lub byłyh studentów WSEI. Równocześnie przykładowe wykłady będą dostępne w formie darmowych podcastów.

Jako wewnętrzne źródło informacji, pracownicy administracyjni WSEI opublikują screencast. Znajdą się w nim m.in. instrukcje dla nowych studentów w jaki sposób zainstalować oprogramowanie, używać systemu zdalnej nauki czy Intranetu Uczelni.

3. Podsumowanie

Zaangażowanie studentów w proces nauczania wymaga nie tylko poświęcenia, ale również odpowiednich zasobów intelektualnych. Połączenie większości zalet, które mają do zaoferowania zaawansowane technologie z treściami przygotowanymi przez środowisko naukowe może owocować łatwymi w odbiorze materiałami dydaktycznymi. Wraz z rosnącym zainteresowaniem podcastingiem i videocastingiem, organizacje naukowe mogą dzielić się wartościową wiedzą w o wiele atrakcyjniejszy dla dzisiejszego studenta sposób.

Bibliografia

[1] Allan B.: Blended Learning: Tools for teaching and training. Londyn, Facet Publishing, 2007.

[2] ARBITRON: The Infinite Dial 2011 5 kwietnia 2011 roku.

[3] Przykładowy ekran podcastu – podcast Polskie Detroit, program Apple iTunes, wersja 11.3 (53), 30 lipca 2014 roku.

[4] Read B. Lectures on the Go. The Chronicle of Higher Education, Section Information Technology, Volume 52, Issue 10, p. A39, 2005.

[5] Richardson W.: Blogs, Wikis, Podcasts and Other Powerful Web Tools for Classrooms. Thousand Oaks, USA, Corwin Press, 2006.

[6] Witryna internetowa firmy Apple: http://www.apple.com/itunes/podcasts/specs.html, 29 lipca 2014.

Zeszyty Naukowe WSEI seria: TRANSPORT I INFORMATYKA, 4(1/2014), s. 103-116

David VALIS

University of Defence, Kounicova 65, Brno, Czech Republic, david.valis@unob.cz

UTILISATION OF SELECTED REGRESSION FUNCTIONS FOR OIL DATA ASSESSMENT

OCENA WYKORZYSTANIA WYBRANYCH FUNKCJI REGRESJI

Summary

The paper deals with application of selected analytical methods for analysing field data from heavy off-road military vehicles. It is the vehicle engine and its oil data which are explored for further utilisation. The pieces of information from the engine oil are interpreted in form of deteriorated oil characteristics and polluting particles in oil. These pieces of information have good technical and analytical potential. However they are not always used for system condition determination therefore has not been explored well yet. Thanks to well working the diagnostics there are the data collected in some fields but never calculated and used of estimation neither prediction. The novelty is to providing inputs for helping to change e.g. the system maintenance policy, system operation and mission planning.

Keywords: tribo-diagnostics; multi--regression; oil field data; maintenance optimisation

Streszczenie

Wzrastające wymagania niezawodności i trwałości środków transportu determinują zastosowanie nowoczesnych systemów diagnostycznych oraz minimalizacji operacji obsługowych. Jednym ze sposobów pozyskania sygnałów diagnostycznych jest okresowe badanie składu oleju silnikowego. Zdaniem autora jest to jeden ważniejszych źródeł pozyskani informacji o stanie obiektu technicznego. W artykule przedstawiono metodologię wykorzystania informacji diagnostycznej pozyskanej z badania oleju silnikowego do optymalizacji czynności obsługowych.

Słowa kluczowe: trybo-diagnostyka, optymalizacja obsługiwania, regresja, analiza oleju silnikowego

1. Introduction

The growing dependability and operation safety requirements for modern equipment together with the increasing complexity and continuous attempts to reduce operation and maintenance costs might be satisfied among others by the consistent use of modern diagnostic systems. The main task of object technical state diagnostics is not only to find out incurred failures, but also to prevent from the failure occurrence with the help of sensible detection and changes localization in the object structure and in its behaviour changes. Many various approaches have been published on system diagnostics and CBM (Condition Based Maintenance).

A tribotechnical system (TTS), friction in it, wear and lubrication, and especially the outcomes of it are the subjects of our major concern. We would like to analyse the outcomes from technical diagnostics of TTS. There exists wide range of data from the TTS which are not analysed further. We find this is a pity. Our main objective is to extract maximum in-formation from the diagnostic of TTS in order to gain tools for optimising: maintenance, cost-benefit processes, operation and mission planning. The authors will apply selected mathematical tools to get some inputs into previously mentioned areas. Regarding the tribotechnical system, the basic information about tribological process, operating and loss variables are provided [2][6][20][22][23][24].

Owing to the TTS we have got a lot of diagnostic oil data. In view of tribo-diagnostics this data is considered to be the final outcome. This data can tell us a lot about lubricants/life fluids quality itself as well as about system condition. Such data are very valuable. System operation, taking the oil samples and the outcomes themselves.

The procedure and results presented below are based on standard mathematical principles – a regression function and a regression analysis. Lead oil data will be our point of interest. From both presumptions we can expect reasonable costs savings [16][18]. As from the military point of view we would like to determine remaining residual life to be able to perform a mission. Following the regression analysis it is possible among others to assess the operating history of an observed vehicle.

2. Objects of diagnostics and methods

The assumed objects of diagnostics in our case the medium lorry T810 engines have not been ready yet in terms of design to use the ON-LINE system, though in practice similar possibilities for other applications have already existed. It results from the information stated above that we are still supposed to use OFF-LINE engine diagnostics system when sampling lubrication fluid at certain intervals, and using known and optimised special tribodiagnostic methods [8].

In our case we use the results and information from atomic emission spectrometry. We concentrate on lead oil data. Following this analysis we can obtain the

information about the presence of the elements of a specific kind and the amount of elements. When evaluating data, the information is transformed many times and provides only estimated reality which might be different from reality itself. If the vagueness in classes distribution is not given by a stochastic character of measured characteristics but by the fact that the exact line among states classes does not exists, it will be later on good to using fuzzy set theory and adequate multi--criteria fuzzy logic. However, we cannot identify their real origin – e.g. as a result of fatigue, cutting or sliding.

Therefore in our further research we try to identify where the elements might come from. We base our assumptions on idea to increase the potential for maintenance optimisation inputs and cost benefit analysis inputs. Usable diagnostic approaches might be found e.g. in [5][28].

3. Oil field data assessment and mathematical model

There is enough statistically important set of field data obtained from the diagnosed objects. It fulfils the basic assumption that we might be capable to solve this problem successfully. Since the data sets are very extensive, we are not going to introduce them here except for a part/example of lead (Pb) particles representing the sample of T810 – it is presented in modified form in Table 1.

Sample/Mh	Pb particles in ppm	Sample/Mh	Pb particles in ppm
1/0	4.34	7/46	4.52
2/8	5.11	8/57	5.28
3/11	5.62	9/64	5.23
4/22	4.92	10/72	5.72
5/26	4.58	11/84	4.62
6/35	5.14	12/95	5.46

Table 1. Input data of Pb particles

We deal with dozens of samples taken and analysed at different types of observed engines. In certain aspects we consider the engine from an infantry fighting vehicle to be a reference object, because the event of a failure type has occurred in it. All tribodiagnostic processes related to the failure occurrence have been recorded. We assume that we have potential for obtaining inputs for system maintenance optimisation and system residual life estimation.

3.1. Utilization of Regression Model – Theory

When analysing diagnostic data, the question arises whether the data is described by only one regression function across all measurement interval, or there are areas of a different regression function in the data record. If the data is divided into more areas for which relevant regression functions will be calculated, in most cases the regression functions will have a different functional value at the point where the data is divided. When looking for a suitable regression function, a continuity requirement is often set. We would like to modify the form of our data regression functions so that they could have the same value at the point where the data is divided into single areas. At the beginning let us presume that the system is in two states. The diagnostic data will be therefore divided into two parts/areas and in each of them we will look for the regression functions whose functional value at a dividing point will be the same.

An independent variable is with a view of its random character represented by a random variable X (motohours), while a dependent variable is represented by a random variable Y (the number of Pb particles). The breakpoint for two parts will be defined by the value x_0. In the first area we are searching for the following regression function:

$$y = \varphi_1(x, \boldsymbol{\beta}^1) = \mathrm{E}(Y \mid X = x) \tag{1}$$

for the data for which it applies: $x < x_0$.

In the second area we are looking for the regression function as follows:

$$y = \varphi_2(x, \boldsymbol{\beta}^2) = \mathrm{E}(Y \mid X = x) \tag{2}$$

for the data for which it applies: $x > x_0$ and at the same time we require that the following equation applies:

$$\varphi_1(x_0, \boldsymbol{\beta}^1) = \varphi_2(x_0, \boldsymbol{\beta}^2) \tag{3}$$

We are looking for both regression functions for the whole area of data concurrently. As for our data, the regression functions will be searched for in a linear form and we will use a linear regression model:

$$y = \sum_{j=1}^{m} \beta_j f_j(x) \tag{4}$$

where $f_j(x)$ are known functions not containing $\beta_{1,\ldots,}\beta_{m,}$. This model is based on the assumptions listed below:

Values x are non-random so the functions $f_j(x)$ acquire non-random values $f_{ji} = f_j(x_i)$ for $j = 1,\ldots..,m$ and $i = 1,\ldots..,n$.

Matrix

$$F = \begin{pmatrix} f_{11} & \cdots & f_{1n} \\ \vdots & \ddots & \vdots \\ f_{m1} & \cdots & f_{mn} \end{pmatrix} \tag{5}$$

of the type (m, n) with the elements f_{ji} is of the rank $m < n$.

Random variable Y_i ,(i-th observation variable Y) satisfies

$$\mathrm{E}(Y_i) = \sum_{j=1}^{m} \beta_j f_{ji} \tag{6}$$

and $\mathrm{D}(Y_i) = \sigma^2 > 0$ for $i = 1,\ldots., n$

Random variables Y_i are non-correlated and have a regular probability distribution for $i = 1, \ldots.,n$. In the single areas we have selected the following regression functions:

for the first area: $m = 2, f_1(x) = 1, f_2(x) = x$, the regression function is:

$$y = \beta_1^1 + \beta_2^1 x \tag{7}$$

for the second area: $m = 2, f_1(x) = 1, f_2(x) = x$, the regression function is:

$$y = \beta_1^2 + \beta_2^2 x \tag{8}$$

For the required dependency we select the model which will fit the data and at the same time will be as simple as possible. We are going to do it like this:

Gradually we will select the value x_0 and a relevant regression function in the following manner. We are looking for the values $b_1^1, b_2^1, b_1^2, b_2^2$ (the point estimation of the parameters: $\beta_1^1, \beta_2^1, \beta_1^2, \beta_2^2$ so that the function:

$$S = \sum_{x<x_0} \left(y - \left(\beta_1^1 + \beta_2^1 x\right)\right)^2 + \sum_{x>x_0} \left(y - \left(\beta_1^2 + \beta_2^2 x\right)\right)^2 \tag{9}$$

could acquire minimum value when:

$$\beta_1^1 + \beta_2^1 x_0 = \beta_1^2 + \beta_2^2 x_0 \tag{10}$$

The minimum is marked as ${}^{x0}S^*_{\min}$.

Out of the regression functions we select the one for which the value ${}^{x0}S^*_{\min}$ is as low as possible. This value will be denoted as $S^*_{\min}$.

Limitation: To keep the model simple, in both areas we have chosen linear dependability which is line-shaped. Owing to a large data spread for low values at motohours, we have selected the value x_0 from the value 20.

3.2. Utilization of Regression Model – Results

Regression model:

$$\min_{x_0}\left\{\min_{\beta_1^1,\beta_2^1,\beta_1^2,\beta_2^2}\left\{\sum_{x<x_0}(y-(\beta_1^1+\beta_2^1 x))^2 + \sum_{x>x_0}(y-(\beta_1^2+\beta_2^2 x))^2\right\}\right\} \tag{11}$$

The lowest value $S^*_{\min}$ we have obtained for x_0 =150.2 (motohours). $S^*_{\min}$ = 293.7. The dependency ${}^{x0}S^*_{\min}$ on the value x_0 is shown in the graph in Fig. 1.

For the value x_0 = 150 we acquired the following regression coefficients values:

First area: Point estimation of the parameter

β_1^1 : b_1^1 = 1.927

β_2^1 : b_2^1 = 0.005556.

Interval estimations: 95% dependability interval for

β_1^1 : ⟨1.695; 2.159⟩

β_2^1 : ⟨0.0009998; 0.0101130⟩.

p-value for the hypothesis H: β_2^1 = 0.01711.

Second area: Point estimation of the parameter

β_1^2 : b_1^2 = 4.512

β_2^2 : b_2^2 = 0.001572

Interval estimations: 95% dependability interval for

β_1^2 : ⟨3.314; 5.711⟩

β_2^2 : ⟨-0.002976; 0.006121⟩.

p-value for the hypothesis H: $\beta_2^2 = 0.4878$.

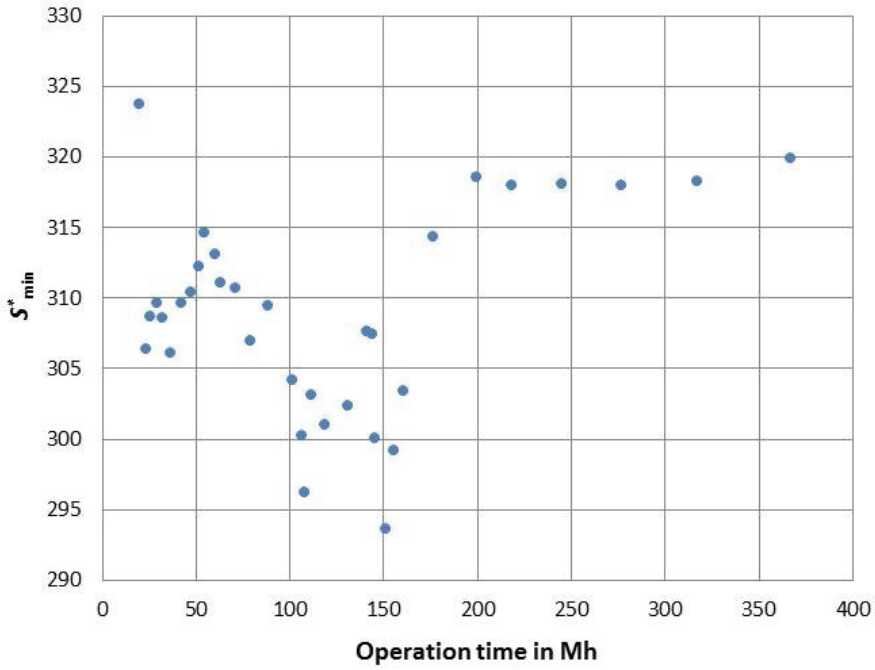


Fig. 1. The dependency of ${}^{x0}S^*_{\min}$ on the value x_0.

As for the values of regression lines, it is important to estimate the coefficient β_2. We are testing the hypothesis H: $\beta_2 = 0$. It results from p-values that in the first area the hypothesis H: $\beta_2^1 = 0$ is rejected at the confidence level of 0.05 (p-value 0.01711). As to the second area the hypothesis H: $\beta_2^2 = 0$ is not rejected at the confidence level 0.05 (p-value 0.4878). The data in the second area then might be approximated with an invariable without diminishing the mathematical value of determination.

Original intention was to create two regression courses for the two sets of clusters in data. This is presented in Fig. 2 – for individual vehicle and in Fig. 3 for mean value. This expression however does not provide continuous course of the curve which is important for further calculations.

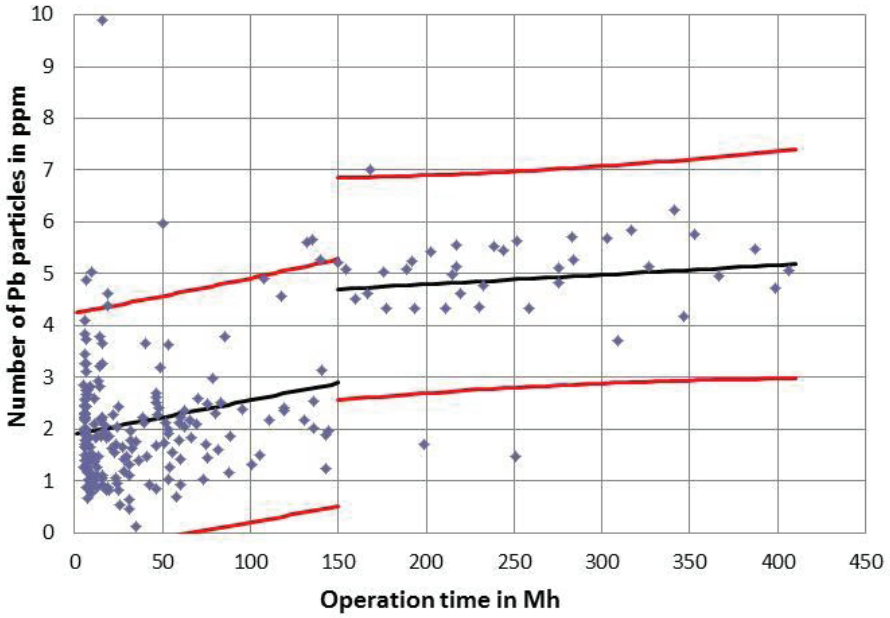


Fig. 2. Course of the broken function with 95% confidence interval for individual value Pb

Regression model:

$$\min_{x_0}\left\{\min_{\beta_1^1,\beta_2^1,\beta_1^2,\beta_2^2}\left\{\begin{array}{l}\sum_{x<x_0}(y-(\beta_1^1+\beta_2^1 x))^2+\sum_{x>x_0}(y-(\beta_1^2+\beta_2^2 x))^2\\ \beta_1^1+\beta_2^1 x_0=\beta_1^2+\beta_2^2 x_0\end{array}\right\}\right\} \tag{12}$$

Therefore the "continuous broken function" for the data was constructed. The confidence intervals might be also drawn for both an individual value *y* (see Fig. 4) and a mean value *y* (see Fig. 5).

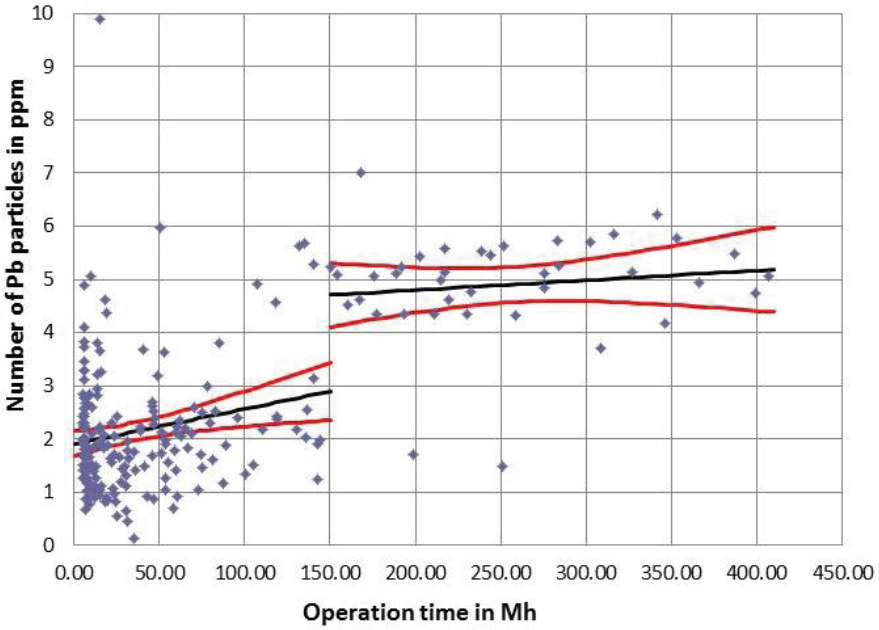


Fig. 3. Course of the broken function with 95% confidence interval for mean value Pb

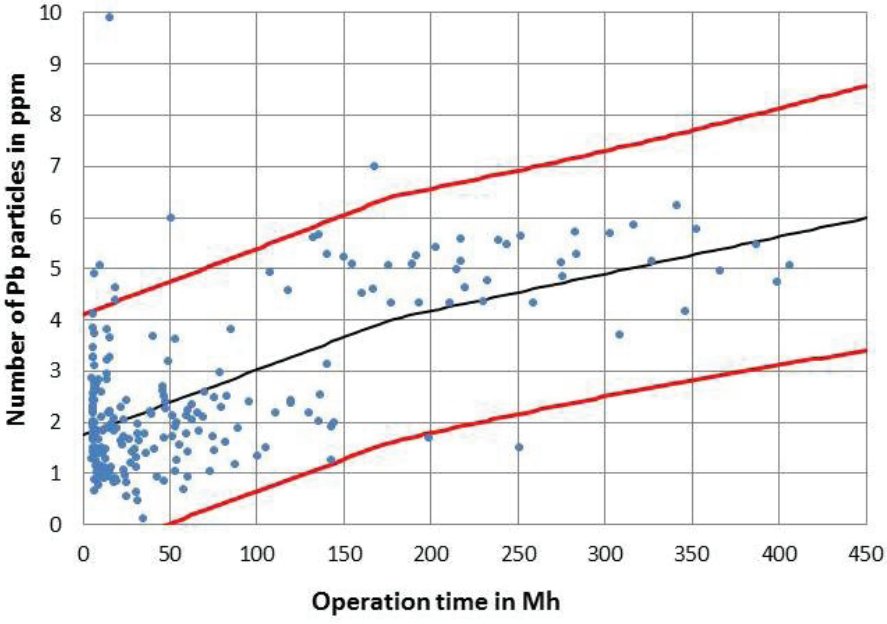


Fig. 4. Course of the broken function with 95% confidence interval for individual value Pb

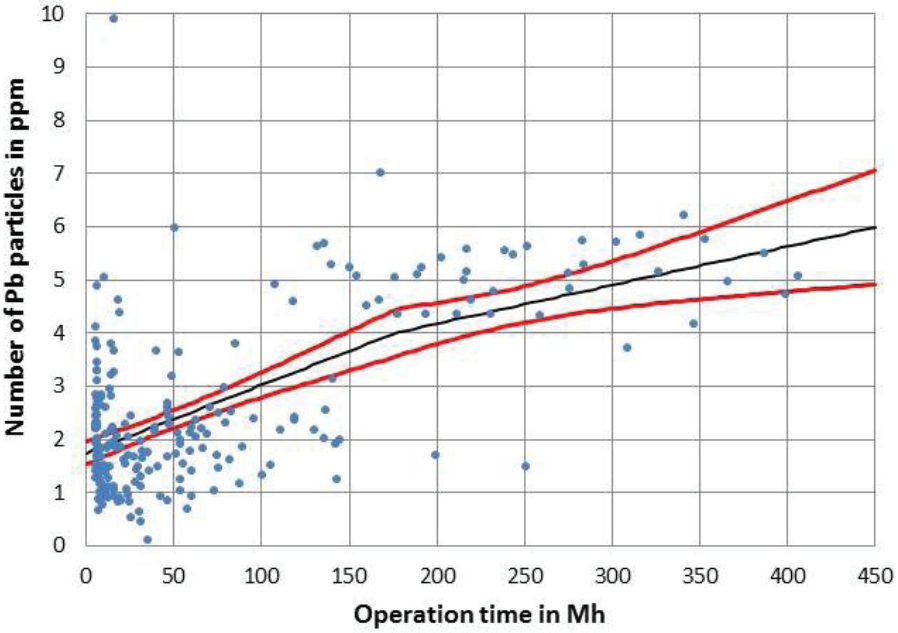


Fig. 5. Course of the broken function with 95% confidence interval for mean value Pb

While using the broken function we try to approximate the data by the regression function with more parameters. The value $S^*_{\min}$ then is lower than if all the area is approximated by the regression function of the same type. For two areas divided in x_0 =177 and regression lines we obtained $S^*_{\min}$ = 320.6. If the data is approximated by one line, then $S^*_{\min}$ = 323.8. When approximating by a second degree polynomial, there is $S^*_{\min}$ = 323.1, while approximating by a third degree polynomial, there is $S^*_{\min}$ = 303.6, and with a fourth degree polynomial there is $S^*_{\min}$ = 294.7.

Approximating by a broken function is then more appropriate than using even a fourth degree polynomial. If we do not take into account continuity requirement, the regression function will have a higher degree of freedom and $S^*_{\min}$ will get lower. For two lines we get the value $S^*_{\min}$ = 293.7.

4. Summary

It is remarkable that the Pb particles generation based on oil field data might have the broken form of course. This co-relation outcome is based on the analysis performed above and will be supported with using the fuzzy approach. The fuzzy approach has the chance to capture the course more precisely and may suit better with the coefficient of determination R^2.

The authors hope for having broadened the possibilities of extracting and utilization of pieces of information from TTS diagnostics – and respective interesting elements indicators. Although the regression analysis and expected utilization of FIS are common mathematical tools they have never been properly applied for analysis of TTS oil data. Whereas the potential of the TTS oil data is actually big valuable. The authors describe just small portion of the capabilities. Application of FIS supports our idea of describing some data generation course by selected regression forms. The authors present capabilities of the analysis results which play significant role as inputs e.g. for the system residual life estimation (RLE), maintenance optimisation, mission planning, etc. Some inspirational proposals for further development especially in terms of diagnostics and maintenance optimisation are mentioned e.g. in [1][3][4][5][7][9][10][11][12][13][14][15][17][19][21][25][27][26][28]. Using the TTS oil data from reference engine it is found that RLE might be determined for the other similar units. This approach will be further extended and developed into more precise RLE approaches.

5. Conclusion

In this paper we were looking for dependencies among measured values using statistical methods. While have applied two non-typical approaches therefore the comparison of acquired results is reasonable. At the beginning we did not know the exact theoretical background of the possible dependence of Pb particles occurrence on operating time. We were looking for this dependence using suitable approximate methods. The regression analysis and the expected use of Fuzzy Inference System serve as starting methods. When dealing with the regression analysis, it is necessary to choose a regression analysis form in advance. If the regression analysis form cannot be deduced theoretically, it is necessary to select relevant regression functions and then compare them with the measured data. In the case of fuzzy methods we will not have to select the form of an expected function, but need to determine the form and the amount of language values. They are important to set the Fuzzy Inference System. For different forms and amount of language values we obtain different dependence forms and shapes. It follows from the other results of other sets of field data obtained from different vehicle types that when combining properly both methods, we can find out that the dependence of measured data correspond with a real process. When dealing with the assessed data, it is advisable to use the FIS first, and then, following the form of a found dependence, select a relevant regression function. Despite taking a different analytical approach when applying single methods, the results are very similar to each other (see e.g. Fig. 3 and 5). It can be then assumed that the processed dependencies can be used for describing contemporary operation of the system and also the expected development under real operating conditions. The

obtained results will be used for further research, e.g. exact determination of end of the "run-in" period, the optimizing of maintenance procedures, mission planning, or the estimation of residual operating units, etc.

6. Conclusion

This paper has been prepared with the support of the Ministry of Defence of the Czech republic, Partial Project for Institutional Development, K-202, Department of Combat and Special Vehicles, University of Defence, Brno and the research project no. 3, „Management Support of Small and Middle-Sized Firms Using Mathematical Methods" of Academy Sting, Business College in Brno.

7. Acknowledgements

This paper has been prepared with the support of the Ministry of Defence of the Czech republic, Partial Project for Institutional Development and Project of Specific Research, K-202, Department of Combat and Special Vehicles, University of Defence, Brno.

Bibliografia

[1] Bartlett L. M., Hurdle E. E. and Kelly E. M.: Intergrated System Fault Diagnostics Utilising Diagraph and Fault Tree-based Approaches, Reliability Engineering and System Safety, 94/9, p. 1371-1380, 2009.

[2] Czichos H., and Habig K-H.: Tribologie-Handbuch; Reibung und Verschleiß, 2nd edition. Weisbaden: Vieweg, 2003.

[3] Edleston O.S.S.T. and Bartlett L.M.: A Tabu search algorithm applied to the staffing roster problem of Leicestershire police force, Journal of the Operational Research Society, 63/4, p. 489-496, 2012.

[4] Jodejko-Pietruczuk A., Mlynczak M.and Zajac M., Assessment of economical lifetime of heavy-duty machines, Case study? Reliability, Risk and Safety: Theory and Applications. London: Taylor & Francis Group, p. 531-534, 2010.

[5] Klimaszewski S. and Woch M.: Modified Hagg & Sankey method to estimate the ballistic behaviour of lightweight metal/composite/ceramic armour and a fuselage skin of an aircraft, Journal of KONES, 19/2, p. 245-252, 2012.

[6] Koucky M. and Valis D.: Suitable approach for non-traditional determination of system health and prognostics, Zeszyty Naukowe WSOWL, 1/159, p. 123-134, 2011.

[7] Kowalski M., Magott J., Nowakowski T. and Werbinska-Wojciechowska S., Analysis of transportation system with the use of Petri nets, Eksploatacja i Niezawodnosc – Maintenance and Reliability, 13/1, p. 48-62, 2011.

[8] Lippay J.: Tribological diagnostics of heavy of road lorries Tatra 815 engines which operate with OA-M6 ADS II oil. Inauguration Thesis, Brno: Military Academy, 1991.

[9] Nowakowski T. and Werbińska-Wojciechowska S., On problems of multicomponent system maintenance modelling, International Journal of Automation and Computing, 6/4, p. 364-378, 2009.

[10] Quigley J. and Walls L.: Trading reliability targets within a supply chain using Shapley´s values, Reliability Engineering and System Safety, 92/10, p. 1448-1457, 2007.

[11] Rak J.R. and Pietrucha K., Risk in drinking water quality control, Przemysl Chemiczny, 87/5, p. 554-556, 2008.

[12] Revie M., Bedford T. and Walls L., Supporting Reliability Decisions During Defence Procurement Using a Bayes Linear Methodology, IEEE Transactions on Engineering Management, 58/4, p. 662-673, 2011.

[13] Stodola J. and Stodola P., Operation Reliability and Diagnostics of Mechanical Systems, Transactions of Famena, 33/1, p. 47-56, 2009.

[14] Stodola J. and Stodola P.: Mechanical System Wear and Degradation Process Modelling, Transactions of Famena, 34/4, p. 19-32, 2010.

[15] Stodola J. and Stodola P.: Tribology´s Contribution to Efficient Maintenance of Military Engines, In: *ICMT´09*. Oprox, Brno, 2009.

[16] Stodola P. and Mazal J: The Trapezoid and S-curve for Real-time Servomotors Control, In Proceedings of the 15th International Conference on Mechatronics – Mechatronika 2012. Prague: Czech Technical University in Prague, pp. 165-168. 2012.

[17] Stodola P., Jamrichova Z. and Stodola J.: Modelling of Erosion Effects on Coating of Military Vehicles Component. Transactions of Famena, 36/3, p. 33-44, 2012.

[18] Stodola P., Mazal J., Mokrá I., and Podhorec M.: The Real-Time Control Algorithm and Control Curves for Servomotors, International Journal of Circuits, Systems and Signal Processing, vol. 7, no. 2, p. 118-125, 2013.

[19] Studzinski A. and Pietrucha-Urbanik K.: Risk Indicators of Water Network Operation. Chemical Engineering Transactions, 26, p. 189-194, 2012.

[20] Valis D., Koucký M. and Zak L.: On approaches for non-direct determination of system deterioration, Eksploatacja i Niezawodnosc – Maintenance and Reliability, 14/1, p. 33-41, 2012.

[21] Valis D., Vintr Z. and Koucky M.: Contribution to highly reliable items' reliability assessment, In. Proc. of the European Safety and Reliability Conference ESREL, Prague, Czech Republic. Reliability, Risk and Safety: Theory and Applications. London: Taylor & Francis 1-3, p. 1321-1326, 2010.

[22] Valis D., Vintr Z. and Malach J.: Selected aspects of physical structures vulnerability – state-of-the-art, Eksploatacja i Niezawodnosc – Maintenance and Reliability, 14/3, p. 189-194, 2012.

[23] Valis D., Zak L. and Pokora O.: On approaches for non-direct determination of system deterioration, Eksploatacja i Niezawodnosc – Maintenance and Reliability, 14/1, p. 33-41, 2014.

[24] Ver-schleiß, Reibung, Definitionen, Begriffe, Prüfung (GfT, Moers)

[25] Vintr Z. and Valis D.: Aircraft gun reliability modelling, Proc. of the European Safety and Reliability Conference ESREL, Stavanger, Norway, Risk, Reliability and Societal Safety. Proceedings and Monographs in Engineering, Water and Earth Sciences 1-3, p. 2769-2774, 2007.

[26] Werbinska S.: Interactions between logistic and operational system – an availability model, Risk, reliability and societal safety. Eds. Terje Aven, Jan Erik Vinnem. Leiden: Taylor and Francis, p. 2045-2052, 2007.

[27] Werbinska-Wojciechowska S.: Time resource problem in logistics systems dependability modelling," Eksploatacja i Niezawodnosc – Maintenance and Reliability, 15/4, p. 427-433, 2013.

[28] Woch M.: Analysis of operating loads on an aircraft's vertical stabilizer on the basis of recorded data, Chemical Engineering Transactions vol. 33, p. 643-648, 2013.

Zeszyty Naukowe WSEI seria: TRANSPORT I INFORMATYKA, 4(1/2014), s. 117-123

Iveta KUBASÁKOVÁ, Bibiána POLIAKOVÁ

Department of road and urban transport, Faculty of Operation and Economics of Transport and Communications, University of Zilina, Univerzitná 8215/1, 010 26 Žilina, iveta.kubasakova@fpedas.uniza.sk, bibiana.poliakova@fpedas.uniza.sk

LEAN DISTRIBUTION FRAMEWORK

SYSTEMY LOGISTYKI ODCHUDZONEJ – LEAN DISTRIBUTION

Summary
This article is written about transport modes, which is possible to combine for environmentally effect of transport. Then also must be written about Lean production, lean distribution and its benefits. This article is contains percentage of road, rail, air and water transport today using in the world.

Streszczenie
W artykule przedstawiono zasady logistyki odchudzonej (lean logistics). Scharakteryzowano pięć elementów tworzących łańcuch odchudzony logistyczny. Omówiono podstawowe korzyści z zastosowania koncepcji logistyki odchudzonej dla wszystkich uczestników łańcucha dostaw.

Keywords: lean production, lean distribution, benefits, results, costs of distribution

Słowa kluczowe: system odchudzony, lean management, logistyka odchudzona, koszty dystrybucji

1.1. Transport modes

Various options for moving products from one place to another are called transportation mode. Road, rail, air, water, and pipelines are considered the five basic modes of transportation by most sources. However, all transport modes may not be applicable or feasible options for all markets and products.

Road transport known as highway, truck, and motor carriage-steadily increased its share of transportation. Road transport became the dominant form of freight transport in the United States, replacing rail carriage and it now accounts for 39,8% of total cargo ton-miles, which is more than 68% of actual tonnage.

Rail carriage accounts for 37,1% of total freight ton-miles (more than 14% of actual tonnage) in the United States, which places railroads after motor carriers as the second dominant mode of transportation. However, in some countries such as the People′s Republic of China, the countries of the former Yugoslavia, and Austria, rail remains the dominant transportation mode. [4],[5]

Air carriers transport only around 0,1% of ton-mile traffic in the United States. Although airfreight offers the shortest time in transit (especially over long distances of any transport mode, most shippers consider air transport as a premium emergency service because of its higher costs. However, the high cost of air transport may be traded off with inventory and warehousing reductions or justified in some periods, and in an emergency. [5], [6], [7], [8], [9]

Water carriage-as the oldest mode of transportation-accounts for 5% of total freight ton-miles (around 3,3% of actual tonnage) in the United States. [5]

The Lean Distribution approach is shown in Figure 1. The five elements of the framework form the solution to a Lean transformation. The top and most critical element is customer lead times, order parameters, and service levels for specific customers, groups of customers, and/or products. All aspects on the Lean approach must be focused on these policies, which are frequently not formalized or well communicated. On the bottom, operational capabilities are the foundation for the approach to ensure the Lean processes can be successfully executed. Operational capabilities can be defined to drive a Lean Distribution approach, even if the current operations are not Lean manufacturing enabled, but the benefits will not be as substantial. [1]

These five elements of the Lean distribution framework contain the paradigm-shifting enablers necessary to break the forecasting barrier to customer service and profits. These eight enablers as a formal service policies, support for Pull, isolate variability, Linkage for Pull, Lead time, Variability, Lot sizes, and Cost trade-offs come from lean manufacturing and supply chain practices as tailored to the distribution environment [1]:

- *Formal service policies.* All organizations have some established “norms” and guidelines for customer service, but few examine and formalize policies to

optimize the entire supply chain. The formal policies required for Lean Distribution revolve around articulated customer needs and key internal capabilities.

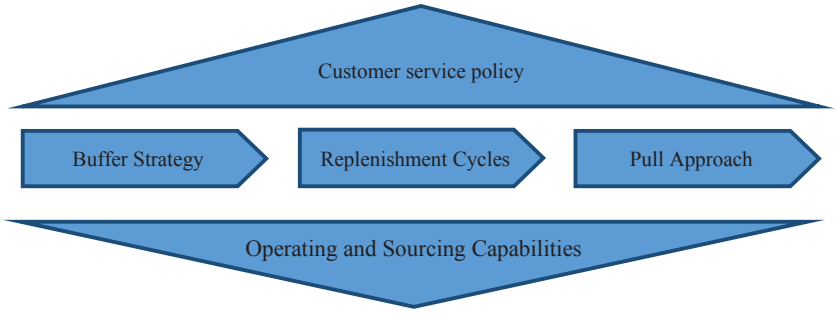


Fig. 1 The Lean Distribution framework [1]

- *Support for Pull.* Customers seek dependable service and generally are willing to allow suppliers more latitude and responsibility to deliver. Support for Pull signifies that the customer recognizes the advantages and allows requirements to flow to the supplier without undue modification or hedging.
- *Isolate variability.* Variability exists in all environments and requires at least some buffer to isolate both customers and internal operations from daily gyrations in forecasts and orders. The trick is to have buffers in the most advantageous places rather than in many or all places customer demand occurs. Strategically placing and managing buffers enables Operations and Sourcing to hit more stationary targets rather than the ever-changing and moving target of a forecast-based plan.
- *Cost trade-offs.* Assess and decide cost trade-offs on a structural level rather than obsess on cost trade-offs for specific transactions every day. It may seem counterintuitive to increase profit by cost optimization of the distribution paths rather than individual replenishment shipments. This more structural approach addresses the variability that is a major barrier to most forecast-and order-driven cost reduction.
- *Linkage for Pull.* Making the links between customer usage or consumption and distribution replenishment processes is the tactical connection required to synchronize the supply chain to consistently meet customer requirements. Pull is more than a Kanban or an "ordering" signal; Pull is the philosophy for replenishment and customer service excellence.
- *Reduced lead times.* Lead times are generally too long. Lead times for internal operations and from suppliers include a high level of safety time to accom-

modate unforeseen events. Lean helps reduce lead times, improving flexibility and responsiveness. Short lead times enable many wonderful cost and service improvements in distribution, particularly when paired with Pull.

- *Reduced variability.* Despite variability existing in all processes, few organizations focus on quantifying and reducing variation in the supply chain. The typical focus is product quality. The first step is to quantify the current variation in order to operate distribution processes based on the limits of current capabilities. For example, distribution center replenishment times may vary, causing Planning to use a "high-end" time for all planned shipments "just to be safe". This longer lead time results in excess inventories and a realization that not all orders must be shipped on the day planned, requiring, expediting and overrides to ensure priority orders are shipped when needed.
- *Reduced lot sizes*. The quantity produced or sourced at one time or lot size has a direct relationship to flexibility and total costs. Larger lot sizes appear to lower costs in sourcing or production but can increase cost and reduce service across the rest of the supply chain. Lean manufacturing practices help reduce lot sizes while eliminating waste, thereby enabling both low product and supply chain costs.

The eight enablers combine to form a cohesive system to improve distribution costs, asset utilization, and customer service. These enablers must be linked and implemented to leverage the overall approach and not as a series of disjoint cost reduction initiatives. The approach is tied together by the Lean waste reduction philosophy and the transition away from forecast to serve daily customer needs. The end result is a new paradigm to view profitability and customer relationships. As lead time, variation, and lot sized decrease, profit approaches gross margin (excluding other general expense items). When these factors decrease sufficiently, there is a net addition to profit above gross margin from the effect of negative working capital, an example being Dell Computer (where accounts payable is three times the amount of inventory and accounts receivable combined.)

This total cost paradigm is more than just adding up all the usual department budgets into a total. It is a view of the drivers of cost rather than the results. Results are the freight, labor, inventory, overhead, and other costs included in financial reports and departmental budgets. These results are driven by other factors, such as a lead times and lot sizes. It is these drivers that require the cost-reducing focus, budgets and financial reports are the measures of results. [1],[2]

2. The benefits of Lean distribution

Lean distribution bears immediate fruit – both tangible and intangible. The benefits are straight for-ward and significant:

- A 10 to 50 percent improvement in labor productivity
- Smoother and more accelerated product and work flows
- Happier and more productive associates, which improves retention
- More capable management team
- Greater facility throughput and capacity
- Avoidance of major capital outlays – i.e., not having to build a new DC to handle incre-mental growth [10].

As in the case of the major big box retailer cited at the outset of this article, the potential savings derived from applying lean to distribution operations run in to the millions of dollars.

The lean distribution approach provides an operational foundation for service excellence and low total costs. A combination of service and cost performance is what differentiates how a Lean approach both simplifies the business and delivers results. Service and cost are typically considered to be conflicting objectives where trade-offs must be made, but Lean focuses efforts on changing the dynamics of this trade-off by reducing cycle time, improving reliability, and increasing flexibility. These changes deliver benefits in customer service, total costs, and asset utilization.

Customer services benefits accrue from improvements in service policies and value provided to customers with Pull. As service policies are formalized and segmented with the Lean distribution approach, benefits relate to service delivery for customers grouped into segments with various levels of service. By formalizing service policies, improvements result from:

- Providing differentiated levels of customer service,
- Improving execution of service delivery,
- Examining and segmenting customers for price/value,
- Directly linking (Pull) with customer usage to improve the customer's material flows.

Within Distribution operations, Lean improves the flow of products to mitigate fluctuations in customer demand. With buffer strategy developments, Lean takes inventory placements and management from a just in case approach to a proactive setting of inventory to ensure service excellence. The buffer "shock absorber" is consolidated, lowering total inventories and protecting against variation. This reduces the pressure to maintain inventory availability so Distribution can focus on executing replenishment to customers and the critical elements of cost.

Benefits also accrue across "downstream" operations both internally and with suppliers. As Pull with customers improves the connection with actual demand, the disruption in planning and customer orders can be decreased, eliminating the snowball and bullwhip effects created as small changes in forecasts and customer orders are magnified. Operations (and supplier) schedules are stabilized by insu-

lating the market variation with the strategic buffers. As schedules stabilize, Lean practices further improve flexibility, reliability, and costs without the distractions of daily disruptions to meet spikes in demand.

The planning function also benefits from a Lean Distribution transformation. Often the attention and effort required to update forecasts and revise plans lessens as forecasts and planning become more strategic and less tactical. Planners spend less time managing the plan ad internal replenishment orders and more time determining and supporting longer-range decision making. The total time required for the Planners may not be less, but certainly more value added time.

Lean benefits may seem counterintuitive to traditional measures; however, they are quantifiable. As lean takes hold, benefits can be quantified along key measures of performance and cost:

- Inventory reductions across the entire supply chain as buffers become strategically located and managed,
- Distribution cost reductions as customer and distribution center replenishments become distribution pipes designed using delivered costs,
- Customer service readily measured against segmented policies and expectations,
- Operations schedule stability as measured by the disruption to schedule within the window of time a specific operation.

These and other benefits can be linked back to key operational parameters and improvements. Cycle time is an example parameter that improves with Lean and can be related to inventory and service costs. As cycle time is reduced, the improvements in inventory and costs are readily apparent. These relationships make lean benefits more transparent and linked with day-to-day operational improvement efforts and measures.[1]

This paper presents results of work supported by the Slovak Scientific Grant Agency of the Slovak republic under the project No. VEGA 1/0331/14.

Bibliography

[1] ZYLSTRA, KIRK, D.: Lean distribution, Published by John Wiley &sons, Inc. Hoboken, New Jersey.2006, ISBN 0-471-74075-6.

[2] STOKŁOSA J. , MARCINIAK A. Modelowanie ryzyka w transporcie. Logistyka nr 3/2014. ISSN 1231-5478.

[3] FAHARANI, ZANJIRANIA REZA, RAZPOUR SHABNAM, KARDAR LALEH: Logistics operations and management, Elsevier insights 2011, first edition, London UK, ISBN 978-0-12-385202-1.

[4] C.L. LEE, M.-H. SHU, Fault-tree analysis of intuitionistic fuzzy sets for liquefied natural gas terminal emergency shut-down system, in: Third Inter-

national Conference on International Information Hiding and Multimedia Signal Processing, IEEE Computer Society, 2007.

[5] UNCTAD, Review of Maritime Transport, United Nations, New York and Geneva, 2003.

[6] J. MUNOZ, N. JIMENEZ-Redondo, J. PEREZ-RUITZ, J. BARQUIN, Natural gas network modelling for power systems reliability studies, in IEEE Bologna PowerTech Conference, Bologna, Italy, 2003.

[7] C. UNSIHUAY-VILA, J. W. MARAGNON-LIMA, A.C.Z. de Souza, I.J. PEREZ-ARRIAGA, A model to long-term, miltiarea, multistage, and integrated expansion planning of electricity and natural gas systems, IEEE Trans, Power Syst.25(2) (2009) 1154-1168 (May 2010).

[8] IEA, South American Gas-Daring to Tap the Bounty, IEA Press, France, 2003.

[9] L.A. Barroso, H. Rudnick, S. Mocarquer, R. Kelman, B. Bazerra, LNG in South America, tha markets, the prices and the security of supply, in: IEEE Power and Energy Society General, Meeting-Conversion and Delivery of Electrical Energy in the 21st Century, Pittsburg, PA, 2008.

[10] http://www.mhi.org/media/members/15155/129660239771047549.pdf.

[11] http://content.aktion.com/file/Lean%20Distribution%20Whitepaper.pdf.