HAKIN9 9/2010

ŚCIĄGNIJ ZA DARMO!

HAKIN9

JAK SIĘ OBRONIĆ HARD CORE IT SECURITY MAGAZINE

nr 9/2010 (64)

TWÓJ BACKUP JEST BEZ SENSU!

PISANIE WIN32 SHELLCODE W KOMPLILATORZE C
NAUCZ SIĘ JAK NAPISAĆ SHELLCODE

SZKODLIWE OPROGRAMOWANIE DLA GIER
W JAKI SPOSÓB CYBERPRZESTĘPCY MOGĄ WEJŚĆ W POSIADANIE WIRTUALNYCH PRZEDMIOTÓW

MONITOROWANIE SIECI
JAK MONITOROWAĆ PRACĘ ELEMENTÓW SIECI UŻYWAJĄC NAGIOS

NEURONY W KOMPUTERZE
SIECI NEURONOWE JAKO SYSTEM ZWIĘKSZAJĄCY BEZPIECZEŃSTWO SIECI KOMPUTEROWEJ

6-7 września 2010 | Warszawa
http://gigacon.org/bin_warszawa

6 października 2010 | Wrocław
http://gigacon.org/bin_wroclaw

Bezpieczeństwo i Niezawodność Systemów Informatycznych

Kontakt: Kamila Tarłowska
SW Konferencje Sp. z o.o. Sp. k.
tel.: 022 427 36 47
e-mail: kamila.tarlowska@software.com.pl

BIN GigaCon™

BEZPIECZEŃSTWO:

B1 Zarządzanie bezpieczeństwem informacji
B2 Systemy audytu bezpieczeństwa (skanery), systemy kontroli włamań
B3 Oprogramowanie i urządzenia szyfrujące
B4 Zarządzanie tożsamością, kontrola dostępu
B5 Systemy firewall i VPN
B6 Oprogramowanie i systemy antywirusowe

NIEZAWODNOŚĆ:

N1 Serwery usług sieciowych
N2 Systemy klastrowe
N3 Bezpieczeństwo i niezawodność systemów baz danych
N4 Fizyczne zabezpieczenie infrastruktury systemów o znaczeniu krytycznym
N5 Pamięci masowe i systemy archiwizacji danych
N6 Ciągłość działania, disaster recovery

WSTĘP **BEZPŁATNY**

www.isecman.org

ISecMan
Information Security Management

CYKL SZKOLEŃ Z OBSZARU ZARZĄDZANIA BEZPIECZEŃSTWEM IT

Mając na uwadze ciągły rozwój systemów bezpieczeństwa informacji za główny cel postawiliśmy sobie dostarczenie Państwu najnowszych i najlepszych praktyk z obszaru zarządzania bezpieczeństwem IT. Robimy to w sposób jasny, przystępny, opierając się o najnowsze normy typu PN-ISO/IEC 27001:2007, PN-ISO/IEC 17799:2007. Wieloletnie doświadczenie i prezentowany poziom merytoryczny stawia nas bezwarunkowo jako liderów na rynku polskim w obszarach szkoleń i certyfikacji bezpieczeństwa.

- **BCM - Zarządzanie Ryzykiem w Organizacji**
  22 lipca 2010
- **Techniki Przełamywania Zabezpieczeń Systemów**
  24–25 sierpnia 2010
- **Szkolenie przygotowujące do egzaminu CISSP (Certified Information Systems Security Professional)**
  18-20 sierpnia, 26-27 sierpnia 2010 (5 dni)
- **Projektowanie Polityki Bezpieczeństwa**
  22-23 września 2010
- **Audyt Bezpieczeństwa Teleinformatycznego zgodny ze standardem PN-ISO/IEC 17799:2007**
  24 września 2010
- **Zarządzanie Ryzykiem w Organizacji zgodnie z normami PN-ISO/IEC 27001 i ISO/IEC 27005**
  październik 2010

Kontakt i szczegóły:
Edyta Szewc
tel. +48 22 427 36 70 | fax +48 22 244 24 59
edyta.szewc@software.com.pl

Information Security ISecMan www.isecman.org Management

Przyjdź i zdobądź:
Generalny Certyfikat Inżyniera Bezpieczeństwa IsecMan

SPIS TREŚCI

NARZEDZIA

ATAK

OBRONA

PRAKTYKA

Miesięcznik **hakin9** (12 numerów w roku)
jest wydawany przez Software Press Sp. z o.o. SK

Prezes wydawnictwa: Paweł Marciniak

Dyrektor wydawniczy: Ewa Łozowicka

Redaktor naczelny:
Katarzyna Dębek *katarzyna.debek@software.com.pl*

Skład i łamanie:
Tomasz Kostro *www.studiopoligraficzne.com*

Kierownik produkcji:
Andrzej Kuca *andrzej.kuca@software.com.pl*

Adres korespondencyjny:
Software Press Sp. z o.o. SK,
ul. Bokserska 1, 02-682 Warszawa, Polska
tel. +48 22 427 36 91, fax +48 22 224 24 59
www.sdjournal.org cooperation@software.com.pl

Dział reklamy: *adv@software.com.pl*

Redakcja dokłada wszelkich starań, by publikowane w piśmie i na towarzyszących mu nośnikach informacje i programy były poprawne, jednakże nie bierze odpowiedzialności za efekty wykorzystania ich; nie gwarantuje także poprawnego działania programów shareware, freeware i public domain.

Wszystkie znaki firmowe zawarte w piśmie są własności odpowiednich firm.
Zostały użyte wyłącznie w celach informacyjnych.

Osoby zainteresowane współpracą prosimy o kontakt:
cooperation@software.com.pl

Słuchawki iBOX HPI 99MV

Producent
iBOX
Typ
Słuchawki
Strona producenta
www.ibox.pl
Recenzent
Artur Żarski

OCENA ★★★★★

Jednym z podstawowych elementów wyposażenia każdego użytkownika komputerów są słuchawki. Modeli słuchawek jest bardzo dużo – nauszne, douszne, przewodowe, bezprzewodowe. Tym razem testujemy słuchawki nauszne firmy iBOX model HPI 99MV.

Słuchawki te wg specyfikacji komunikują się z urządzeniami za pomocą kabla, ale na szczęście jest on całkiem długi – 2,4 m zupełnie wystarczy w większości sytuacji. Złącze typu mini-Jack pozwala na użycie ich prawie we wszystkich dostępnych urządzeniach – komputerach, wieżach HiFi czy nawet w niektórych telefonach komórkowych. Słuchawki te maja wbudowany mikrofon, dzięki czemu automatycznie rozwiązują problem komunikacji VoIP. Dodatkowo na przewodzie zainstalowany jest pilot, za pomocą którego możemy sterować głośnością. Jego podstawowe dane (wg producenta) są następujące: pasmo przenoszenia wynosi 20Hz-20kHz, czułość mikrofonu -58+-2 dB a skuteczność słuchawek to 105 +- 1dB. Takie parametry wystarczają zupełnie do tego, aby przyzwoicie słuchać muzyki, grać w gry komputerowe czy prowadzić konwersację za pomocą komunikatorów internetowych (nie są to jakieś super wyniki, ale za tą cenę jest naprawdę OK).

Pierwsze wrażenie jest pozytywne – ładnie zapakowane w plastikowe, przeźroczyste pudełko. Drugie wrażanie trochę gorsze – bez ostrego noża lub nożyczek otwarcie opakowania jest niemal niemożliwe – natomiast nie testujemy opakowania a słuchawki. Wracając do słuchawek – pierwsze wrażenia pozytywne – dobra cena (ok 45 PLN), przyzwoity wygląd, jakość wykonania również dobra.

Pierwszy test – zakładamy słuchawki na głowę – duże i miękkie nauszniki, dzięki czemu nawet przy dłuższym słuchaniu nie uwierają. Nie mniej jednak wydają się na pierwszy rzut oka nie wygodne – nacisk na uszy czy ułożenie mogą sprawiać wrażenie trudnych do użycia. Tutaj mała uwaga dla osób z większą głową mogą sprawiać wrażenia jakby się miały rozlecieć.

Słuchawki testowane były około tygodnia i w tym czasie działały bardzo dobrze – nie było żadnych problemów. Nie wiadomo jak zachowują się po dłuższym użyciu, czy głośniki brzęczą czy mikrofon nie zniekształca dźwięków. Innym testem, który zakończył się sukcesem było starcie pomiędzy dwuletnim dzieckiem a słuchawkami – po kilkunastu minutach zabawy, naciągania mikrofonu oraz kabla nie udało się ich zepsuć – dobrze wróży na długie użycie

Reasumując – słuchawki firmy iBox to dobra alternatywa dla droższych modeli znanych firm. Nie wiemy ja będą działać po czasie ciągłego użycia ale jakość do ceny jest dobry.

iBOX Wee – niska cena to nie wszystko

Importer
Impet Computers Sp. z o.o.
Typ
Odtwarzacz MP4
Strona producenta
www.ibox.pl
Recenzent
Patryk Krawaczyński

OCENA ★★★★★

iBOX Wee jest małym odtwarzaczem mp4 (wymiary: 80 mm x 41 mm x 8 mm) z 1,72" wyświetlaczem CSTN (65 tys. kolorów). Model otrzymany do testów został wyposażony w wbudowaną pamięć o pojemności 2 GB. Dotarł zapakowany w czarno – szare pudełko, na którym zostały zawarte niektóre z jego funkcjonalności: akumulator, wbudowane radio, dyktafon, możliwość odczytu eBooków oraz obsługa technologii Microsoft PlayFX, która ma za zadanie zapewnienie lepszej jakości odtwarzanego dźwięku. Po dokładnym przestudiowaniu opakowania, oprócz specyfikacji technicznej w pięciu językach (rosyjskim, słowackim, niemieckim, angielskim i polskim) można znaleźć dość często spotykany napis „Made in China". Nie jest to żadnym uchybieniem dla sprzętu, ale jest to jedynym punktem zaczepienia jeśli chodzi o informacje o producencie. Importerem na rynek polski jest firma Impet Computers sp. z o.o. z siedzibą w Warszawie. Próżno też szukać informacji o producencie na stronie internetowej (www.ibox.eu), która mimo europejskiej domeny kieruje na polską wersję językową (również przy zmianie języka przeglądarki) dość skromnej strony. Fakt ten nie budzi zbyt dużego zaufania u potencjalnego klienta szanującego sobie wiarygodność elektronicznych zakupów i dalszego serwisowania sprzętu – szczególnie, że na odtwarzacz wystawiona jest bezpośrednia 24 miesięczna gwarancja. Ponadto na bocznej stronie opakowania, która poświęcona jest angielskiemu opisowi w oczy rzuca się pojedynczy polski napis o pojemności pamięci, co dość dziwnie komponuje się zresztą treści. W opakowaniu zostały zawarte: odtwarzacz, słuchawki, płyta mini-CD, gwarancja, instrukcja oraz przewód mini-USB.

Krótka instrukcja obsługi została napisana w dwóch językach – polskim i angielskim. Po lekturze pierwszych stron rażąco rzucają się w oczy literówki oraz luki w tłumaczeniach (np. antenna). Sam odtwarzacz prezentuje się ładnie. Po usunięciu foli ochronnej lustrzana obudowa nabyła syndrom znany z laptopów pewnej firmy, które są idealną powierzchnią do zdejmowania odcisków palców. Nie trafionym pomysłem jest także mała naklejka gwarancyjna obok gniazda zasilania, którą można bardzo szybko zniszczyć ze względu na mobilne przeznaczenie urządzenia. Z funkcjonalnością urządzenia jest już trochę lepiej. Poruszanie się po menu jest dość intuicyjne i łatwe do zapamiętania, chociaż i ono nie zostało w pełni przetłumaczone (np. Language). Instrukcja milczy na temat blokady klawiatury – informacje te można dopiero uzyskać poprzez lekturę Internetu (Play+M). Maksymalny czas pracy z baterią (odtwarzanie muzyki) zakładany przez producenta to 11 godzin. Po pełnym załadowaniu urządzenia za pomocą portu USB z komputera (+/- 45 minut); ustawienia: wyłączania wyświetlacza po 5 sekundach, jasności 1 na 5, oszczędnym trybie wyświetlania oraz głośności 16 na 32 pozwoliły na uzyskanie wyniku około 6 godzin odtwarzania plików muzycznych. Jakość odtwarzania z załączonymi słuchawkami dousznymi oraz włączonym (z siedmiu trybów) MS Play FX Pure Bass jest bardzo przyzwoita. Jeśli znudzi znam lista odtwarzania możemy przełączyć się na radio z ręcznym lub automatycznym trybem wyszukiwania stacji. Podobnie jak z muzyką Wee bardzo dobrze radzi sobie z wyświetlaniem obrazów w formacie jpeg lub bmp. Gorzej jest już z plikami video. Wyświetlacz jest na tyle mały, że sprawia dyskomfort przy oglądaniu krótkiego klipu (pasek nawigacyjny oraz postępu nie ukrywają się w trybie pełnotekstowym). W dodatku dołączone oprogramowanie do konwersji plików video wiesza się przy większych plikach.

Niska cena tego odtwarzacza powinna rekompensować niektóre jego wady. Jednak podczas jego użytkowania pozostaje przykre wrażenie, że większość rzeczy była robiona po prostu po bardzo niskich kosztach lub z mniejszym profesjonalizmem niż to ma miejsce w przypadku innych odtwarzaczy tej klasy. Wiele rozwiązań jest na dobrym poziomie, ale szczegóły sprawiają, że użytkownikowi może przejść przez głowę myśl o złej decyzji zakupu.

Pisanie WIN32 shellcode w kompilatorze C

Didier Stevens

Napisanie shellcodu jest trudne. Dlatego opracowane metody w tym artykule pozwolą wygenerować WIN32 shellcode z kompilatora C. Aby w pełni korzystać z treści tego artykułu, należy mieć doświadczenie w pisaniu programów w WIN32 C/C++ i WIN32shellcode, i rozumieć różnice między tymi podejściami.

Dowiesz się:

- Jak napisać shellcode w czystym C

Powinieneś wiedzieć:

- Podstawowe wiadomości o tym co jest shellcode, i umiejętność programowania w C

Dla celów niniejszego artykułu, określam shellcode jako niezależny od pozycji kod urządzenia. Normalnie shellcode jest napisany w asemblerze, a deweloper zwraca uwagę na tworzenie kodu niezależnego od pozycji. Innymi słowy, że shellcode zostanie wykonany poprawnie bez względu na jego adres w pamięci.

Kompilator jaki używam to Visual C++ 2008 Express. Jest on darmowy i obsługuje asemblację inline. Shellcode generowany przy użyciu tej metody jest dynamiczny: nie używa stałej adresacji API, która ogranicza shellcode do poszczególnych wersji systemu Windows. Metoda wykorzystuje kod Dave Aitel do wyszukiwania adresów niezbędnych funkcji API opisanych w książce "Shellcoder's Handbook".

Możliwość debugowania shellcode wewnątrz Visual C + + w okresie rozwoju jest ważnym wymogiem dla mnie. Rozwijanie shellcode nie jest łatwe, możemy korzystać z wszystkich pomocy jakie możemy uzyskać, wizualny debugger jest z pewnością mile widziane. Kilka projektowych decyzji było podjętych z powodu tego wymogu.

Metoda nie próbuje wygenerować kompaktowych shellcode. Jeśli rozmiar jest problemem, zacznij od małych, ręcznie napisanych i zoptymalizowanych shellcodów i wywołaj generator shellcode przez kompilator C na późniejszym etapie.

Kod źródłowy, dla którego kompilator C wygeneruje shellcode znajduje się między funkcjami `ShellCodeStart` i `main` (linie 45-213). Funkcja `main` i następujące po niej funkcje nie są częścią generowanego shellcode, dostarczają one wsparcia dla testowania i debugownia shellcode i automatycznie wyodrębniają shellcode z generowanego pliku PE (plik *.exe*).

Jak nazwy wskazują, `ShellCodeStart` (linia 45) jest początkiem naszego shellcode. Wywołuje funkcję `ShellCodeMain`, a następnie wraca. To wszystko, co robi. Ponieważ ten prosty kod jest rzeczywiście napisany w assemblerze, a nie w C, musimy o tym poinformować kompilator. Konstrukcja `__asm` jest tym, czego potrzebujemy do osiągnięcia tego celu:

```
__asm
{
    call ShellCodeMain
    ret
}
```

Kiedy kompilator C emituje kod maszynowy do normalnego funkcjonowania, to dodaje instrukcje instalacji i podział ramki stosu (wewnętrznej struktury danych C do przechowywania argumentów i zmiennych automatycznych na stosie). Kod ten nazywa się odpowiednio prolog i epilog. My nie potrzebujemy ramki stosu funkcji `ShellCodeMain`. Aby poinformować kompilator C że ma pominąć epilog i prolog, dekorujemy `ShellCodeMain` atrybutem `__declspec` (naked).

Cel funkcji `ShellCodeStart` jest dwojaki: zrobić pierwszy bajt shellcode punktem wejścia, a także zapewnienie adresu startowego, aby wyodrębnić shellcode z pliku PE.

`ShellCodeMain` (linia 196) jest główną funkcją naszego shellcode. Zapewnia pamięć do przechowywania danych, zwraca kod do wyszukiwania adresów funkcji API które potrzebujemy i wykonuje nasz kod rdzenia.

Następujący wiersz (linia 198) rezerwuje pamięć na stosie dla naszych danych:

```
SHELL_CODE_CONTEXT scc;
```

`SHELL_CODE_CONTEXT` jest strukturą zawierającą wszystkie dane potrzebne w całym naszym shellcode, jak adresy funkcji Win32 API. Jest przekazywana do wszystkich funkcji (które jej potrzebują) przez wskaźnik. Przechowujemy strukture na stosie (jako automatyczna zmienna), aby nasz shellcode był niezależny od stanowiska. Deklarowanie zmiennej dla struktury jako statycznej poleci kompilatorowi C, aby przechowywał zmienną w segmencie danych, co nie jest niezależne od stanowiska, a tym samym nie nadaje się do naszego shellcode. Dla naszego shellcode `MessageBox`, struktura zawiera takich członków (linia 29):

```
struct SHELL_CODE_CONTEXT
{
    TD_LoadLibraryA FP_LoadLibraryA;
    TD_GetProcAddressA FP_GetProcAddressA;

    char szEmptyString[1];

    HMODULE hmUSER32;
    TD_MessageBoxA FP_MessageBoxA;
};
```

`FP_LoadLibraryA` i `FP_GetProcAddressA` są zmiennymi (a dokładniej zmiennymi wskaźnika funkcji), aby zapisać adres kernel32 wexportowania `LoadLibraryA` i `GetProcAddressA`. Pamiętaj, że piszemy dynamiczny shellcode, nie używamy sztywnych adresów API.

`szEmptyString` to zmienna do przechowywania pustych łańcuchów. "" - jest łańcuchem pustym, to jest różne od `NULL`. Nie możemy użyć ciągów bezpośrednio w naszym shellcode, ponieważ kompilator C będzie przechowywał je w segmencie danych. Obejściem jakiego ja używam jest "zbudowanie" łańcucha znaków wraz z kodem i zapisanie go w zmiennych na stosie. W ten sposób ciągi są częścią naszego shellcode.

Jako że pusty ciąg jest ciągiem potrzebnym w kilku funkcjach, postanowiłem zapisać pusty ciągu w strukturze shellcode.

`hmUSER32` to zmienna do przechowywania adresu załadowanej *user32 dll*. Zmienna `FP_MessageBoxA` jest wskaźnikiem funkcji `MessageBoxA`.

Po utworzeniu zmiennych na stosie, musimy je zainicjować. Funkcja ShellCodeInit (linia 57) realizuje kod "The Shellcoder's Handbook" aby dynamicznie wyszukiwaniać adresy `kernel32` dla `LoadLibraryA` i `GetProcAddressA`. Kod wykonuje to przez przeszukiwanie procesu struktury danych, które zawierają listę załadowanych modułów i eksportowanych funkcji. W celu uniknięcia stosowania łańcuchów znaków dla nazw funkcji, wykorzystuje hashe (linie od 19 do 21):

```
#define KERNEL32_HASH 0x000d4e88
#define KERNEL32_LOADLIBRARYA_HASH 0x000d5786
#define KERNEL32_GETPROCADDRESSA_HASH 0x00348bfa
```

Te dwie funkcje API są wszystkim czego nam potrzeba do wyszukiwania innych funkcji API. Kod stosowany do tej pory to szablon który będzie ponowne wykorzystany we wszystkich innych shellcode opracowanych tą metodą.

Inicjowania pustego łańcucha jest łatwe (linia 202):

```
scc.szEmptyString[0] = '\0';
```

Teraz musimy znaleźć adres `MessageBoxA` który za pomocą `LoadLibraryA` i `GetProcAddressA`. `MessageBoxA` jest eksportowany przez user32.dll. Musimy odwołać się do tego modułu, a może załadować go, jeśli nie jest już załadowany wewnątrz procesu, w którym nasz shellcode zostanie wykonany. Robimy to przy pomocy `LoadLibraryA` z argumentem "user32". "user32" musi być łańcuchem znaków, ale pamiętaj, że nie możemy zapisać "user32" literalnie w naszym kodzie C, ponieważ kompilator C będzie przechowywał ciąg "user32" w miejscu niedostępnym dla naszego shellcode. Trikiem używanym w celu uniknięcia tego jest zainicjowanie łańcucha przy pomocy tablicy znaków (linia 205):

```
char szuser32[] = {'u', 's', 'e', 'r', '3', '2', '\0'};
```

Niniejszy zapis wymusza aby kompilator wyemitował kod w którym każdy znak łańcucha „user32” (wraz z kończoncym 0) będzie przechowywany na stosie w zmiennej szuser32, tak by przy starcie był dynamicznie budowany napis. Zaletą tego sposobu jest to, że słowa są nadal czytelne w naszym kodzie źródłowym. Wadą jest to, że shellcode wygenerowany w ten sposób nie jest kompaktowy: emitowanie instrukcji kodu maszynowego dla każdego pojedynczego znaku zajmuje miejsce na dysku. Zwróć uwagę aby zawsze kończyć każdą tablicę znaków przy pomocy znaku null ('\ 0');

Tworzenie napisu "MessageBoxA" odbywa się w ten sam sposób:

```
char szMessageBoxA[] = {'M', 'e', 's', 's', 'a', 'g',
                        'e', 'B', 'o', 'x', 'A', '\0'};
```

Listing 1a. *Pełen kod źródłowy shellcode do wyświetlenia okna komunikatu.*

```
/*
   ShellCodeTemplate v0.0.1: MessageBox demo
   Source code put in public domain by Didier Stevens, no Copyright
   Except for the code in ShellCodeInit, which is released under the GNU PUBLIC LICENSE v2.0
   http://didierstevens.com
   Use at your own risk

   Shortcommings, or todo's ;-)
      - find fix for function allignment

   History:
      2008/10/24: start
      2010/02/02: cleanup
*/

#include <windows.h>
#include <stdio.h>

#define KERNEL32_HASH 0x000d4e88
#define KERNEL32_LOADLIBRARYA_HASH 0x000d5786
#define KERNEL32_GETPROCADDRESSA_HASH 0x00348bfa

typedef HMODULE (WINAPI *TD_LoadLibraryA)(LPCTSTR lpFileName);
typedef FARPROC (WINAPI *TD_GetProcAddressA)(HMODULE hModule, LPCTSTR lpProcName);

// Add your API function pointer definitions here:
typedef int (WINAPI *TD_MessageBoxA)(HWND hWnd, LPCTSTR lpText, LPCTSTR lpCaption, UINT uType);

struct SHELL_CODE_CONTEXT
{
   TD_LoadLibraryA FP_LoadLibraryA;
   TD_GetProcAddressA FP_GetProcAddressA;

   char szEmptyString[1];

   // Add your module handles and API function pointer members here:
   HMODULE hmUSER32;
   TD_MessageBoxA FP_MessageBoxA;
};

void ShellCodeMain(void);
int WriteShellCode(LPCTSTR, PBYTE, size_t);
void *ShellCodeData(void);

void __declspec(naked) ShellCodeStart(void)
{
   __asm
   {
      call ShellCodeMain
      ret
   }
}

#pragma warning(push)
#pragma warning(disable:4731)

void ShellCodeInit(TD_LoadLibraryA *pFP_LoadLibraryA, TD_GetProcAddressA *pFP_GetProcAddressA)
{
   TD_LoadLibraryA FP_LoadLibraryA;
   TD_GetProcAddressA FP_GetProcAddressA;

   // Shellcode functions to lookup API functions, based on
   // The Shellcoder's Handbook http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0764544683.html'
   // Released under the GNU PUBLIC LICENSE v2.0

   __asm
   {
      push KERNEL32_LOADLIBRARYA_HASH
      push KERNEL32_HASH
      call getfuncaddress
      mov FP_LoadLibraryA, eax

      push KERNEL32_GETPROCADDRESSA_HASH
      push KERNEL32_HASH
      call getfuncaddress
      mov FP_GetProcAddressA, eax

      jmp totheend

   getfuncaddress:
      push ebp
      mov ebp, esp
      push ebx
      push esi
      push edi
      push ecx
      push fs:[0x30]
      pop eax
      mov eax, [eax+0x0c]
      mov ecx, [eax+0x0c]
```

Listing 1b. *Pełen kod źródłowy shellcode do wyświetlenia okna komunikatu.*

```
091   nextinlist:
092      mov edx, [ecx]
093      mov eax, [ecx+0x30]
094      push 0x02
095      mov edi, [ebp+0x08]
096      push edi
097      push eax
098      call hashit
099      test eax, eax
100      jz foundmodule
101      mov ecx, edx
102      jmp nextinlist
103   foundmodule:
104      mov eax, [ecx+0x18]
105      push eax
106      mov ebx, [eax+0x3c]
107      add eax, ebx
108      mov ebx, [eax+0x78]
109      pop eax
110      push eax
111      add ebx, eax
112      mov ecx, [ebx+28]
113      mov edx, [ebx+32]
114      mov ebx, [ebx+36]
115      add ecx, eax
116      add edx, eax
117      add ebx, eax
118   find_procedure:
119      mov esi, [edx]
120      pop eax
121      push eax
122      add esi, eax
123      push 1
124      push [ebp+12]
125      push esi
126      call hashit
127      test eax, eax
128      jz found_procedure
129      add edx, 4
130      add ebx, 2
131      jmp find_procedure
132   found_procedure:
133      pop eax
134      xor edx, edx
135      mov dx, [ebx]
136      shl edx, 2
137      add ecx, edx
138      add eax, [ecx]
139      pop ecx
140      pop edi
141      pop esi
142      pop ebx
143      mov esp, ebp
144      pop ebp
145      ret 0x08
146
147   hashit:
148      push ebp
149      mov ebp, esp
150      push ecx
151      push ebx
152      push edx
153      xor ecx,ecx
154      xor ebx,ebx
155      xor edx,edx
156      mov eax, [ebp+0x08]
157   hashloop:
158      mov dl, [eax]
159      or dl, 0x60
160      add ebx, edx
161      shl ebx, 0x01
162      add eax, [ebp+16]
163      mov cl, [eax]
164      test cl, cl
165      loopnz hashloop
166      xor eax, eax
167      mov ecx, [ebp+12]
168      cmp ebx, ecx
169      jz donehash
170      inc eax
171   donehash:
172      pop edx
173      pop ebx
174      pop ecx
175      mov esp, ebp
176      pop ebp
177      ret 12
178
179   totheend:
180   }
181
182   *pFP_LoadLibraryA = FP_LoadLibraryA;
183   *pFP_GetProcAddressA = FP_GetProcAddressA;
184 }
185
186 #pragma warning(pop)
187
188 // Write your custom code in this function.
189 // Add extra functions as needed.
190 void ShellCodePayload(SHELL_CODE_CONTEXT *pSCC)
191 {
192   char szHello[] = {'H', 'e', 'l', 'l', 'o', '\
                  0'};
193   pSCC->FP_MessageBoxA(NULL, szHello, pSCC-
```

Listing 1c. *Pełen kod źródłowy shellcode do wyświetlenia okna komunikatu.*

```
                >szEmptyString, 0);
194 }
195
196 void ShellCodeMain(void)
197 {
198    SHELL_CODE_CONTEXT scc;
199
200    ShellCodeInit(&(scc.FP_LoadLibraryA), &(scc.FP_
                GetProcAddressA));
201
202    scc.szEmptyString[0] = '\0';
203
204    // Add your own API function initialization code
                here:
205    char szuser32[] = {'u', 's', 'e', 'r', '3', '2',
                '\0'};
206    char szMessageBoxA[] = {'M', 'e', 's', 's',
                'a', 'g', 'e', 'B', 'o', 'x', 'A',
                '\0'};
207    scc.hmUSER32 = scc.FP_LoadLibraryA(szuser32);
208    scc.FP_MessageBoxA = (TD_MessageBoxA)scc.FP_
                GetProcAddressA(scc.hmUSER32,
                szMessageBoxA);
209
210    ShellCodePayload(&scc);
211 }
212
213 int main(int argc, char **argv)
214 {
215    size_t dwSize;
216    char szBinFile[MAX_PATH];
217
218    dwSize = (PBYTE)main - (PBYTE)ShellCodeStart;
219    printf("Shellcode start = %p\n",
                ShellCodeStart);
220    printf("Shellcode size = %08x\n", dwSize);
221    sprintf_s(szBinFile, MAX_PATH, "%s.bin",
                argv[0]);
222    printf("Shellcode file = %s\n", szBinFile);
223    if (0 == WriteShellCode(szBinFile,
                (PBYTE)ShellCodeStart, dwSize))
224       printf("Shellcode file creation successful\
                n");
225    else
226       printf("Shellcode file creation failed\n");
227
228    // Calling ShellCodeMain to debug shellcode
                inside Visual Studio
229    // Remove this call if you don't want to execute
                your shellcode inside Visual Studio
230    ShellCodeMain();
231
232    return 0;
233 }
234
235 // Function to extract and write the shellcode to
                a file
236 int WriteShellCode(LPCTSTR szFileName, PBYTE
                pbShellCode, size_t sShellCodeSize)
237 {
238    FILE *pfBin;
239    size_t sWritten;
240
241    if (S_OK != fopen_s(&pfBin, szFileName, "wb"))
242       return -1;
243    sWritten = fwrite(pbShellCode, sShellCodeSize,
                1, pfBin);
244    fclose(pfBin);
245    if (sWritten != 1)
246       return -2;
247    return 0;
248 }
```

Teraz możemy znaleźć adres modułu user32, w którym musimy odnaleźć adres funkcji API MessageBoxA (linia 207):

```
scc.hmUSER32 = scc.FP_LoadLibraryA(szuser32);
```

Następnie wyszukujemy adres MessageBoxA (linia 208):

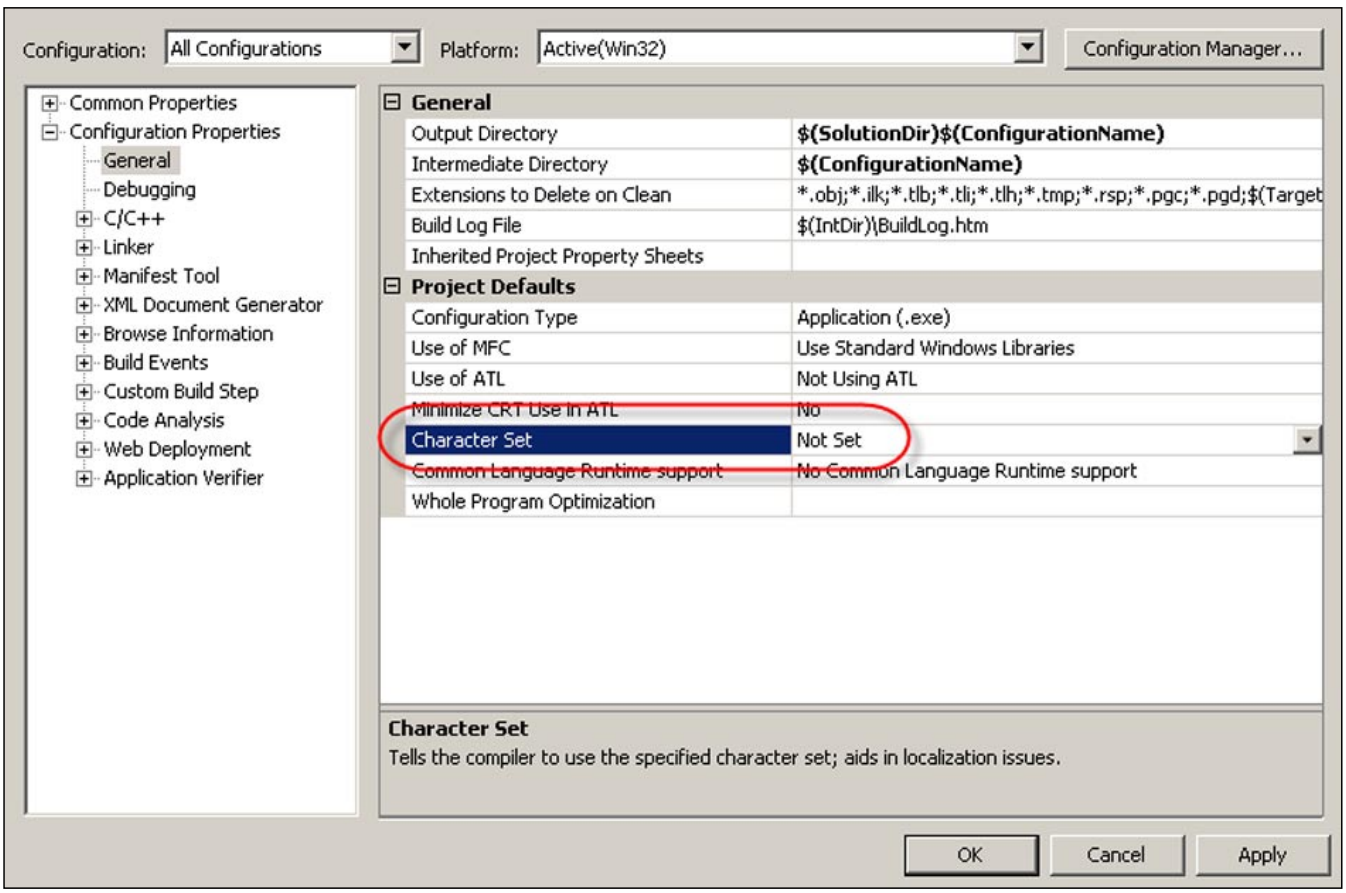
```
scc.FP_MessageBoxA = (TD_MessageBoxA)scc.FP_GetProcAd
                dressA(scc.hmUSER32, szMessageBoxA);
```

Dla każdej potrzebnej z Win32 API funkcji (jak MessageBoxA), trzeba będzie określić typ wskaźnika funkcji:

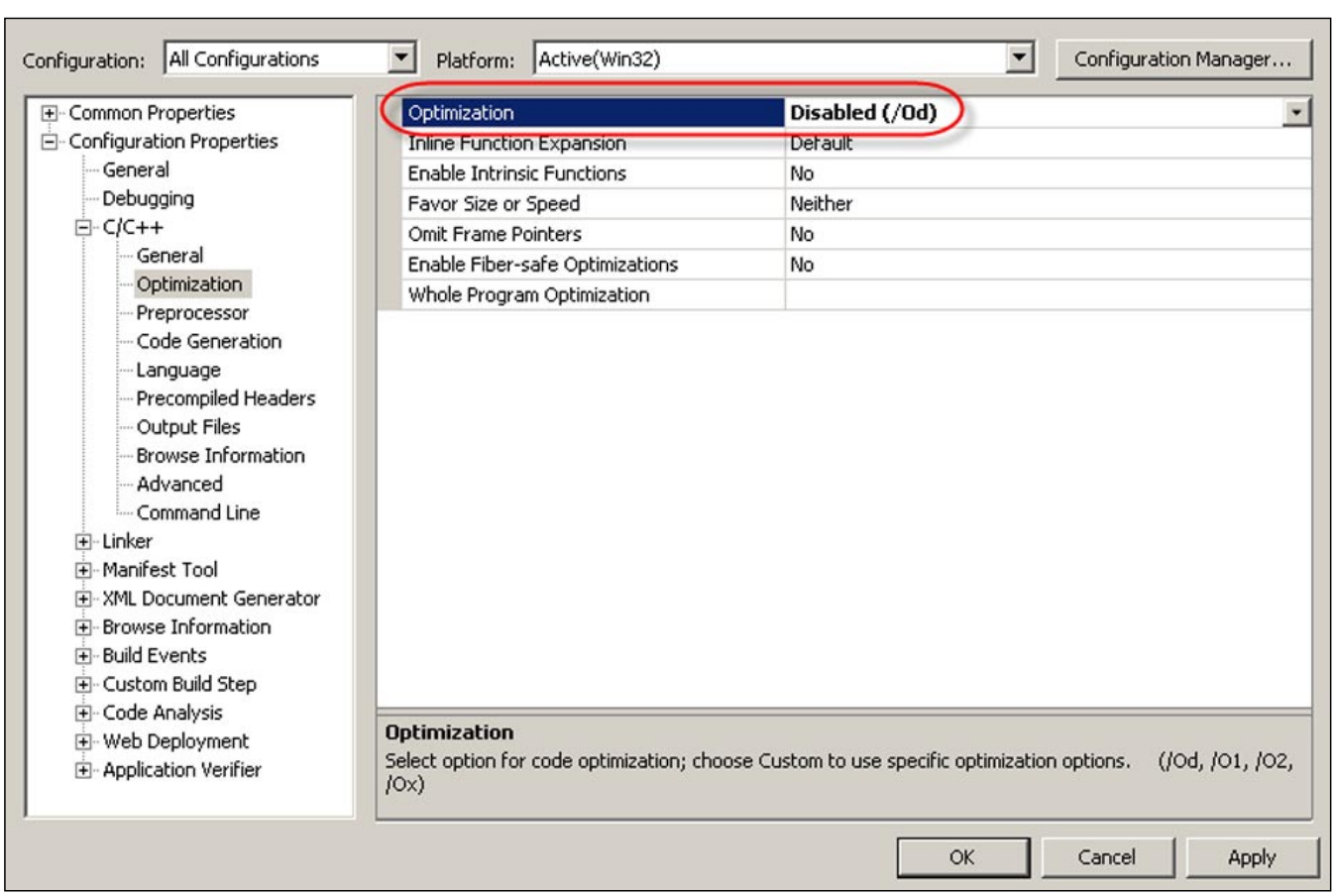
```
typedef int (WINAPI *TD_MessageBoxA)(HWND hWnd,
                LPCTSTR lpText, LPCTSTR lpCaption,
                UINT uType);
```

Aby dokładnie wiedzieć jakiego typu są to wartości zwracane i typy argumentów do zadeklarowania, spójrz do API funkcji w witrynie MSDN (*http://msdn.microsoft.com/en-us/library/ms645505 28VS.85% 29.aspx%*). Zwróć szczególną uwagę na API funkcji, która jako argumenty przyjmują ciągi znaków. Są 2 wersje tych funkcji: wariant ASCII i wariant UNICODE. Ja używam wariantu ASCII MessageBox: MessageBoxA.

To jest wszystko, czego potrzebujemy do instalacji środowiska aby nasz shellcode wykonał się prawidłowo (w naszym przykładzie, wywołanie MessageBox).



Rysunek 1. *Character Set: Nie ustawiono*



Rysunek 2. *Wyłączyć optymalizację*

Jeśli zastanawiasz się, dlaczego postanowiłem wyszukiwać `MessageBoxA` z `GetProcAddressA` i łańcuch znaków, a nie z kodu "The Shellcoder's Handbook" i odpowiedniego hasha, podnosisz ważną kwestię. Nie ma powodu, dlaczego nie można by użyć metody "The Shellcoder's Handbook" do wyszukiwania MessageBoxA. Ale to oznacza, że masz do obliczenia skrót GetMessageBoxA i napisać kilka linijek kodu, aby połączyć kod z wywołaniem getfuncaddress.

I to jest coś, czego chciałem uniknąć, kiedy opracowywałem mój sposób generacji shellcode. Moja metoda, nie wymaga pisania kodu assemblera, ale możesz, jeśli chcesz.

Teraz nasze środowisko jest gotowe, wykonajmy nasz kod rdzenia. Robimy to poprzez wywołanie `ShellCodePayload` i przekazanie jako wskaźnik do struktury kontroli shellcode (linia 210):

```
ShellCodePayload(&scc);
```

`ShellCodePayload` (linia 190) jest łatwa do zrozumienia:

```
char szHello[] = {'H', 'e', 'l', 'l', 'o', '\0'};
pSCC->FP_MessageBoxA(NULL, szHello, pSCC-
                >szEmptyString, 0);
```

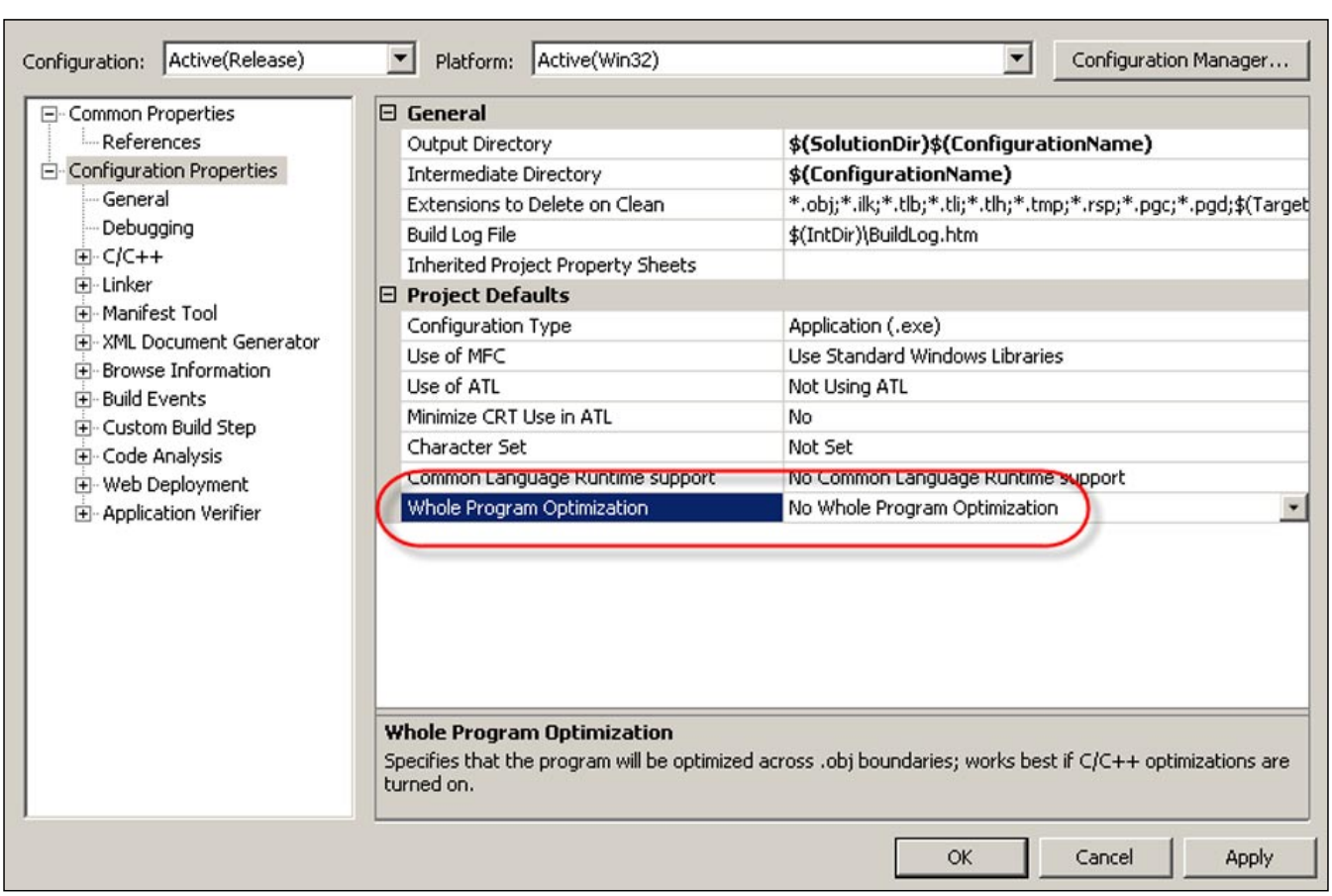
Deklarujemy i wypełniamy łańcuch wartości "Hello", a następnie wywołujemy MessageBoxA z wskaźnikiem funkcji FP_MessageBoxA i przekazuje jej niezbędne wartości, takie jak łańcuch znaków do wyświetlenia.

Aby wygenerować shellcode, kompilujesz program w C. Aby wyodrębnić shellcode z wygenerowanego pliku PE, uruchamiasz program. Shellcode zostanie zapisany w tym samym katalogu co plik .exe, z rozszerzeniem .bin. W przykładzie także wywołuję ShellCodeMain w funkcji main (linia 230). To wykonuje shellcode po uruchomieniu programu i pozwala debugować shellcode wewnątrz Visual Studio Express przy wykorzystaniu wszystkich jego wspaniałych funkcji debugowania! Jeśli nie chcesz, aby shellcode był wykonywany podczas kompilacji, usuń z funkcji main wywołanie ShellCodeMain.

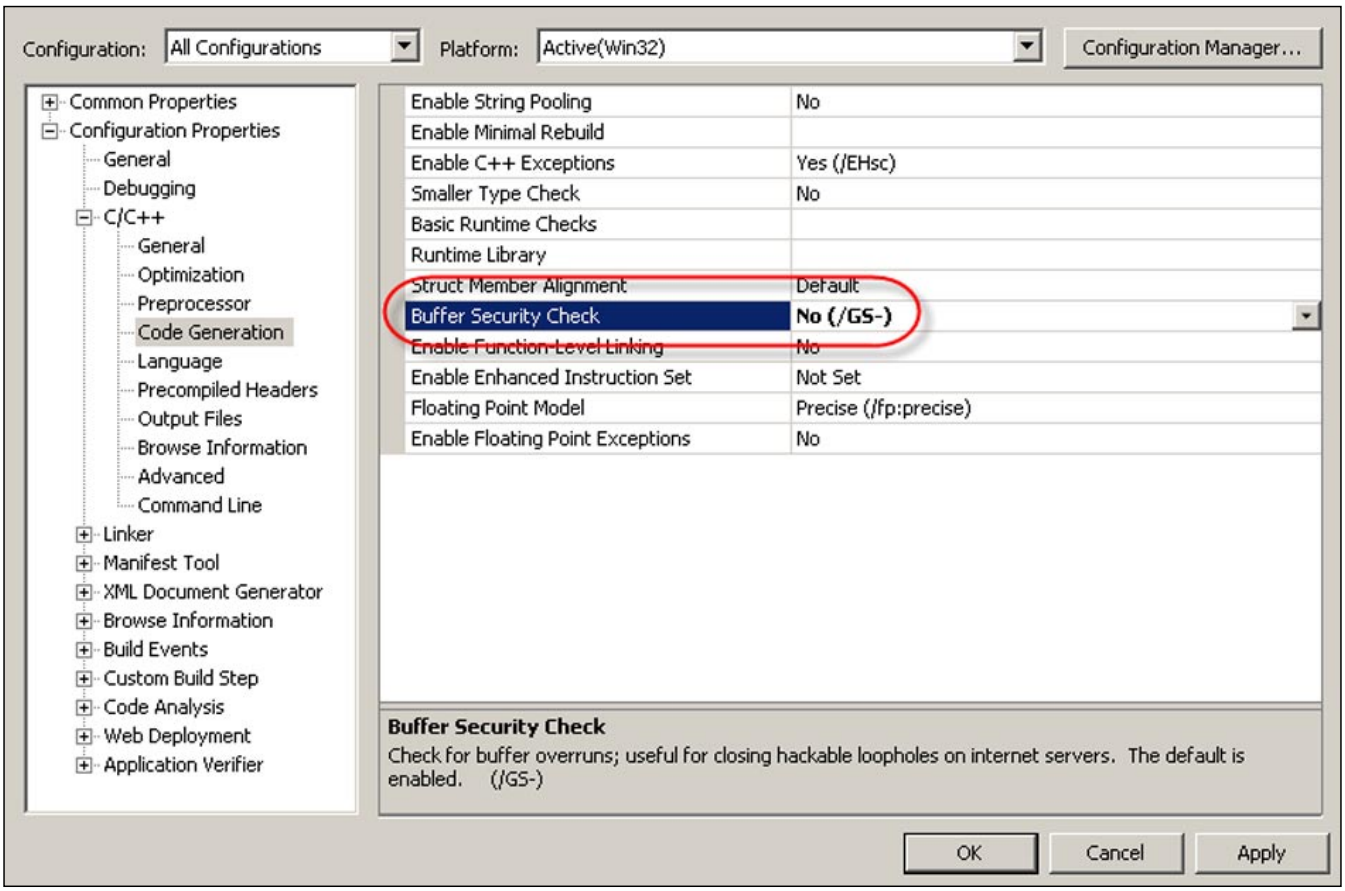
Należy również zmienić kilka właściwości projektu Visual Studio, aby poinstruować kompilator C do wyemitowania odpowiednich kodów które mogą być wykorzystane jako shellcode. Kompilator C nie może emitować plików binarnych w UNICODE, nie może optymalizować kodu i nie może dodawać do kodu ochrony stosu.

Ustaw następujące właściwości projektu jak to widać na Rysunku 1.

Gdy jesteś gotowy do generowania wersji ostatecznej poleć kompilatorowi nie emitowanie kodu debugowania. Przełącz na Release zamiast Debug i usuń wszystkie punkty przerwania jakie masz ustawione.



Rysunek 3. *Wyłączyć optymalizację całego programu*



Rysunek 4. *Wyłączenie ochrony stosu*

Jeszcze jedno na co trzeba zwrócić uwagę to: nie używać funkcji z biblioteki standardowej C, jak strcpy. Jeśli potrzebujesz tych funkcji, albo napisz je samemu albo skorzystaj z podobnych funkcje w ntdll.dll.

Typowa funkcja znajdująca się w shellcode jest wtrąceniem danych (np. pliku) na koniec shellcode. Jest także możliwe napisanie kodu C do osiągnięcia tego celu. Poniżej jest przykład z MessageBox, który wyświetla napis załączony na końcu shellcode.

Musimy dodać element do struktury shellcode aby przechowywać wskaźnik do załączonych danych:

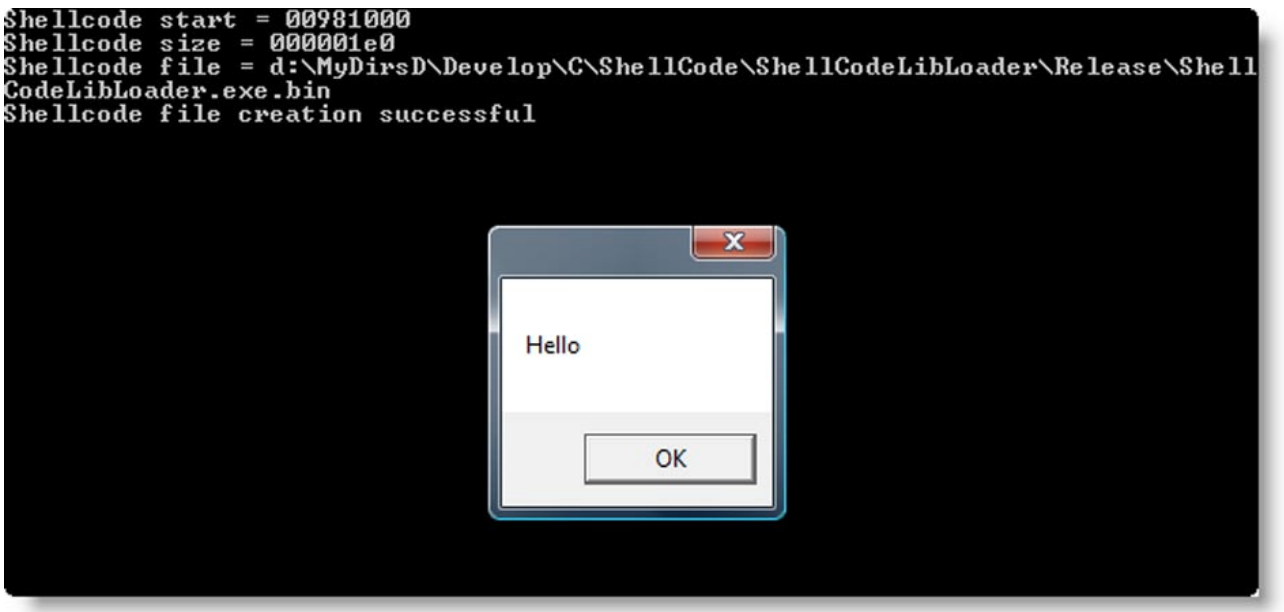
```
void *vpData;
```

Dodaj następujący wiersz do `ShellCodeMain`:

```
scc.vpData = ShellCodeData();
```

`ShellCodeData` to funkcja którą musimy dodać po `ShellCodeMain`:

```
void __declspec(naked) *ShellCodeData(void)
   {
      __asm
      {
         call WhereAmI
      WhereAmI:
```



Rysunek 5. *Kiedy wykonamy program, shellcode jest wyodrębniony i zapisany, a następnie wykonany*

```
                pop eax
        add eax, 5
        ret
        _emit 'R'
        _emit 'e'
        _emit 'p'
        ...
        _emit 0x00
    }
}
```

Ostatnim krokiem jest połączenie MessageBox z tym ciągiem znaków w funkcji ShellCodePayload:

```
pSCC->FP_MessageBoxA(NULL, (LPCTSTR)pSCC->vpData,
                pSCC->szEmptyString, 0);
```

Uruchomienie tego shellcode (z załączonym ciągiem znaków) wyświetla MessageBox z napisem na koniec shellcode (począwszy od pierwszego wystąpienia `_emit`). Jeśli chcesz zmienić ten napis, można po prostu otworzyć plik z shellcode w hex-edytorze i zamienić ciąg znaków w łańcuchu. Nie ma konieczności zmiany kodu źródłowego i rekompilacji projektu.

Po de-asemblacji shellcode generowanego przy użyciu tej metody, można zauważyć, że istnieje szereg wystąpień 0xCC lub INT3 których nie dodaliśmy do naszego kodu źródłowego. Te 0xCC bajty są dodawane przez kompilator aby wyrównać każdą funkcje do granicy 16-bajtów. Powoduje to że shellcode jest większy niż potrzeba.

Jeśli chcesz przekonwertować shellcode z jego binarnego formatu do kodu, należy użyć disassemblera. Ponieważ moim ulubionym asemblerem jest NASM (bezpłatny), używam jego uzupełnienie NDISASM jako disassemblera. Wymaga to pewnego ręcznego czyszczenia kodu, zanim będzie można użyć go w NASM.

W naszym przykładzie nasz shellcode wychodzi poprzez zwrot (oświadczenie ret). Ale można zakodować inne wyjścia:

- wywołanie ExitProcess,
- wywołanie ExitThread,
- ustawienie SEH i spowodowanie wyjątku.

Użyłem tej metody do generowania shellcode dla programu MemoryLoad Joachima Baucha. Jest to kod C, który ładuje biblioteki DLL z pamięci do pamięci. Dostosowałem jego kod źródłowy do mojej metody i byłem w stanie wygenerować shellcode, który ładuje biblioteki DLL z pamięci do pamięci. DLL musi być dołączona na końcu shellcode.

Możesz pobrać szablony i przykłady z mojego bloga: *http://blog.DidierStevens.com/software/shellcode*

DIDIER STEVENS

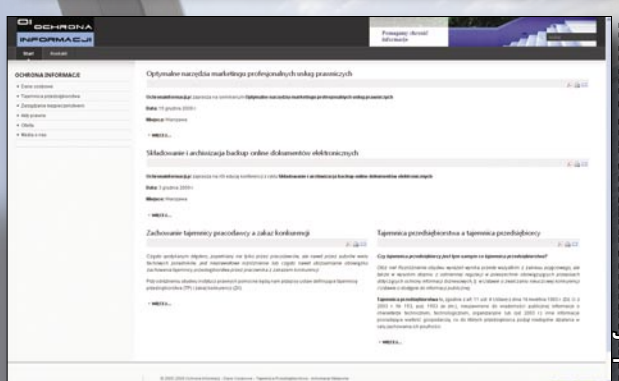
Jest Specjalistą IT Security, specjalizuje się w bezpieczeństwie aplikacji oraz złośliwym oprogramowaniu (malware). Didier pracuje dla Contraste Europe NV. Wszystkie jego narzędzia są open source.

www.mgibki.wordpress.com

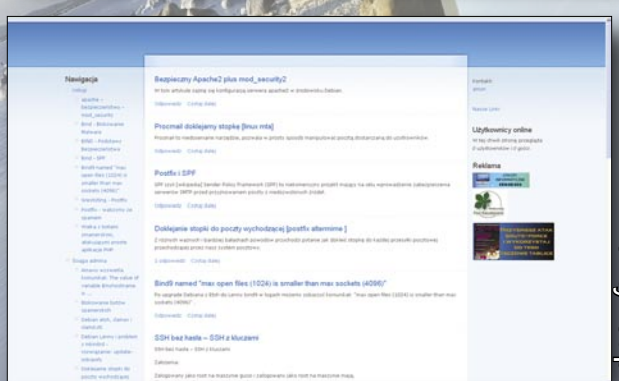
www.hakerzy.net

www.osdev.pl

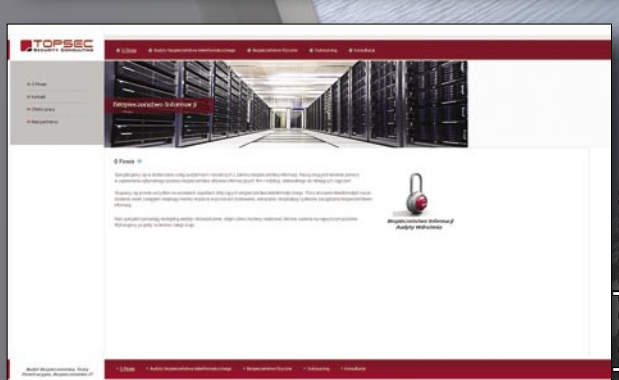
www.ccns.pl



www.ochronainformacji.pl

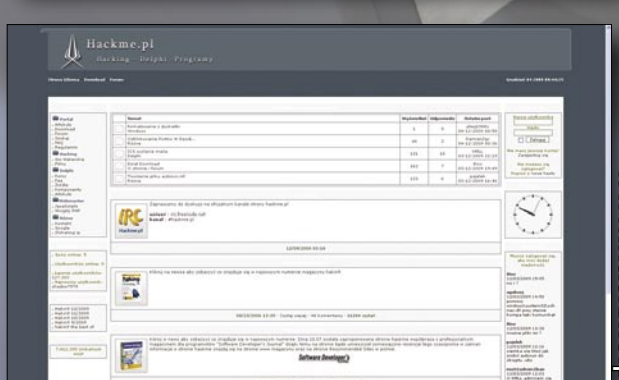


www.antynet.pl

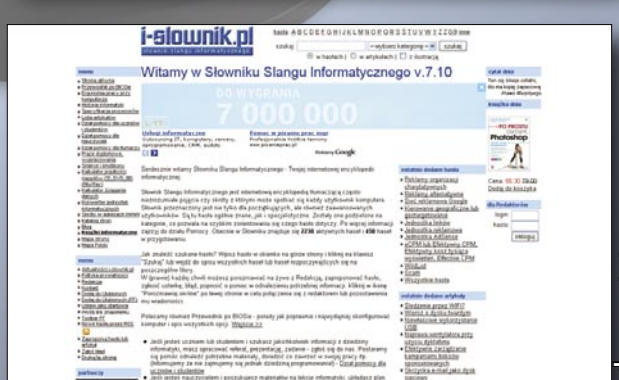


www.topsec.pl

www.hcsl.pl



www.hackme.pl



www.i-slownik.pl

Szkodliwe oprogramowanie dla gier – kompletny ekosystem

Piotr Kupczyk

Jeszcze kilka lat temu nie do pomyślenia było, że ktoś mógłby chcieć ukraść nam wirtualne przedmioty, pieniądze, postaci lub konto z gry wideo. Teraz to rzeczywistość, która dodatkowo stała się poważnym, zorganizowanym biznesem cyberprzestępczym. Biznesem, na którym hakerzy zarabiają prawdziwe, nie wirtualne, pieniądze. A gracze i inni użytkownicy komputerów cierpią...

Dowiesz się:

- W jaki sposób cyberprzestępcy mogą wejść w posiadanie wirtualnych przedmiotów, pieniędzy czy postaci z gier? Dlaczego gracze online są atrakcyjnym celem dla hakerów? Ile jest szkodliwych programów, które mają na celowniku graczy?

Powinieneś wiedzieć:

- Powinieneś posiadać podstawową wiedzę na temat zagrożeń IT.

Na początek dobra wiadomość. Liczba trojanów atakujących użytkowników gier zmniejszyła się na przestrzeni minionych kilku miesięcy. Jednak gracze online wciąż powinni mieć się na baczności. Blisko 12 milionów graczy World of Warcraft stanowi atrakcyjny cel, szczególnie że niedługo ma pojawić się kolejna wersja tego hitu. Po premierze eksperci spodziewają się kolejnego wzrostu liczby tego typu szkodników.

Granie w World of Warcraft (WoW) stało się teraz trochę bezpieczniejsze. W ciągu ostatnich miesięcy analitycy zagrożeń zauważyli spadek liczby ataków na graczy online. Przyczyna tego trendu, co dziwne, jest ekonomiczna i wiąże się z przesyceniem rynku wirtualnych przedmiotów z gier. Obecnie jednak obserwujemy niewielką falę ataków na gry MMORPG (Massively Multiplayer Online Role-Playing Games), takie jak WoW czy Aion. Do ataków cyberprzestępcy wykorzystują najczęściej sprawdzone techniki, takie jak phishing. Okazuje się, że sfałszowany e-mail i spreparowana strona WWW wystarczą, aby oszukać wielu użytkowników Internetu i ukraść ich dane. Obecna fala ataków cybernetycznych pokazuje, że na kradzieży informacji dotyczących gier online wciąż można zarobić dużo pieniędzy.

To, że w ostatnich kilku miesiącach liczba trojanów dla gier online zmniejszyła się, nie oznacza, że jest ich mało. Ile zatem faktycznie jest takich szkodników? Poniższy wykres przedstawia liczbę szkodliwych programów identyfikowanych jako Trojan-GameThief wykrytych na komputerach polskich użytkowników.

Jak widać, do marca liczba trojanów dla gier online wykrywanych w Polsce rosła, osiągając niemal 307 tys., po czym zaczęła powoli spadać, aż do 215 tys. w czerwcu. Warto jednak zwrócić uwagę, że trend w dalszym ciągu jest rosnący – między styczniem a czerwcem nastąpił wzrost o ponad 30 tys. Łącznie, od stycznia do czerwca 2010, trojany dla gier online spowodowały ponad 1,5 mln infekcji na polskich komputerach. Z pewnością będziemy obserwować ciągłe, choć może nie gwałtowne, zwiększanie się popularności tego typu zagrożeń, ponieważ graczy online jest coraz więcej.

Na kolejnym wykresie możemy zaobserwować jak w ciągu roku zmienił się odsetek trojanów dla gier online wśród wszystkich szkodliwych programów infekujących komputery użytkowników – dane dla całego świata. W pierwszym kwartale 2009 r. szkodniki typu Trojan-GameThief były odpowiedzialne za 5,95% infekcji, natomiast w pierwszym kwartale 2010 r. już za 7,63% zarażeń. Warto zwrócić uwagę na to, że są to jedyne trojany, które zanotowały w tym czasie istotny wzrost w swojej kategorii.

Gry rządzą się własną ekonomią rynkową

Popularne gry online stworzyły własną ekonomię rynkową, która rządzi ich wirtualnymi przedmiotami. Za ukończenie misji dobrzy wojownicy i stratedzy są nagradzani wartościowymi przedmiotami i wirtualną gotówką, przy pomocy których mogą jeszcze bardziej rozwinąć swoje postacie online. Gracze, którzy inwestują dużo czasu i wysiłku w takie gry, mogą stworzyć niezwykle wartościowe awatary online. Te wirtualne kosztowności naturalnie stanowią pożądane przedmioty i mogą być sprzedawane za "prawdziwe" pieniądze. Szczególnie w Korei Południowej i Ameryce Północnej istnieje duża liczba profesjonalnych graczy. Takie osoby grają przez cały dzień, aby zgromadzić wirtualne przedmioty, które następnie zamieniają na gotówkę.

Włamania na konta

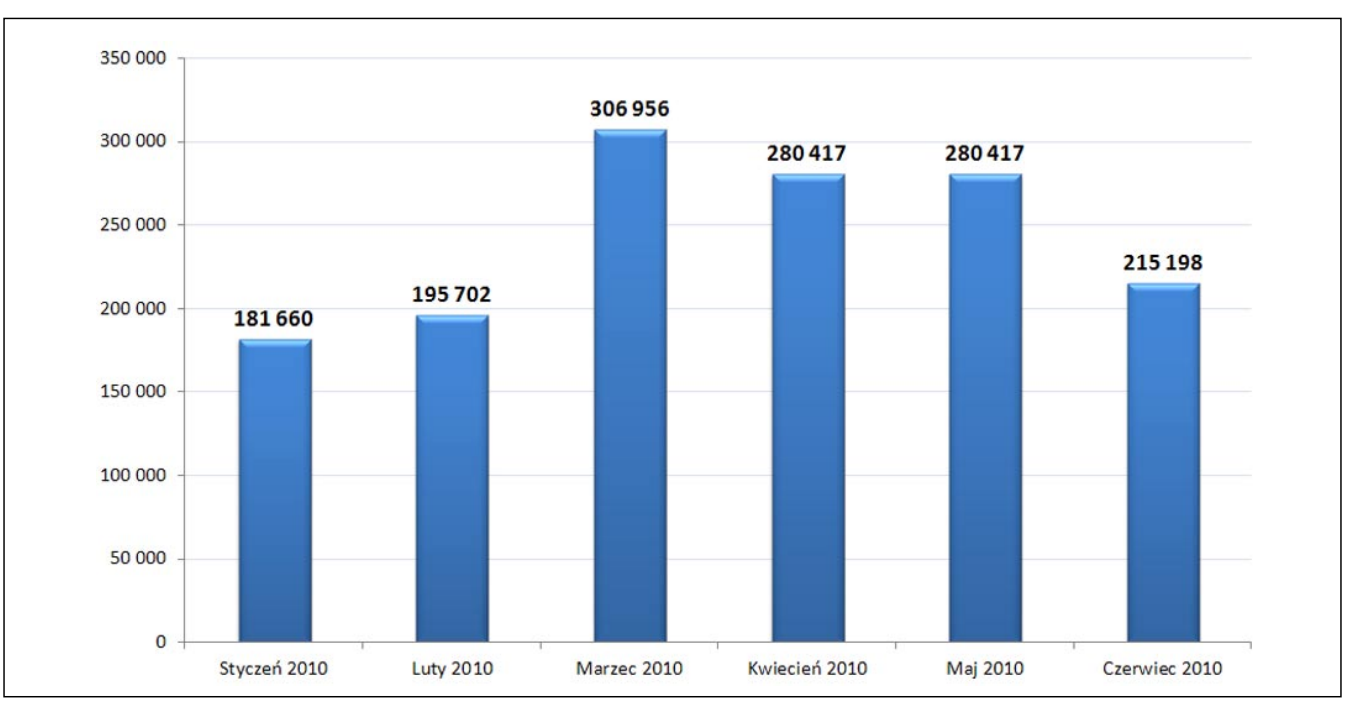
Jednym ze sposobów zdobycia bogactw należących do innych graczy jest włamanie się do ich kont. Według danych Kaspersky Lab, istnieje około 1,73 miliona wyspecjalizowanych szkodników dla komputerów, które próbują uzyskać hasła i dane dostępu do gier online. W ten sposób hakerzy chcą osiągnąć kilka rzeczy. Osoby grające w gry online są uważane za ekspertów komputerowych. Ta grupa nie używa Internetu jedynie do grania. Oprócz tego dokonują również zakupów i transakcji online. Dlatego cyberprzestępcy mogą zebrać wiele informacji, łącznie z danymi dotyczącymi karty kredytowej i hasłami do innych serwisów, takich jak portale społecznościowe.

Kolejną korzyścią dla złodziei jest to, że gracze online zwykle posiadają bardzo wysokiej jakości sprzęt komputerowy i szybkie łącza internetowe. Dlatego stanowią atrakcyjny cel operatorów botnetów, które często są określane jako sieci komputerów zombie. Komputery wchodzące w skład takich sieci nie są już kontrolowane przez ich właścicieli: zdalną kontrolę za pośrednictwem Internetu sprawują nad nimi przestępcy. Komputery zombie mogą być na przykład wykorzystane jako broń do przeprowadzenia ataków na inne systemy komputerowe. Wysoce wydajne komputery PC, takie jak te wykorzystywane przez graczy online, są naturalnie preferowanym celem operatorów botnetów.

Kradzież rzeczywista i wirtualna

Jednak w jaki sposób przestępcy atakują graczy online? Wykorzystywane są różne metody - jedną z nich jest wspomniany wcześniej phishing. Niedawno, na przykład, wielu użytkowników platformy Steam (`http://store.steampowered.com`) otrzymało fałszywe wiadomości z prośbą o aktualizację ich informacji osobowych. Wiadomości zawierały również odsyłacz, który należało kliknąć, aby uzyskać dostęp do własnego profilu. Podobnie jak w przypadku serwisów bankowości online i innych, dostawcy platform dla graczy nigdy nie wysyłają tego typu wiadomości pocztą elektroniczną. Dlatego mamy tu wyraźnie do czynienia ze złodziejami haseł, którzy próbują wyłudzić dane logowania użytkowników. Zamieszczony odsyłacz naturalnie nie prowadził do profilu użytkownika, ale do fałszywej strony internetowej stworzonej w celu przechwytywania informacji logowania.

Oprócz kradzieży informacji logowania "w realnym świecie" gracze padają również ofiarą "wirtualnej" kradzieży. `ENISA`(European Network and Information Security Agency) odnotowała niepokojący wzrost liczby



Rysunek 1. *Liczba trojanów dla gier wykrytych na komputerach polskich użytkowników w roku 2010*

oszustw w świecie wirtualnym, a około jedna trzecia graczy padła kiedyś ofiarą kradzieży własności wirtualnej. Fakt, że w 2007 roku na samej sprzedaży wirtualnych przedmiotów zarobiono około 2 miliardów amerykańskich dolarów, nie pozostawia wątpliwości, że jest to bardzo lukratywny rynek - taki, w którym hakerzy próbują przyjąć tożsamość online użytkowników, aby uzyskać dostęp do wirtualnego bogactwa.

Serwisy wymiany plików, które odwiedzają gracze, aby zaopatrzyć się w najnowsze cheaty, stanowią popularne źródło szkodliwego oprogramowania, podobnie jak luki w zabezpieczeniach serwerów gier. Ponieważ oficjalne serwery gier zwykle są dobrze utrzymane, korzystający z nich gracze są stosunkowo bezpieczni. Jednak serwery pirackie, które pozwalają graczom bezpłatnie brać udział w grach multiplayer to zupełnie inna historia.

Nowe włamania

Oszuści online są niezwykle twórczy jeżeli chodzi o łamanie nowych systemów bezpieczeństwa. Widać to na przykładzie generatora kodów TAN dla WoW. Jest to niewielki dodatkowy sprzęt, który generuje losowe kombinacje numerów, przy pomocy których gracze mogą na krótki czas zalogować się do gry. Niedawno, przy użyciu trojana cyberprzestępcom udało się ukraść konto gracza za pośrednictwem tak zwanego ataku "man-in-the-middle" (`http://forums.wow-europe.com/thread.html?topicId=12730404058&sid=1&pageNo=1`). Świadczy to zarówno o aktywności jak i szybkim reagowaniu cyberprzestępców. Dlatego ważne jest, aby użytkownicy nie mieli fałszywego poczucia bezpieczeństwa z powodu obecnego spadku liczby szkodliwych programów. Za każdym razem, gdy publikowane są nowe dodatki do gier, eksperci przewidują natychmiastowy wzrost liczby ataków. Kolejnym niecierpliwie oczekiwanym wydarzeniem jest premiera Diablo III.

Zabezpiecz się

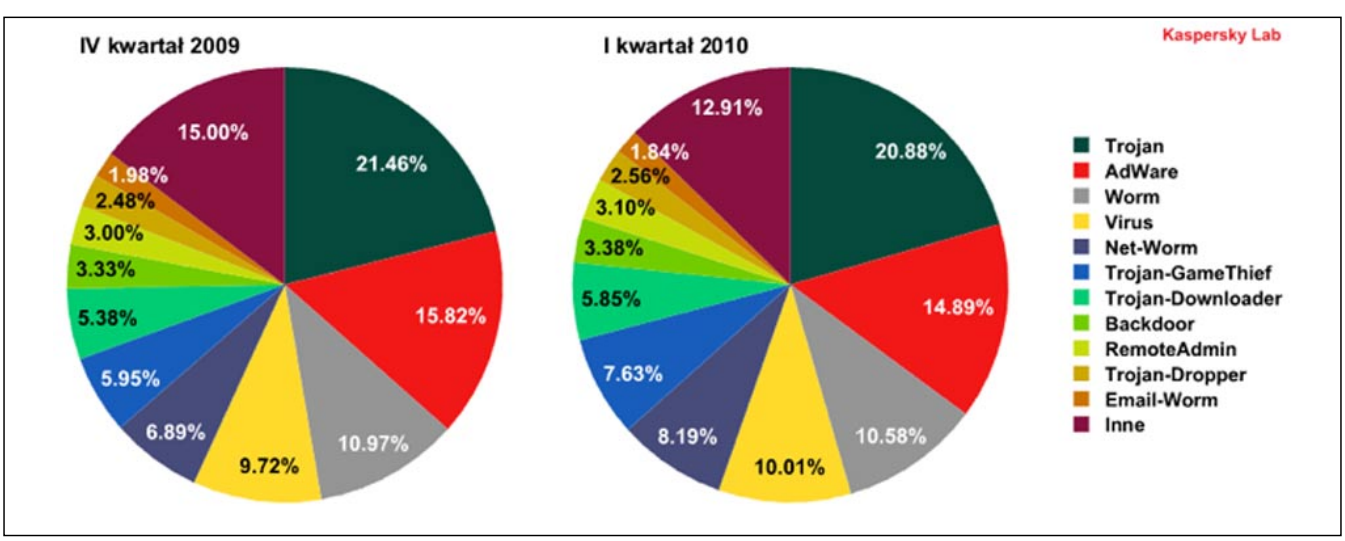
Jest kilka rzeczy, o których powinni pamiętać gracze online, aby zabezpieczyć siebie i swoje systemy. Tak jak w przypadku innych aktywności internetowych, uaktualnione oprogramowanie antywirusowe jest koniecznością dla osób, które grają w gry online, i nie powinno zostać wyłączone podczas gry. Ważne jest również to, aby pakiet oprogramowania antywirusowego automatycznie pobierał sygnatury zagrożeń.

Optymalna ochrona antywirusowa dla graczy

Jeżeli chodzi o oprogramowanie antywirusowe, gracze mają szczególne wymagania. Programy bezpieczeństwa muszą zapewnić optymalną ochronę, a przy tym nie powinny zakłócać gry ani wpływać na wydajność systemu. Kaspersky Internet Security posiada „Tryb gracza", w którym nie są wyświetlane okienka wyskakujące przypominające o aktualizacjach i skanowaniu. W ten sposób oprogramowanie bezpieczeństwa nie zakłóca gry, ale działa w tle. Program nie wywiera wpływu na wydajność systemu, gra przebiega bez zakłóceń, a użytkownik posiada optymalną ochronę.

Uwaga na cheaty

Gracze powinni mieć się na baczności przez cały czas, nawet gdy nie grają. Cheaty mające za zadanie ułatwienie gry można rozpoznać po tym, że wiele obiecują, lepiej jednak całkiem omijać takie programy. Podobnie jak banki, dostawcy gier online nie wysyłają e-maili żądających informacji osobowych: takie wiadomości mogą pochodzić jedynie od złodziei haseł. Nie powinniśmy również ujawniać innym graczom naszych danych wykorzystywanych podczas logowania i dokonywania płatności. Aby zapewnić sobie niczym niezmąconą rozrywkę, najlepiej korzystać tylko z oficjalnych serwerów gier. Pirackie serwery obiecujące darmowe gry są kuszące, często jednak zawierają mnóstwo luk bezpieczeństwa.



Rysunek 2. *Rozkład szkodliwych programów odpowiedzialnych za infekcje wg kategorii*

Przykład z życia wzięty

W marcu 2010, gdy obserwowaliśmy nasilenie aktywności cyberprzestępców, Internet zalała fala wiadomości phishingowych wycelowanych w fanów dwóch gier online: World od Warcraft oraz Aion. Oczywiście, takie ataki nie są żadną nowością, jednak te konkretne zostały przygotowane niezwykle starannie.

Adres nadawcy wiadomości phishingowych był spreparowany i mógł wyglądać następująco: „noreply@blizzard.com". Temat oszukańczych wiadomości brzmiał „World of Warcraft - Account Change Notice" a ich treść była przygotowana bardzo profesjonalnie. Po pierwsze, e-maile cechowały się bezbłędnością językową, po drugie, wygląd portalu internetowego, na którym przechwytywane były dane uwierzytelniające, do złudzenia przypominał oryginał. Z pewnością nie była to „tania" podróbka, z jakimi spotykaliśmy się wcześniej wiele razy. Nawet doświadczeni użytkownicy mogli paść ofiarą takiego oszustwa.

Odsyłacz widniejący w sfałszowanej wiadomości e-mail nie prowadził do oryginalnej strony WWW, ale do skryptu, który kierował użytkownika do witryny stworzonej przez oszustów po to, aby gracze wprowadzili dane uwierzytelniające dostęp do swojego konta gry online. W ten sposób oszuści przechwytują dane uwierzytelniające do kont gier MMORPG, a następnie sprzedają je na aukcjach internetowych. Istnieje wiele trojanów kradnących dane uwierzytelniające użytkowników. Wiele z nich przechwytuje klawisze wciskane podczas wchodzenia przez użytkownika do okna logowania się do gry. Informacje te są zapisywane w zwykłym pliku tekstowym wysyłanym do autora trojana.

W ciągu ostatnich 3-4 lat cyberprzestępcy uzyskali spore korzyści finansowe przy pomocy oszustw związanych z grami online, w szczególności wykorzystując konta, postacie oraz przedmioty z gry World of Warcraft. Konta z dobrze rozwiniętymi postaciami mogą kosztować nawet 1 000 - 3 000 dolarów amerykańskich.

Jak zmniejszyć ryzyko infekcji?

Najważniejszym i niezbędnym ogniwem ochrony jest aktualna aplikacja antywirusowa, a najlepiej typu Internet Security. Dotyczy to wszystkich użytkowników Internetu. Poniżej krótka listę porad, które pomogą w podniesieniu bezpieczeństwa i zmniejszą prawdopodobieństwo infekcji.

1. Uaktualniaj system Windows i aplikacje innych producentów.
2. Regularnie twórz kopię zapasową swoich danych na płycie CD, DVD lub zewnętrznym nośniku USB.
3. Nie odpowiadaj na e-maile ani wiadomości wysyłane za pośrednictwem portali społecznościowych, jeżeli nie znasz ich nadawcy.
4. Nie klikaj załączników do wiadomości e-mail ani obiektów wysyłanych za pośrednictwem portali społecznościowych przez nieznanych nadawców.
5. Nie klikaj odsyłaczy w wiadomościach wysyłanych za pośrednictwem poczty elektronicznej ani komunikatorów internetowych. Wpisuj adresy bezpośrednio do przeglądarki internetowej.
6. Nie podawaj informacji osobowych w odpowiedzi na maile, nawet jeżeli wyglądają na oficjalne wiadomości.
7. Dokonuj zakupów oraz transakcji bankowych na bezpiecznych stronach. Ich adresy zaczynają się od ciągu znaków „https://", a w oknie przeglądarki wyświetlana jest złota kłódka.
8. Używaj innych haseł na różnych stronach internetowych lub serwisach, z których korzystasz. Twórz hasła, które składają się z więcej niż 5 znaków i zawierają liczby, znaki specjalne oraz wielkie i małe litery. Nie twórz nowych haseł na podstawie starszych (np. „hasło1", „hasło2") i unikaj takich, które można łatwo odgadnąć (np. imię Twojej mamy).
9. Nie zdradzaj nikomu swoich haseł.
10. Nie zapisuj swoich haseł w postaci niezaszyfrowanej, na przykład w pliku tekstowym na dysku komputera.

Podsumowanie

Spadek liczby zagrożeń dla gier online, który zaobserwowaliśmy w ostatnich miesiącach może być przysłowiową ciszą przed burzą. Popularność gier nie maleje, przed nami kilka premier, na które czeka wielu pasjonatów elektronicznej rozrywki, a aukcje z przedmiotami z gier ciągle się mnożą. Nic nie wskazuje na to, aby trojany dla gier miały zniknąć z cyberprzestępczego krajobrazu. Pozostaje nam mieć się na baczności, korzystać wyłącznie z oficjalnych serwerów dla graczy i zapewnić należytą ochronę naszych komputerów. Przecież stracić możemy nie tylko wirtualne przedmioty...

PIOTR KUPCZYK

Dyrektor działu prasowego w Kaspersky Lab Polska (producent rozwiązań do ochrony danych). Z firmą związany od pierwszego dnia funkcjonowania jej polskiego przedstawicielstwa w 2000 roku. Z wykształcenia jest informatykiem. Ukończył kierunek Techniki Multimedialne i Metody Sztucznej Inteligencji na Politechnice Częstochowskiej. Interesuje się wszystkim, co związane z zagrożeniami internetowymi. Prywatnie pasjonat rocka progresywnego oraz gier wideo.

Kontakt z autorem: piotr.kupczyk@kaspersky.pl

Monitorowanie sieci

Sylwester Zdanowski

Monitoring sieci pozwala na zautomatyzowane informowanie administratora o pojawiających się problemach. Jednak w chwili otrzymania informacji o problemie jest on już odczuwalny dla użytkowników. Można jednak wykryć część problemów zanim staną się odczuwalne.

Dowiesz się:
- Jak wykorzystać LMS do lokalizacji usterek w sieci
- Jak monitorować pracę elementów sieci używajac NAGIOS

Powinieneś wiedzieć:
- Jak poruszać się w konsoli Linuksa

Stwórzmy na nasze potrzeby sieć w niewielkim biurze. Przede wszystkim mamy router ADSL stanowiący połączenie ze światem. Dalej dwa switche, każdy obsługujące inne piętro oraz jeden AP. Oczywiście biuro ma kierownika z dwoma komputerami, w tym jednego laptopa. Następnie marketing i zaopatrzenie po jednym komputerze. Zostaliśmy my jako informatyk i księgowość mająca dwa komputery.

W ramach naszego monitoringu stworzymy za pomocą LMS mapę sieci pokazującą połączenia oraz aktywność komputerów. Dzięki narzędziu NAGIOS stworzymy monitoring zasobów systemów. Dzięki temu będziemy wiedzieć kiedy ilość wolnego miejsca na hdd dowolnej maszyny ulegnie zmniejszeniu lub gdy pojawi sie podejrzanie duża ilość zalogowanych użytkowników.

Mając tak niewielką sieć można pokusić się o monitorowanie dosłownie wszystkiego co może spowodować problemy. Podejście takie niema jednak zastosowania w większych sieciach. Przede wszystkim uwagi wymagają urządzenia istotne dla pracy firmy. Monitorowanie wszystkiego spowoduje chaos informacyjny i trudności w określaniu priorytetów. W prosty sposób może dojść do sytuacji gdzie technik zasypany informacjami o zapełniającym się dysku szeregowego pracownika przeoczy komunikat o awarii głównego routera.

Dlatego też konieczne jest planowanie monitoringu. Zarówno pod względem jego zakresu jak i metod powiadamiania zależnie od znaczenia usług.

LMS

Lan Managment System jest przeznaczony dla firm oferujących dostęp do internetu. Na nasze potrzeby użyteczne będą jedynie niektóre z jego możliwości. Prócz określania położenia urządzeń możliwe jest dodanie monitoringu ilości przesyłanych danych, wysyłanie korespondencji zbiorowej, przyjmowanie zgłoszeń czy stworzenie kolejki zadań do wykonania.

Sama instalacja LMS jest niezwykle prosta i dobrze opisana na stronach projektu. Zasadniczo potrzebny będzie serwer LAMP. Do uruchomienia LMS wystarczy konfiguracja vhosta oraz skopiowanie bazy danych.

Wizualne efekty wykorzystania LMS można zobaczyć na ilustracji pierwszej. Pokazuje ona naszą sieć znacznie lepiej niż wcześniejszy opis.

Przed stworzeniem takiej mapy warto wprowadzić małą zmianę w konfiguracji. W interfejsie użytkownika można dodać parametry konfiguracji. Na nasze potrzeby wystarczy dodanie *allow_mac_sharing* z wartością 1 do sekcji *phpui*. Dzięki temu łatwiejsze będzie dodawanie urządzeń dla których nie znamy adresu MAC.

Zaczniemy jednak od dodania klientów którzy zostaną posiadaczami komputerów. W naszym przypadku najlepsze będzie dodanie nazw działów wraz z numerami pokojów. W następnym kroku musimy przejść do *Sieci IP* aby dodań nową sieć. Dzięki temu dodając nowe urządzenia będziemy mogli wybierać wolny

IP z listy. W tym momencie jasne powinno być iż musimy ustawić maszyny ze stałymi adresami IP lub wykorzystać DHCP z adresami IP przypisanymi do adresów MAC.

Korzystając z widoku klientów możemy dodawać im nowe komputery. Jednak nie mamy ich jeszcze gdzie podłączyć. Do dodawania urządzeń służy opcja *osprzęt sieciowy*. Należy pamiętać o dodawaniu ilości połączeń jakie mogą występować z danym urządzeniem. Same połączenia można tworzyć w trakcie dodawania lub edycji urządzeń wybranych z listy.

Ostatnim etapem jest dodanie komputerów użytkowników i ich podłączenie. Warto pamiętać o dodatkowej opcji dla oznaczenia połączeń bezprzewodowych. Ma to zasadnicze znaczenie dla użyteczności mapy przy rozwiązywaniu problemów.

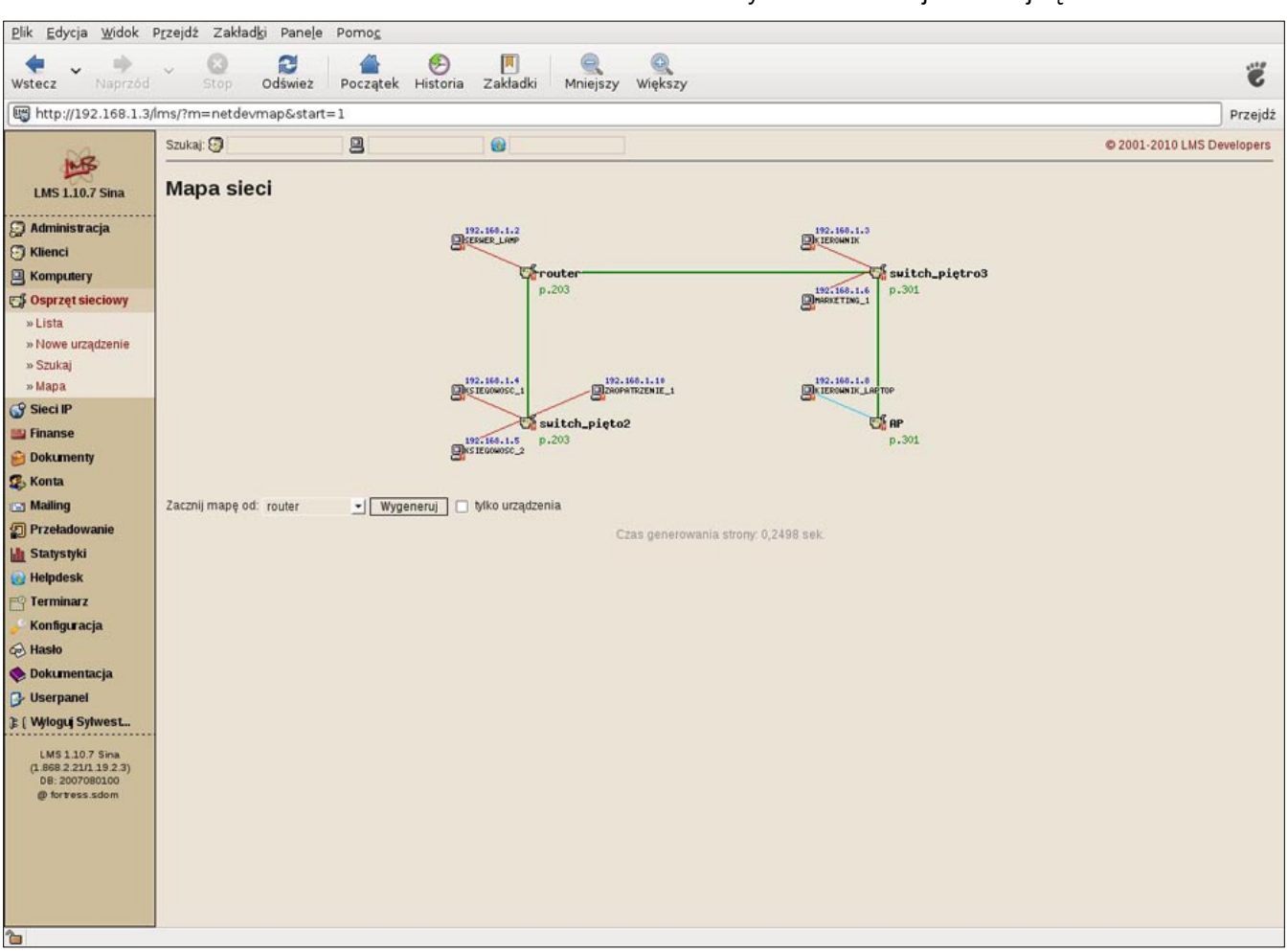
Aby całość była użyteczna należy dodać sprawdzenie które maszyny są dostępne. W katalogu *lms/bin* znajduje się skrypt *lms-fping*. Wykorzystując program *fping* sprawdzi on aktywność wszystkich komputerów. Dzięki temu przy wszystkich pracujących komputerach zniknie znak zapytania. To samo stanie się z urządzeniami podłączonymi do działających maszyn. Jeżeli za jakiś czas przestaną one odpowiadać ikony staną się ciemne.

Listing 1. *host_template.cfg*

```
define host{
        name                            lan-host
        notifications _ enabled   0
        event _ handler _ enabled         0
        flap _ detection _ enabled        1
        failure _ prediction _ enabled    1
        process _ perf _ data             1
        retain _ status _ information     1
        retain _ nonstatus _ information  1
        check _ period                    24x7
        check _ interval                  5
        retry _ interval                  1
        max _ check _ attempts            10
        check _ command           check-host-alive
        register                          0
    }
```

Jeżeli dodamy skrypt do *cronetab* będziemy wiedzieć na bieżąco które urządzenia działają. Przy większych sieciach mechanizm taki niezwykle ułatwia określenie miejsca w którym występuje problem.

Mając szybko skleconą mapę sieci zobaczmy co jeszcze możemy zrobić z LMS. Idąc po kolei przez menu, jako że nie sprzedajemy internetu finanse, dokumenty i konta raczej nie znajdą zastosowania. Je-



Rysunek 1. *Widok sieci w LMS*

żeli dodamy do systemu dane pracowników wraz z ich adresami e-mail, wysłanie korespondencji seryjnej może się okazać użyteczne. Tym bardziej że możliwe jest połączenie bazy MySQL z pakietem OpenOffice. Następna opcja przeładowania jest potrzebna przy wykorzystaniu serwera *pppoe* co nas nie dotyczy. Przydatna może być funkcja *helpdesk* oraz terminarz. Szczególnie jeżeli mamy do czynienia z kilkoma rozproszonymi sieciami, ich użytkownicy mogą zgłaszać swej pytania i problemy za pomocą panelu użytkownika współpracującego z LMS.

Panel użytkownika dzięki modułowej konstrukcji bardzo łatwo ograniczyć do samego zgłaszania problemów. Dzięki czemu użytkownicy nie będą mieli pojęcia o jego wszystkich możliwościach. Rozwiązanie to jednak wymaga utworzenia kont dla pracownikół przez ich dodanie do LMS jako klientów.

NAGIOS na skróty

W porównaniu do LMS dokumentacja stworzona dla NAGIOSa jest niewiele warta. Trzeba jednak się z nią męczyć jako jedynym źródłem informacji. Aby ułatwić sobie zdanie warto wyszukać instrukcję instalacji dla posiadanej dystrybucji. Będzie ona szczególnie przydatna dla instalacji dodatkowych pakietów.

Dla ograniczenia zamieszania przy instalacji do polecenia *configure* warto dodać | *grep gdImagePng.* Jeżeli nie otrzymamy informacji potwierdzającej odnalezienie bibliotek należy je zainstalować przed przejściem dalej. Gdy po instalacji okaże się iż niemożna korzystać z opcji mapy i trendu naprawa problemu będzie kosztować znacznie więcej czasu.

Mając zainstalowane wtyczki i ustawione hasło zgodnie z dokumentacją NAGIOSa pozostaje jego uruchomienie. W katalogu *bin* należy uruchomić program NAGIOS podając jako parametr ścieżkę do pliku konfiguracyjnego.

Pozostaje nam stworzenie konfiguracji monitorującej jak najwięcej parametrów pracy wszystkich elementów sieci. Możemy do tego wykorzystać pliki *localhost.cfg* oraz *switches.cfg*. Przy czym plik *switches.cfg* należy dodać do *nagios.cfg*. W naszym prostym wypadku wystarczy dodanie hostów na zasadzie analogi do *localhost.cfg*.

Zaczynając od routera konieczne będzie doinstalowanie wtyczki odpowiedzialnej za SNMP. Z kolei do jej skompilowania w systemie muszą się znajdować biblioteki net-snmp. W następnej kolejności switch i nasz serwer. Dla switchy nieposiadających własnych adresów IP można nie podawać żadnych usług.

Listing 2. *host.cfg*

```
define host{
            use                     lan-host
            host_name               castle
            alias                   castle
            address                 192.168.1.101
    }
```

Listing 3. *service.cfg*

```
define command{
            command_name    lan_ssh_disk
             command_line    $USER1$/check_by_ssh -p $ARG1$ -l sylwester -i /home/nagios/.ssh/id_rsa
-H $HOSTADDRESS$ -C '~/check_disk -w $ARG2$ -c $ARG3$ -p $ARG4$'
    }
    define service{
            use                             generic-service
            notifications_enabled           0
            host_name                       castle
            service_description             SSH check Disk
            check_command                   lan_ssh_disk!22!10%!5%!/dev/hda2
    }
```

Listing 4. *Zastosowanie polecenia snmptranslate*

```
snmptranslate -IR -On IP-MIB::ipDefaultTTL
No log handling enabled - turning on stderr logging
Undefined OBJECT-GROUP (udpHCGroup): At line 486 in /usr/share/snmp/mibs/udp-mib.mib
.1.3.6.1.2.1.4.2
```

W efekcie będą one widoczne jako element sieci bez monitoringu.

Sam serwer stanowi najprostszy element. Jeżeli na nim właśnie znajduje się NAGIOS od ręki będzie on monitorowany. Wystarczy dodanie parametru *parents* aby został połączony z routerem.

Jeżeli na pozostałych komputerach zainstalowany jest system Windows będą one wymagały dodatkowych czynności. Konieczne jest na nich zainstalowanie *NSClient++* który umożliwi monitoring. Również konfiguracja wymaga zastosowania oddzielnych poleceń.

Po zakończeniu konfiguracji nie tylko możemy kontrolować sieć. Niezwykle użyteczna może okazać się opcja trendów. Szczególnie gdy po 5minutowej awarii wszechwiedzący kierownik stwierdzi że sieć ciągle nie działa i chce wprowadzać zmiany kadrowe.

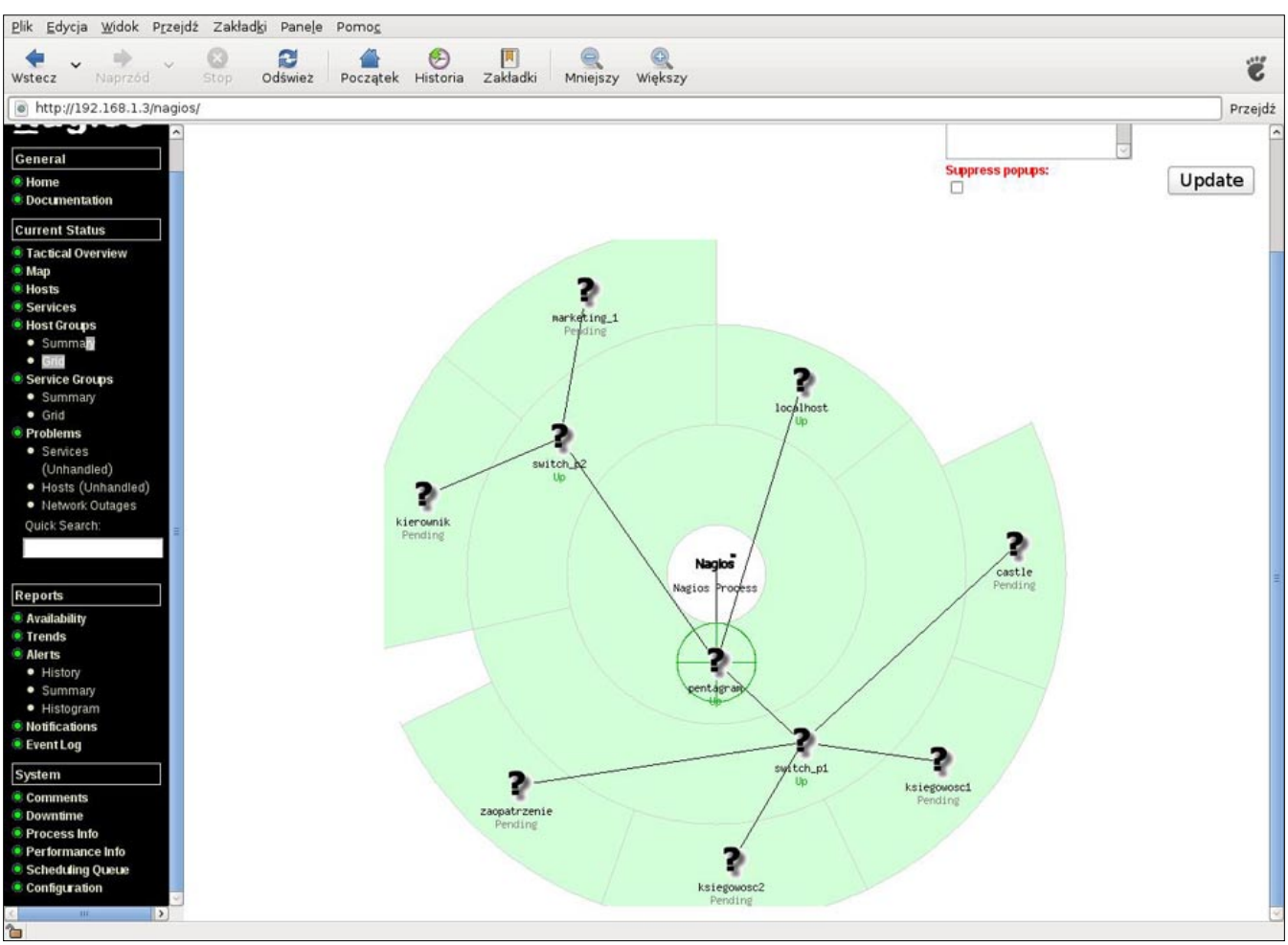
Idea wielkiej sieci

Podejście na skróty do NAGIOSa niema racji bytu przy implementacji dla większej sieci. Konieczne staje się zapoznanie z zasadami monitoringu i grupowaniem, tak usług jak i osób za nie odpowiedzialnych. Obejmują one nie tylko kwestie konfiguracji ale również ustalenie priorytetów. Mając pod swoją opieką sieć z setką usług monitorowanie wszystkich doprowadzi jedynie do chaosu. Monitoring powinien ograniczać się do elementów istotnych dla działania firmy. Z dużym prawdopodobieństwem konieczne będzie rozpoczęcie od usług widocznych dla klientów. Wysoki priorytet otrzymają również serwery których awaria wpłynie na firmę lub całe jej oddziały. W dalszej kolejności wielofunkcyjne urządzenia z interfejsami sieciowymi.

Punktem wyjścia samej konfiguracji jest oczywiście plik *nagios.cfg*, nie obędzie się bez dodania nowego pliku konfiguracyjnego. Możliwe jest również dodanie całego katalogu zawierającego dane obiektów. Dodamy więc wpis *cfg_dir=/usr/local/nagios/etc/objects/lan.* W katalogu przydadzą nam się na początek dwa pliki, *lan_templates.cfg* i *lan_hosts.cfg*. W razie potrzeby dla zachowania przejrzystego podziału można się pokusić o apache_servers czy dmz_hosts, jest to kwestia potrzeb.

Należało by stworzyć szablon dla naszych komputerów w lokalnej sieci.

Na nasze teoretyczne potrzeby stworzyliśmy szablon wyłączający wysyłanie powiadomień. W efekcie puki ktoś nie sprawdzi NAGIOSa nie dowie się o pro-



Rysunek 2. *Prosta sieć w NAGIOSie*

blemie. Wyłączony jest również *event handler* pozwalający na wykonanie dodatkowych poleceń przy zmianie stanu usługi. W kolejnym pliku możemy dodać definicje hosta. Jednak w tym momencie nie dowiemy niczego poza tym czy jest włączony.

Na nasze szczęście drugi komputer korzysta z systemu Linux. Wystarczy więc utworzyć klucz publiczny oraz przenieść *check_disk* na sprawdzany komputer. Definicja polecenia składa się z nazwy polecenia oraz wykonywanej komendy. Analizując fragment listingu 3

$USER1$ - zmienna równoznaczna z katalogiem zawierającym *check_by_ssh*, /usr/local/nagios/libexec.

$ARG1$ - podobnie jak ARG2-4 przyjmuje parametry przy wywołaniu z definicji usługi są to kolejno numer portu ssh wartość dla ostrzeżenia i wartości krytycznej oraz monitorowany dysk.

W ramach definicji usługi istotne są grupy. Nazwę hosta można zastąpić nazwą grupy co jest jednoznaczne z dodaniem wszystkich należących do niej komputerów. Taka organizacja pozwala zaoszczędzić bardzo dużo czasu przy uruchamianiu monitoringu nowego parametru pracy.

Pominęliśmy jednak możliwość informowania o problemach. Do tego celu potrzebne są definicje użytkownika grupy oraz samej metody powiadamiania. Należy wsiąść pod uwagę ze informacja nie dotrze do celu jeżeli awarii uległ serwer pośredniczący.

Pierwszą możliwością są komunikatory IM. Po instalacji komunikatora mogącego przyjmować adresata i tekst wiadomości w formie polecenia należy zdefiniować polecenie. Nie zależnie od sposobu powiadamiania składać się ono będzie z dwóch elementów, ułożenia tekstu powiadomienia i przekazania go do programu wysyłającego.

Niewątpliwie najbardziej skuteczną metodą powiadamiania jest sms. Jego wykorzystanie wymaga programu działającego podobnie do *sms-pl* oraz konta u operatora sieci. W przypadku poczty najłatwiej posłużyć się konfiguracją *exim4* dla gmail. Posiadając własny serwer można ją łatwo dostosować.

CACTI

CACTI stanowi znakomite narzędzie do uzyskiwania bieżących informacji o wykorzystaniu zasobów jak pamięć procesor czy przepustowość. Problemem w wykorzystaniu tego narzędzia polega na specyfice SNMP. Protokół prostego monitoringu sieci korzysta z mało przyjaznej metody określania cóż chcemy monitorować. W najprostszym wariancji możemy skorzystać z domyślnie dostępnych funkcji jak pamięć routera. Wówczas pozostaje na udostępnić na urządzeniu możliwość monitoringu i wybór protokołu. Jedynie SNTP v3 zapewnia bezpieczeństwo dostępu do urządzenia przez podanie hasła.

Jeżeli chcemy monitorować specyficzne parametry urządzenia SNTP zaczyna gryźć użytkowników. Przede wszystkim system monitorujący musi być wyposażony w *snmp* do którego należy dograć pliki MIB. Zawierają one obiekty możliwe do monitorowania parametry w formie obiektów. Tutaj mamy drugi i największy problem jakim jest ustalenie który obiekt odpowiada za interesujący nas parametr.

Skoro już wiemy jaki obiekt potrzebujemy musimy ustalić numer OID. Pomoże nam tu polecenie *snmptranslate*.

Cyfry które uzyskaliśmy będą potrzebne do stworzenia pliku XML wykorzystywanego przez CACTI do monitoringu.

Od strony samego protokołu, każda z podanych cyfr stanowi wybór opcji na kolejnym poziomie drzewa. Dla przykładu, zgodnie z listingiem 3: 1-iso, 3-org, 6-dod, 1- internet. Z pomocą polecenia *snpwalk* możliwe jest przejście przez wszystkie opcje SNMP, sprawdzając co jest dostępne na naszym urządzeniu. Sprawdza się ono również jako metoda samej kontroli połączenia z urządzeniem.

Podsumowanie

Zarówno NAGIOS jak i LMS nie są jedynymi możliwymi narzędziami do wykorzystania. W połączeniu niezależnie od wielkości sieci umożliwiają szybkie wykrywanie problemów i ich lokalizację. Dzięki temu każda minuta zainwestowana w monitoring zwróci się wielokrotnie w postaci czasu zaoszczędzonego na kontrolowaniu i diagnozowaniu sieci. Mechanizm można rozbudować o CACTI czy mrtg jednak przy specyficznych parametrach urządzeń wymagają znacznie więcej pracy.

Żadna sieć na dłuższą metę nie obejdzie się bez monitoringu. Nawet w niewielkich sieciach składających się z kilku urządzeń warto wykorzystać przedstawione rozwiązania. Dzięki temu nie tylko bieżące zarządzanie stanie się prostsze, przy rozbudowie sieci unikniemy sytuacji w której nikt tak naprawdę nie wie jak jej elementy są połączone.

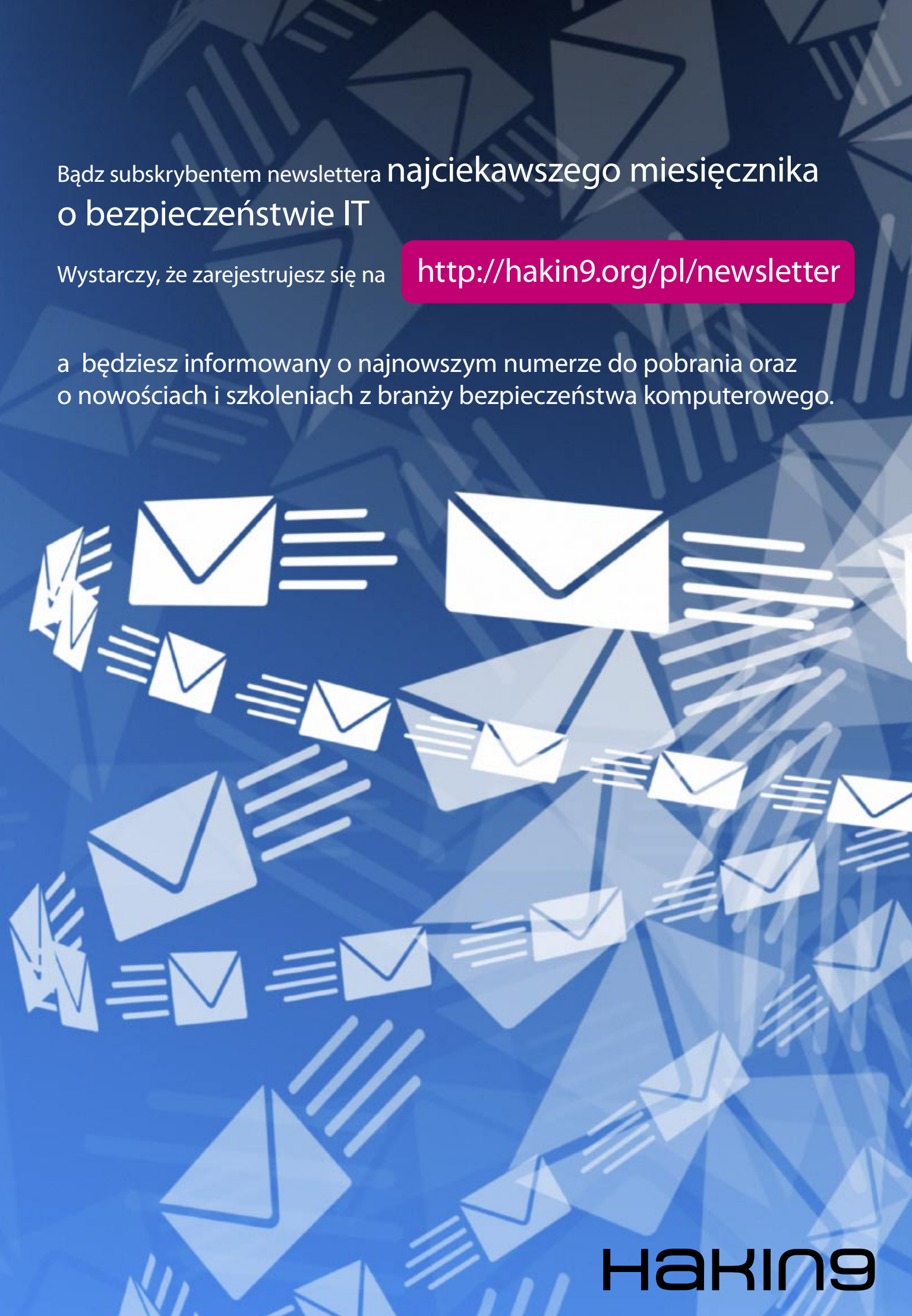
W Sieci

- *http://www.lms.org.pl/* – LMS
- *http://www.nagios.org/* – NAGIOS
- *http://wiki.debian.org/GmailAndExim4* – Exim4 dla Gmail

SYLWESTER ZDANOWSKI

Autor książki „Debian Linux. System operacyjny dla każdego. Pierwsze starcie." Student Europeistyce na Wydziale Humanistycznym Uniwersytetu Szczecińskiego. Obecnie pracuje nad projektem SPRS, http://code.google.com/p/sprs/ .

Kontakt z autorem: sylwesterzdanowski@o2.pl



Twój backup jest bez sensu!

Waldemar Konieczka

Poświęcasz godziny na konfigurację sprawdzanie logów programów backupowych? Wydajesz pieniądze na modernizację serwera? To i tak nie daje pewności, że w razie awarii odzyskasz pliki. Wiesz o tym! A przecież są prostsze sposoby zabezpieczenia najważniejszych dla Twojej firmy danych.

Dowiesz się:
- czym jest backup internetowy
- dla kogo przeznaczone są usługi backupu przez Sieć
- jak trudna jest instalacja tego rozwiązania

Powinieneś wiedzieć:
- czym jest kopia zapasowa
- jakie są podstawowe potrzeby zabezpieczania danych
- co to jest: NAS, storage serwer oraz FTP

Artykułów opisujących sposoby wykonywania kopii zapasowych jest w sieci na prawdę wiele. Podobnie jak oprogramowania, które – za darmo, bądź za pieniądze – pozwoli nam zabezpieczyć pliki przed zniszczeniem.

Ten artykuł nie ma być kolejnym, przedstawiającym założenia konfiguracji mniej lub bardziej chałupniczych systemów archiwizacji danych. Jego celem jest omówienie, w kontekście typowych rozwiązań, nowoczesnych systemów tworzenia kopii zapasowych on-line.

W dzisiejszych czasach, przy co raz większej świadomości zarówno użytkowników jak i administratorów co do konieczności zabezpieczania danych, przychodzi pora na rzetelne rozważanie nie tyle samej idei backupu ile racjonalności i konkurencyjności poszczególnych metod jego wykonywania.

Należy pamiętać o tym, że wybór rozwiązań w tej dziedzinie zależny jest od wielu czynników. Oczywistym jest, że inne metody zastosujemy dla dużych ilości danych czy też dużych sieci, inne zaś dla małych i średnich.

Na szczęście dostępne dzisiaj rozwiązania w zakresie kopii zapasowych pozwalają stworzyć sprawny system kopii zapasowych, którego koszty będą na tyle rozsądne, by mogły je ponieść nie również małe i średnie firmy a nawet indywidualni użytkownicy.

Sposób „na piechotę”

Najprostszym sposobem wykonywania kopii zapasowych jest ręczne tworzenie duplikatów plików. Ten archaiczny sposób nadal wykorzystuje część indywidualnych użytkowników.

Oczywistymi wadami takiego rozwiązania są: całkowicie manualna obsługa, konieczność pamiętania o wykonywaniu kolejnych kopii czy brak możliwości szybkiego i kompleksowego przywrócenia plików.

Z wspomnianych wyżej powodów implementacja tego rachitycznego systemu w skomplikowanych strukturach informatycznych jest praktycznie nie możliwa. Niestety mimo że trudno także wyobrazić sobie jego funkcjonowanie nawet w małych firmach, zdarzają się jeszcze administratorzy wykonujący backupy raczej według wskazań swojego przeczucia niż wedle prawideł IT.

Opcja „sprzęt”

W wielu przypadkach użytkownicy komputerów uciekają się do różnorodnych rozwiązań sprzętowych. Bardzo często powtarzana jest obiegowa opinia, że niedrogą i dobrą metodą zabezpieczania danych jest standard RAID.

Niestety, twierdzenie to nie jest słuszne. Macierze RAID mają służyć przede wszystkim wydajniejszemu i szybszemu dostępowi do danych zgromadzonych na wielu dyskach twardych. Ich konstrukcja i zastosowana nadmiarowość nie służy zabezpieczeniu owych danych przed ich utratą bądź uszkodzeniem.

Oczywiście, RAID w standardach powyżej 0 oferuje wprawdzie pewną nadmiarowość danych, pamiętaj-

my jednak, że jest to tak na prawdę integralny element działania macierzy, forma zabezpieczenia jej przed utratą funkcjonalności w wyniku uszkodzenia fizycznego jednego lub więcej dysków nie zaś system kopii bezpieczeństwa.

Innym sprzętowym rozwiązaniem, o które można się pokusić jest stworzenie serwera backupu dla naszej sieci. Pomysł ten wydaje się dość atrakcyjny. Na rynku dostępnych jest wiele darmowych programów, które potrafią wykonać kopię zapasową, a i opisów ich konfiguracji jest aż nadto.

Nim jednak pochwalimy to wyjście z sytuacji musimy zwrócić się ku kwestii niezwykle ważnej dla backupu – kosztów tworzenia kopii zapasowej. Każdy taki system – oprócz tego, że jest niezbędny powinien być także dokładnie skalkulowany i opłacalny.

Jak owa kalkulacja wygląda? Koszt w miarę stabilnej maszyny mogącej służyć za serwer backupu (przy założeniu, że komputer posiadać będzie dwurdzeniowy procesor low cost, 4 GB RAM oraz 2 dyski x 1TB) to około 2000 zł. Do tego urządzenie to musiałoby synchronizować się dla bezpieczeństwa z bliźniaczą maszyną – kolejne 2000zł. Koszt zużytej rocznie energii dla jednego urządzenia (przy założeniu, że zużyje ono 100 W, a cena za 1kWh to 0,3 zł) wyniósłby nieco ponad 500 zł.

Łącznie to 5000 zł, które, jeśli podzielimy je przez 12 miesięcy roku, dadzą miesięczny wydatek wahający się w granicach 400 zł. I to tylko przy uwzględnieniu podstawowych kosztów, nie doliczając m.in. odpowiednich UPS.

Należy pamiętać przy tym, że o ile można obejść koszty oprogramowania, a zakup serwerów to wydatek (nie licząc amortyzacji) jednorazowy to utrzymanie w ruchu takiego rozwiązania wiąże się z ponoszeniem dodatkowych wydatków. Musimy bowiem doliczyć fundusze na ewentualne awarie oraz bieżącą obsługę całego sprzętu i oprogramowania. Ponadto pozostaje ryzyko, że w przypadku pożaru, zalania czy np. rabunku pomieszczeń, w których znajduje się sieć i nasz backupowy serwer prawdopodobnie, pomimo tworzenia kopii zapasowych, stracimy wszystkie dane.

Rozwiązanie (prawie)kompleksowe

Sposobem wykonywania kopii zapasowych, który zyskuje co raz większą popularność jest backup przez internet. Można wykonywać go na kilka sposobów. Pierwszą z możliwości jest proste rozwiązanie stanowiące połączenie oprogramowania backupowego działającego na komputerze z kontem ftp albo serwerem zdalnym.

Jednym z wariantów takiego rozwiązania jest zastosowanie programu Cobian Backup i konta na serwerze FTP. Tego typu metoda wykonywania kopii zapasowych należy do jednych z tańszych. Oplata za duży hosting waha się w granicach 400 zł rocznie, a sam program często nie kosztuje nic. Niestety, pomimo że, konfiguracja taka bez problemu radzi sobie z regularnym wykonywaniem kopii, to posiada także kilka mankamentów.

Przede wszystkim nie ma możliwości sprawnego zarządzania kopiami bezpieczeństwa. Przesłane przez program pliki po prostu *leżą* w lokalizacji docelowej. Jedyną możliwością zarządzania nimi jest praktycznie opcja ręcznego umieszczenia plików w katalogach tworzonych wg klucza daty czy godziny.

Kolejnym problemem jest brak funkcji przywracania danych, nie mówiąc już o możliwości przywracania danych w zależności od zaistniałej sytuacji.

Ponadto, choć w ustawieniach można m.in. wyłączyć pewne rozszerzenia plików z wykonywania kopii zapasowej, istnieje też wspomniana wyżej opcja rozdzielania plików lub kopii zapasowych za pomocą daty i godziny oraz pakowania każdego pliku do osobnego archiwum, to nie ma możliwości łatwego w obsłudze wersjonowania plików.

Zarządzanie backupem odbywa się niejako na dwóch oddzielnych płaszczyznach – programu wykonującego kopie oraz serwera składowania danych. Każdy z nich trzeba skonfigurować osobno i oddzielnie dbać o jego poprawne funkcjonowanie. Nie ma również możliwości automatycznego, okresowego sprawdzenia poprawności archiwum.

Administratorów systemów Windows zmartwi brak wsparcia dla natywnego backupu Exchange, baz MS SQL nie wspominając o bardziej zaawansowanych aplikacjach.

Kolejnym mankamentem całego rozwiązania jest fakt, że serwer ftp w firmie hostingowej nie jest dedykowany do składowania plików. Część hostingów (jak. np. DreamHost) zastrzega już w momencie zakupu limity w podziale przestrzeni. Przeznaczanie hostingu www wiąże się często z również z innymi ograniczeniami – m.in brakiem indywidualnego mirroru konta czy też wielkością plików, które można wysłać na serwer. To ostatnie obostrzenie wymusza z kolei dzielenie pliku backupu na mniejsze części, co dodatkowo komplikuje sytuację podczas przywracania kopii.

Nie zapominajmy też o tym, że pomimo wdrożonych w firmach hostingowych zabezpieczeń przed awariami – nie są to serwery backupu i nie mają w regulaminie zapisanej nadrzędnej funkcji ochrony danych.

„All in one"

Alternatywą dla wymienionych wyżej rozwiązań są usługi kompleksowego backupu przez internet. Wśród firm oferujących takie usługi przodują te zza oceanu, jednak na szczęście w wyborze dostawcy nie ograniczają nas kraje ani kontynenty.

Nim przejdziemy od opisu najważniejszych funkcjonalności tego typu produktów warto wspomnieć o ich najważniejszej zalecie – bezpieczeństwie stworzonych kopii zapasowych. W przypadku backupu internetowego dane znajdują się w oddzielnych serwerowniach a konta są mirrorowane lub nawet wielokrotnie duplikowane na wielu serwerach, co dodatkowo zapewnia zabezpieczenie na wypadek awarii po stronie serwera.

Ponadto na bezpieczeństwo plików wpływa także fakt, iż nie przechowujemy backupu w tej samej lokalizacji, w której znajdują się zabezpieczane nim maszyny. Wyklucza to ryzyko uszkodzenia danych w wyniku pożaru i innych zdarzeń losowych a także kradzieży. Zwłaszcza, że dane są przechowywane w profesjonalnie przygotowanych pomieszczeniach z zabezpieczeniami przeciwpożarowymi, przeciwwłamaniowymi oraz z pełnym wsparciem na wypadek utraty zasilania.

Po co to wszystko?

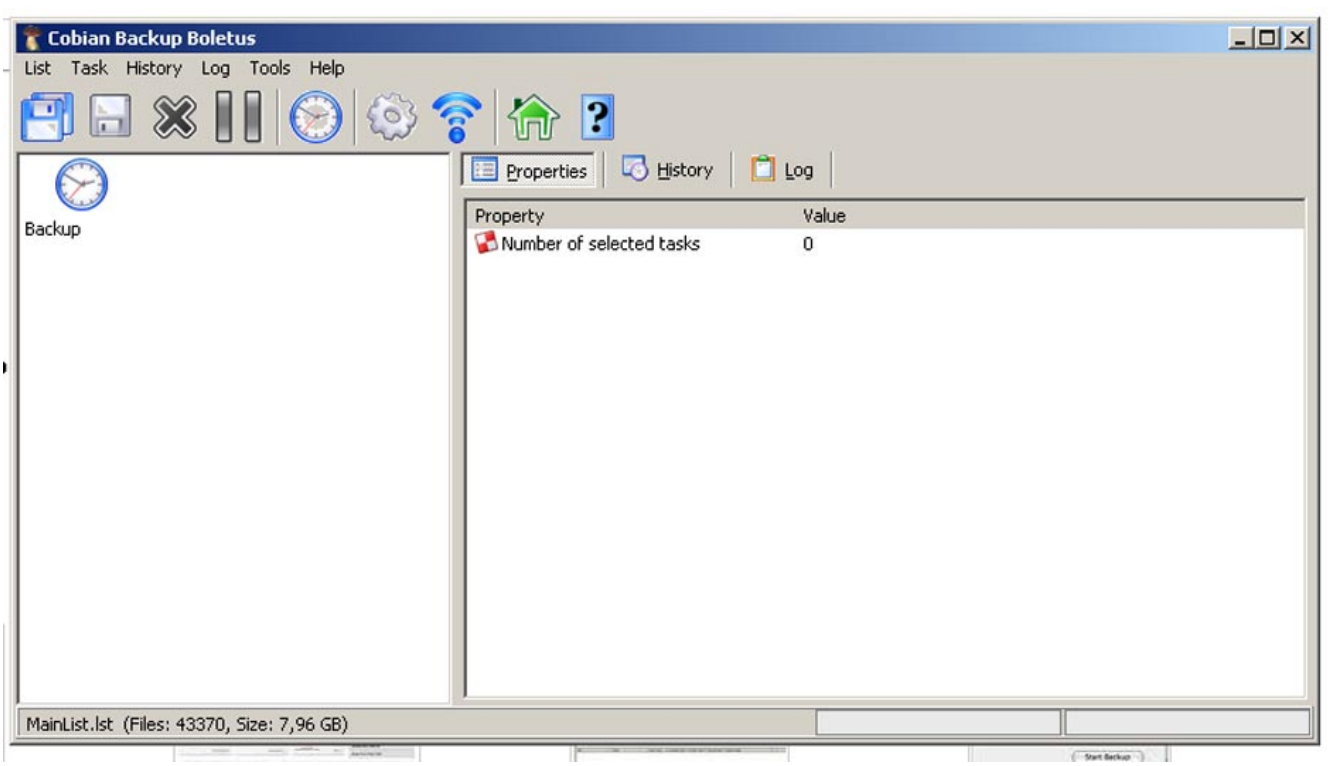
Ideą backupu internetowego jest stworzenie mechanizmu, który pozwoli w wygodny sposób automatycznie wykonać kopie zapasową danych. Dostępne na rynku systemy w opierają się, w gruncie rzeczy, na tym samym mechanizmie, co opisywana wyżej metoda wykorzystująca Cobian Backup. Istotne różnice pojawiają się gdy szczegółowo przeanalizujemy funkcjonalności poszczególnych rozwiązań.

Pierwszą z różnic jest fakt, iż software instalowane na dysku komputera klienta jest oprogramowaniem dedykowanym producenta a same serwery dostosowano konfiguracją do obsługi zadań backupowych i nie trzeba ich dodatkowo ustawiać. Kolejnym ważnym założeniem, szczególnie z punktu widzenia końcowego użytkownika, jest prostota obsługi całego systemu. Owszem, dostępne są niezwykle rozbudowane aplikacje, jednak ideą większości programów klienckich (niezależnie od stopnia skomplikowania systemu) jest bezproblemowość obsługi. W przypadku administracji dużą siecią jest to ważne także dla administratorów, którzy nie muszą długo borykać się z zaznajamianiem użytkowników z interfejsem, jednocześnie mogą pozwolić użytkownikom na pewną swobodę w dobieraniu danych, które chcieliby zarchiwizować.

Potencjał i możliwości

Niezależnie od szczytnych założeń, o przydatności danego rozwiązania decyduje przede wszystkim elastyczność oraz możliwości. W przypadku internetowego backupu większość firm oferuje oprogramowanie i usługi dostosowane wymagań wykonywania kopii zapasowych dla użytkownika końcowego pracującego na możliwie standardowo skonfigurowanej stacji roboczej.

Istnieje oczywiście opcja definiowania danych i obszarów, których kopie zapasowe chcielibyśmy wykonać. Większość dostępnych na rynku programów pozwala dowolnie wybierać pliki i katalogi do backupu oferując przy tym opcję filtrowania zawartości backupowanych danych pod kątem rozszerzeń czy wielkości plików. Dodatkowo niektórzy producenci oferują zestawy predefiniowanych ustawień backupu takich jak ko-



Rysunek 1. *Konsola konfiguracji programu Cobian Backup*

pie zapasowe skrzynek pocztowych programu Outlook czy skrótów pulpitu Windows.

Software radzi sobie bez większych problemów z wykonaniem zadanych mu czynności, niemniej jednak predefiniowane przez producentów ustawienia często mają swoje ograniczenia. Ponadto oferta części dostawców tego typu usług skonfigurowana jest tak, by swoją funkcjonalnością obejmować wykonywanie kopii zapasowych plików nie zaś całych rozwiązań softwerowych takich jak zwirutalizowane maszyny czy systemy bazodanowe.

Przydatne drobiazgi

Większość firm zajmujących się backupem online oddaje do dyspozycji użytkowników kilka ciekawych funkcjonalności, które ułatwiają życie przy backupach.

Pierwszym z tych udogodnień jest dostępne niemal u wszystkich firm wersjonowanie plików. Korzyści wynikające z zastosowania tego mechanizmu są nieocenione. Pozwala on nie tylko przywrócić utracone wersje pliku w przypadku utraty bądź uszkodzenia danych, ale także odtworzyć stan wcześniejszy pliku jeśli w razie błędu użytkownika dokonano w nim niechcianych zmian.

Wersjonowanie odbywa się według różnych zasad. Jedne systemy (jak np. MozyBackup) przyjmują kryterium czasowe przechowując wersje pliku przez określoną liczbę dni (w przypadku Mozy jest to okres 30 dni). Inni dostawcy wybierają ograniczenia ilościowe. W jednych przypadkach, np. w IDrive jest to z góry ustalona liczba 30 wersji, w kolejnych to definiowana przez użytkownika wartość w ramach narzuconych przez dostawcę ograniczeń.

Pozostałe dodatki obejmują przede wszystkim wykonywanie zadań backupowych według harmonogramu oraz dostęp do kopii zapasowych poprzez różnorodne interfejsy.

Zarówno wspomniane wcześniej opcje konfiguracyjne jak i łatwe ustawianie harmonogramu przy prostocie obsługi całości systemu przez użytkownika końcowego wpisują się w zasadę *ustaw i zapomnij.*

Oprogramowanie wykona wszystkie zadania w tle, a użytkownik zostanie poinformowany – najczęściej komunikatem, a u niektórych dostawców mailowo – o wykonanej kopii.

Przywracanie danych

Nie należy zapominać, że najważniejszym zadaniem backupu jest przede wszystkim przywrócenie utraconych danych w razie ich skasowania albo uszkodzenia. W większości programów zadbano o to, by odzyskanie zasobów było równie proste jak stworzenie ich kopii zapasowej.

Zasadniczo w rozwiązaniach tego typu odzyskiwanie odbywa się poprzez to samo oprogramowanie dostarczone przez firmę, która świadczy dla nas usługi, a które wykorzystywane jest do robienia kopii zapasowych. Niektórzy producenci dodatkowo umożliwiają integrację odzyskiwania z eksploratorem Windows, co czyni przywracanie danych jeszcze bardziej intuicyjnym.

Różnice pomiędzy software poszczególnych producentów dotyczą najczęściej ilości opcji, które mamy w przypadku odzyskiwania. Niektórzy z nich proponują przywrócenie jedynie ostatniej wersji plików, inni – jak np. Mozy pozwalają swobodnie przeglądać strukturę plików i przywracać je na dysk

Rysunek 2. *Wizualizacja wirtualnego pulpitu jednego z dostawców internetowego backupu*

Dla wymagających ciekawe rozwiązania

Stosowanie backupu dla użytkowników końcowych to oczywiście ważny element zabezpieczania danych. Niemniej jednak kopie zapasowe często muszą obejmować także rozwiązania specjalistyczne. Wśród dostępnych usług z zakresu internetowego backupu są także takie, które dzięki swojej funkcjonalności pozwalają wykonywać kompleksowe kopie bezpieczeństwa.

Firma Novastor dostarcza kilka gotowych pakietów usług. Są w śród nich m.in. rozwiązania pozwalające backupować serwerowe systemy firmy Microsoft w wersjach 2008 i 2003 (wraz z rozwiązaniami SBS i 2008 R2). Produkt pozwala tworzyć kompletny backup systemu pozwalający na jego późniejsze odtworzenie w razie awarii dysku.

Ponadto otrzymujemy także możliwość backupu systemów Exchange i SQL. Novastor proponuje także dwa ciekawe rozwiązania dotyczące backupu w sieciach komputerowych. Pierwszym z nich jest *Scalable Network Backup Software for all Platforms* będące zestawem klient – serwer umożliwiającym nie tylko wykonywanie kopii na centralnym serwerze (i późniejsze jej wysłanie na serwer firmy backupowej) ale także realizację tej czynności niezależnie od platformy systemowej, która obsługuje stację kliencką. Na plus tego oprogramowania zaliczyć należy obsługę systemów NetWare.

Drugim z wartych uwagi rozwiązań jest scentralizowany backup oparty o storage server. Producent dostarcza oprogramowanie, które umożliwia wykonanie backupu lokalnie i przez internet na wspomnianym serwerze a następnie przesłanie go w inne miejsce. Rozwiązanie to jest o tyle wygodne, że jednym systemem backupu objąć można zarówno sieć lokalną w siedzibie firmy jak i placówki czy też pracujących zdalnie pracowników. Dodatkowo oprogramowanie pozwala wykonywać równolegle backup zapisywany lokalnie na każdej z maszyn.

Backupowy kombajn

Niezwykle rozbudowana ofertę posiada również Ahsay. Jej software działa we wszystkich liczących się systematach operacyjnych (wliczając w to oczywiście darmowego Linuxa oraz MacOs i NetWare). Dodatkowo producent zapewnia, że system ruszy na wszystkich platformach wspierających język Java.

Oprogramowanie może wykonywać kopie zapasowe systemów bazodanowych MS Exchange, MS SQL, Oracle, Lotus Domino/Notes, MySQL oraz m.in. plików Microsoft Outlook / Outlook Express. Ponadto platforma działa także na NASach oraz wspiera Windows Active Directory.

Sam producent daje także wiele możliwości ułatwiających samo wykonywanie kopii zapasowych, które wychodzą poza przyjęty standard. Należą do nich m.in. możliwość ustawienia indywidualnego publicznego URL, z którym będzie można połączyć się w celu wykonania backupu czy też przydzielanie określonej części pasma pojedynczym użytkownikom.

Dobre, ale za ile?

Opcje i możliwości, jakie daje backup internetowy wyglądają niezwykle się obiecująco. Pozostaje odpowiedzieć sobie na pytanie jakie koszty należy ponieść, aby móc zaopatrzyć się w tego typu rozwiązania.

Oczywiście ceny uzależnione są od stopnia skomplikowania usługi, którą chcielibyśmy zamówić. Część firm oferuje nawet darmowe produkty, jednak w większości ich oferta skierowana jest do użytkowników domowych.

Propozycja Mozy dla firm obejmuje opłatę miesięczną na poziomie ok. 4 USD oraz 0,5 USD za każdy GB danych. Podobny możliwościami Ibackup będzie kosztował nas minimum 10 USD za pakiet 10 GB danych. Mamy więc tutaj do czynienia z cenami oscylującymi ok. 50 zł miesięcznie przy backupie wielkości 10 GB.

Polscy dostawcy oferujący usługi backupowe sprzedają swoje produkty w zróżnicowanych cenach. Dla przykładu Msejf w ofercie dla firm za licencję na jeden komputer każe sobie płacić 180 zł netto rocznie dając przy tym 2 GB miejsca na kopie zapasowe. Za dodatkowe miejsce i więcej komputerów trzeba zapłacić dodatkowo.

Tabela 1. *Stosunek funkcjonalności poszczególnych rozwiązań do ich skomplikowania i kosztów*

Nazwa	Funkcjonalność	Prostota	Koszt	Elastyczność
serwer backup	wysoka	niska	wysoki	wysoka
Cobian + FTP	niska	niska	niski	niska
MozyBackup	średnia	wysoka	niski	Średnia
Ahsay	wysoka	niska	średni	wysoka
Novastor	wysoka	niska	średni	wysoka

najtańsze rozwiązanie
rozwiązanie najbardziej funkcjonalne w stosunku do kosztów
rozwiązanie najbardziej funkcjonalne

Rozwiązania profesjonalne są droższe. Ich ceny wahają się na poziomie od kilkunastu do kilkudziesięciu dolarów miesięcznie i często również zależne są od wielkości plików w backupie.

To przecież proste!

Internetowy backup sprawia, że zabezpieczenie naszych danych staje się co raz prostsze. Nieskomplikowana instalacja oraz łatwość konfiguracja prowadzą do tego, że nawet w mniejszych firmach można bez kłopotów wdrażać tego typu rozwiązania.

Bez wątpienia pamiętać należy o tym, że stopień skomplikowania implementacji zależy od tego co chcemy zabezpieczać a także od tego, jak duża i skomplikowana jest np. struktura sieciowa, w której pracują poszczególne jednostki.

Uczciwie przyznać należy, że prostota backupów internetowych jest także pewną ich wadą. Trudno zaprzeczyć, że w niektórych przypadkach niewielkie, zautomatyzowane oprogramowanie nie zastąpi nam profesjonalnie skonfigurowanego i dobrze wyposażonego serwera kopii zapasowych nadzorowanego przez doświadczonego informatyka.

Pamiętajmy jednak, że nie można pomijać kwestii racjonalności wydatków. Firmy oferujące backup przez internet dają nam dobrze i bezpieczne rozwiązanie w rozsądnej cenie, która stanowi ułamek kosztów utrzymania dedykowanego rozwiązania. Dlatego dla większości firm i administratorów te rozwiązania są wystarczająco dobrze zaprojektowane, bezpieczne i mało awaryjne, a przede wszystkim tańsze w zakupie.

Każdy krok jest ważny

Każde zastosowane rozwiązanie backupowe zwiększa bezpieczeństwo danych. Każdy kolejny skopiowany plik to krok do uniknięcia gigantycznych kłopotów, a w wielu przypadkach wymiernych strat materialnych.

Stosowanie internetowego systemu kopii zapasowych to nieskomplikowany sposób na odpowiedzialne podejście do sprawy bezpieczeństwa naszych plików.

Choć rozwiązanie to z pewnością nie zadowoli profesjonalnych administratorów i wielkich firm, to znacznej większości użytkowników sprzętu komputerowego może dać wewnętrzny spokój i oszczędzić wielu nieprzyjemnych sytuacji.

Backup przez sieć jest rozwiązaniem idealnym dla małych i średnich przedsiębiorstw. Sprawdzi się wszędzie tam, gdzie nie nie ma potrzeby (bądź technicznych i finansowych możliwości) zatrudniania informatyka i budowania indywidualnego systemu kopii zapasowych.

Wszyscy ci, którzy posiadają w firmach jedną lub kilka stacji roboczych nie korzystających ze skomplikowanych systemów wirtualizacji czy centralizacji pracy poprzez rozwiązania oparte o serwery czy bazy danych mogą z powodzeniem stosować produkty Mozy czy Ibackup – proste w konfiguracji i bezproblemowe w działaniu. Oferują one rozsądny zakres możliwości konfiguracyjnych, które jednocześnie nie komplikują obsługi całego systemu.

Administratorzy nieco bardziej rozbudowanych i skomplikowanych sieci muszą skorzystać z oferty usługodawców takich jak Ahsay czy Novastor. Ich funkcjonalność jest na tyle wysoka, że poradzą sobie z większością backupowych zadań, nawet tych dość specyficznych jak kopie bezpieczeństwa serwerowych systemów Microsoftu. Niestety, stopień skomplikowania konfiguracji tych rozwiązań sprawia, że nie poradzą sobie z nimi niedoświadczeni użytkownicy.

Niezależnie od wybranej opcji czy dostawcy usług internetowy backup przyniesie firmie sporo korzyści i umożliwi jej spokojne oraz stabilne funkcjonowanie przy rozsądnych nakładach finansowych.

WALDEMAR KONIECZKA

Autor jest Głównym Specjalistą ds. Informatycznych w firmie AKTE z Poznania od 10-ciu lat.
Na co dzień łączy wiedzę teoretyczną z praktycznym zastosowaniem wiedzy z zakresu wdrożeń systemów IT.
Autor na łamach tego pisma dzieli się swoim wieloletnim doświadczeniem teoretycznym i praktycznym, zdradza tajniki wiedzy informatycznej oraz każe nam się przyjrzeć na co zwrócić szczególną uwagę aby nasza praca w IT była bardziej świadoma, a co za tym idzie bardziej komfortowa.

Firma Akte świadczy usługi Outsourcingu IT oraz Profesjonalnego Odzyskiwania i Archiwizacji Danych komputerowych.
W ramach działań operacyjnych firma wdraża systemy archiwizacji i bezpieczeństwa danych, gdzie autor nadzoruje projekty od strony informatyczno-biznesowej.
Po godzinach gra na gitarze w zespole rockowym.
Kontakt z autorem: akte@akte.com.pl
Strona autora: http://www.akte.com.pl

Jak uchronić się przed potopem cyfrowych danych

Stanisław Rejowski, Dyrektor Działu Produkcji Serwerów Actina Solar w ACTION S.A.

Tempo przyrostu ilości cyfrowych danych zwiększa się z roku na rok. Eksperci z International Data Corporation (IDC) prognozują, że do 2020 roku objętość cyfrowego wszechświata zwiększy się 67-krotnie, a w samym tylko 2010 roku świat zaleje 1,2 zettabajtów informacji elektronicznych. Co zrobić, by nie utonąć w potopie elektronicznych informacji oraz jak efektywnie przechowywać dane?

Dowiesz się:

- Jakie jest tempo wzrostu cyfrowego wszechświata
- Jakie są główne problemy związane z przechowywaniem danych
- czym jest duplikacja, deduplikacja, archiwizacja, replikacja, cloud storage oraz backup

Powinieneś wiedzieć:

- podstawowa wiedza z zakresu przechowywania danych

Cyfrowy wszechświat stale rozszerza swoje granice, a specjaliści alarmują, że w kolejnych latach możemy spodziewać się jeszcze większego tempa przyrostu informacji cyfrowych. Wg prognoz IDC 2020 roku ilość danych cyfrowych wzrośnie aż 67-krotnie. Co ciekawe, jedynie połowa cyfrowego wszechświata jest wynikiem bezpośrednich działań ludzi – wysyłania maili, zakładania stron internetowych czy też rozmów telefonicznych w technologii VoIP. Drugą połowę stanowi tzw. cyfrowy cień – pozostałość, ślad podejmowanych przez nas działań zapisany w rejestrach rozmów, historii przeglądanych stron czy nagraniach kamer monitoringu.

Za tak dynamicznym przyrostem ilości danych stoi cały szereg zjawisk - od rosnącej popularności urządzeń mobilnych, dynamicznej ekspansji internetu i cyfryzacji coraz większej liczby dziedzin życia, po procesy globalizacji, ciągły rozwój społeczeństwa informacyjnego oraz gospodarki opartej na wiedzy, w której informacje – obok ludzi – stanowią najcenniejsze aktywa firm. Jednak tak szybki przyrost ilości cyfrowych danych niesie ze sobą także pewne wyzwania - chcąc uniknąć sytuacji, w której wielkość cyfrowego świata przewyższy możliwości nośników informacji, nieodzowne stanie się wprowadzenie odpowiedniej polityki zarządzania danymi, zarówno w wymiarze korporacyjnym, jak i w przypadku użytkowników domowych.

Nieaktywne dane zabierają miejsce

Istotnym problemem współczesnego storage'u jest fakt, iż znaczna część przechowywanych danych - niezależnie czy na firmowych serwerach, czy na dyskach twardych komputerów domowych – to informacje nieaktywne, niewykorzystywane przez użytkowników. Dobrym rozwiązaniem problemu niekontrolowanego przyrostu zbędnych plików jest wdrożenie systemu zarządzania danymi – profesjonalnego w przypadku firmy, racjonalnej polityki optymalizującej wykorzystanie zasobów dyskowych w odniesieniu do użytkowników domowych.

W przypadku przedsiębiorstw pierwszym i podstawowym krokiem wiodącym do optymalizacji storage'u jest określenie kluczowych z punktu widzenia firmy danych i wymagań odnośnie ich przechowywania (np. z uwagi na uregulowania prawne czy poziom poufności). Istotne jest także zapewnienie niezakłóconej pracy aplikacji roboczych oraz wdrożenie inteligentnych rozwiązań zapisu i magazynowania danych. Ważną rolę odgrywają także technologie pozwalające na maksymalne wykorzystanie pojemności posiadanych dysków – mowa tu m.in. o technologiach deduplikacji i archiwizacji. Nieodzowne z punktu widzenia przechowywania danych w firmach są również procedury backupu oraz zabezpieczania danych.

Powyższe zasady znajdują swoje zastosowanie także w przypadku magazynowania danych przez użytkowników domowych, na komputerach których również znajdują się cenne cyfrowe dane.

Nieuzasadniona duplikacja

Kolejną kwestią, z którą przyjdzie nam się zmierzyć w obliczu dynamicznego przyrostu danych cyfrowych jest nieuzasadniona duplikacja. O ile planowa i celowa replikacja danych jest jednym z podstawowych narzędzi backupu, o tyle niczym nieuzasadniona duplikacja stanowi istotny problem zarówno dla centrów danych, jak i dla pojedynczych serwerów storage'owych czy dysków twardych. Duplikacja to zjawisko występowania jednej lub wielu kopii tej samej porcji danych w obrębie jednego nośnika lub na różnych nośnikach. Z duplikacją danych mamy do czynienia nie tylko w firmach, ale także i na prywatnych, domowych komputerach. Często ten sam plik, zdjęcie, utwór czy film występują na kilku komputerach, urządzeniach cyfrowych czy nośnikach w obrębie jednego gospodarstwa domowego. Dobrym rozwiązaniem, eliminującym multiplikowane porcje danych, jest stworzenie małego, domowego centrum danych. Za centrum tego typu posłużyć może niewielki serwer storage'owy tzw. Home Server, do którego dostęp uzyskają wszyscy domownicy. Wyposażony w wyjście HDMI może on stanowić centrum multimedialne. Podobne funkcje może pełnić dysk sieciowy NAS.

NAS (Network Attached Storage) to technologia umożliwiająca podłączenie zasobów zgromadzonych na dyskach bezpośrednio do sieci komputerowej. Jest to macierz dyskowa (lub pojedynczy dysk twardy) podłączony bezpośrednio do sieci lokalnej. W ramach infrastruktury serwer-klient pełni rolę serwera, posiada procesor i okrojoną wersję systemu operacyjnego. Zasoby serwerów NAS są udostępniane uprawnionym użytkownikom. Rozwiązanie to umożliwia łatwy dostęp do danych zmagazynowanych w jednym miejscu z różnych punktów sieci. Silną stroną dysków sieciowych NAS jest możliwość stosowania ich w heterogenicznych środowiskach sieciowych bazujących na różnych rozwiązaniach klienckich, dzięki czemu dane dostępne są niezależnie od zainstalowanego systemu operacyjnego.

Ponadto publiczny adres IP czy też usługi VPN oferują użytkownikowi dostęp do zgromadzonych zasobów z dowolnego miejsca, także za pomocą urządzeń mobilnych.

Rysunek 1. *Komputer Actina z monitorem LG*

Odpowiedzią na nieuzasadnioną duplikację są procedury deduplikacji. Polegają one na eliminacji powtarzających się informacji poprzez zastąpienie kopii odnośnikami do oryginalnej, źródłowej porcji danych. Deduplikacja to operacja wykonywana automatycznie przez oprogramowanie, najczęściej w macierzy dyskowej, dzięki której możliwe jest wyeliminowanie powtarzających się danych (ich duplikatów). Umożliwia ona przechowywanie jedynie unikatowych plików lub bloków z danymi, generując tym samym znaczne oszczędności przestrzeni dyskowej.

Wyróżnia się dwa rodzaje deduplikacji: na poziomie systemu plików i na poziomie bloków dyskowych. Deduplikacja na poziomie bloków dyskowych gwarantuje lepsze rezultaty, gdyż nie zależy od ilości lub rodzaju plików ani od systemu operacyjnego, na którym bazuje dany system informatyczny.

Funkcje deduplikacji oferowane są przez niektóre aplikacje odpowiedzialne za backup. Oprogramowanie to często pozwala ponadto na jednoczesną na kompresję danych, co oczywiście daje dodatkowy zysk. Przy szybko rosnącej ilości przechowywanych informacji funkcje te zaczynają nabierać szczególnie istotnego znaczenia - większa ilość przechowywanych danych archiwizuje się dłużej, a czasu na backup pozostaje tyle samo. Tym bardziej, że liczne regulacje coraz częściej zmuszają firmy do archiwizowania danych nawet przez dziesiątki lat.

Archiwizacja

Gdy mamy do czynienia z danymi, z których na co dzień nie korzystamy, ale z różnych względów uznajemy je za cenne i nie chcemy się z nimi rozstać, z pomocą przychodzą nam procedury archiwizacji. Statystyki pokazują, że z 80% stworzonych danych firma nigdy już nie korzysta bądź korzysta bardzo rzadko. Archiwizacja umożliwia przeniesienie owych danych na wolniejsze, a przez to tańsze od dysków twardych nośniki. Archiwizacja to proces przenoszenia danych z systemów komputerowych na inne nośniki w celu zredukowania ich ilości. Archiwizacja może przybierać formę kopii analogowych (wydruki dokumentów), kopiowania danych na nośniki wymienne (np. płyty CD-R, płyty DVD-R czy Blu-ray), kopiowania plików na inny dysk tego samego komputera (np. RAID), kopiowania na taśmę magnetyczną, kopiowania na inny komputer (np. kopia zwierciadlana, serwer plików) lub zautomatyzowanej archiwizacji online (np. Przeklej.pl, Plikus.pl).

Standardowe sposoby archiwizacji to metody ręczne i półautomatyczne, wymagające od użytkownika nakładów pracy, czasu i pieniędzy. Archiwizacja w obrębie sieci lokalnej wiąże się z koniecznością zakupu oraz regularnego unowocześniania sprzętu, ponoszenia kosztów związanych z eksploatacją infrastruktury. Ponadto wymaga wygospodarowania bezpiecznej przestrzeni na

składowanie zapełnionych nośników, co niesie ze sobą kolejne koszty (instalacja alarmu, zatrudnienie dodatkowej obsługi). Doskonałe rozwiązanie archiwizacyjne powinno umożliwiać nieprzerwany dostęp do zarchwizowanych danych, stąd najczęściej stosowanymi rozwiązaniami archiwizacyjnymi są dyski twarde, dyski zewnętrzne i serwery storage'owe czy biblioteki taśmowe. Do celów archiwizacji nadają się także nośniki optyczne (DVD, Blu-ray) – ze względu na swoją stosunkowo dużą pojemność i relatywnie niski koszt stanowią dobre rozwiązanie zarówno dla firm, jak i dla zaawansowanych użytkowników domowych. Niestety, czytniki tego typu utrudniają dostęp do zarchiwizowanych danych oraz pozostawiają wiele do życzenia, jeśli chodzi o bezpieczeństwo zarchiwizowanych informacji z uwagi na ograniczone możliwości kontroli dostępu.

Coraz więcej firm decyduje się na usługi archiwizacji online. Outsourcing procedur archiwizacyjnych to wygodne rozwiązanie, zdejmujące z barków firmy lub użytkownika konieczność inwestowania w urządzenia i infrastrukturę magazynowania danych. Sceptycy archiwizacji onlne za jej główną słabość uznają długi - zależny od przepustowości łącza internetowego- upload danych.

Jak przechowywać, to z głową… w chmurze

Coraz większą popularnością cieszą się rozwiązania tzw. cloud storage – przechowywania danych w chmurze. Cloud storage stanowi model przechowywania danych online, w którym informacje magazynowane są na wirtualnych serwerach hostowanych przez dostawców usług. W modelu tym użytkownik płaci tylko za powierzchnię dyskową, z której rzeczywiście korzysta oraz za faktycznie wykorzystywane zasoby. Przechowywanie danych w modelu cloud niesie ze sobą liczne korzyści. Cloud storage nie wymaga od użytkowników ponoszenia nakładów na lokalną infrastrukturę fizyczną, zdejmuje z niego także ciężar przeprowadzania backupów, replikacji, konserwacji urządzeń. Z drugiej jednak strony istnieją pewne wątpliwości odnośnie bezpieczeństwa danych (szczególnie tych wrażliwych) magazynowanych w chmurze. Specjaliści zwracają także uwagę na fakt, iż dostęp do cyfrowych informacji przechowywanych w ten sposób jest uzależniony od dostępności i jakości połączenia internetowego.

Rysunek 2. *Wnętrze serwera*

Specjaliści z IDC prognozują, że technologia cloud storage będzie zyskiwać coraz większy udział w rynku. Obecnie szeroko pojęte usługi cloud storage stanowią 9% rynku rozwiązań cloud computing, którego całkowita wartość szacowana jest na ok. 17,4 mld USD. W 2013 wartość rynku cloud storage ma wzrosnąć do 6,2 mld USD. Polscy użytkownicy nadal z ograniczonym zaufaniem podchodzą do technologii cloud storage – dość niechętnie godzą się z myślą, że ich cenne dane przechowywane będą z dala od nich.

Bezpieczeństwo danych - backup

Wszystkie technologie przechowywania danych mogą być narzędziem backupu. Kopie zapasowe można przechowywać w chmurze, na płytach, na dyskach twardych, dyskach zewnętrznych, na serwerach własnych lub wynajętych od zewnętrznej firmy. Tworzenie zapasowych kopii bezpieczeństwa ma na celu umożliwienie odtworzenia danych utraconych w przypadku ich utraty bądź uszkodzenia. Choć pierwotnie termin backup odwoływał się do zautomatyzowanych i usystematyzowanych form tworzenia kopii zapasowych, obecnie pojęciem tym określamy wszelkie, nawet amatorskie, sposoby sporządzania kopii bezpieczeństwa plików. Kopie bezpieczeństwa najczęściej przechowywane są na dyskach HDD lub nośnikach taśmowych.

Wiele programów, np. edytorów tekstów, automatycznie tworzy kopie opracowywanych plików, przez co umożliwia odtworzenie ich zawartości w sytuacji awarii komputera lub niespodziewanego odcięcia dopływu prądu. Mechanizmy backupu są standardem w przypadku środowisk serwerowych, gdzie od bezpieczeństwa danych zależeć może funkcjonowanie całej firmy.

Wyróżniamy kilka rodzajów backupu: całościowy, przyrostowy, różnicowy, lokalny oraz sieciowy. Każdy z nich ma swoje mocne i słabe strony, decyzja o wybranej metodzie tworzenia kopii zapasowych zawsze należy do użytkownika.

Backupowi całościowemu (full backup) poddawane są wszystkie zapisane na nośniku dane, a bit „archive" plików ustawiany jest w stan „0". Backup całościowy zdecydowanie ułatwia wyszukiwanie danych – wszystkie znajdują się na jednym nośniku, a w przypadku awarii odtworzenie systemu nie zajmuje zbyt wiele czasu. Z drugiej jednak strony full backup nie pozwala na efektywne wykorzystanie nośników – permanentnie tworzone są kopie bezpieczeństwa rzadko modyfikowanych danych. Kolejną komplikacją może być długi czas wykonywania operacji.

Drugą techniką tworzenia kopii bezpieczeństwa jest tzw. backup przyrostowy (incremental backup). W przeciwieństwie do swojego całościowego odpowiednika, backup przyrostowy dokonuje zapisu jedynie nowopowstałych danych lub tych, które uległy zmianie od czasu wcześniejszego backupu. Backupowane są pliki, które bit „archive" mają ustawione w stan „1", a po tej operacji bit „archive" jest przestawiany w stan „0". Silną stroną backupu przyrostowego jest krótki czas jego przeprowadzania oraz efektywne wykorzystanie nośników. Backup przyrostowy nie jest jednak metodą idealną, poza długim czasem odtwarzania systemu jego podstawową słabością jest trudność wyszukiwania danych - do odnalezienia zbioru są potrzebne wszystkie nośniki z backupami przyrostowymi oraz ostatni nośnik z backupem całościowym.

Kolejnym sposobem tworzenia kopii zapasowych jest backup różnicowy (differential backup), w którym zapisywane są te dane, które uległy zmianie od czasu ostatniego całościowego lub przyrostowego backupu. Backupowi poddawane są pliki, które bit „archive" mają ustawione w stan „1". Po tej operacji nie ulega on zmianie. Ten typ backupu umożliwia łatwe wyszukiwanie potrzebnych w danej chwili danych, odtworzenie systemu przeprowadza się stosunkowo szybko. Również sama procedura backupu trwa krócej niż w przypadku backupu całościowego. Także i ten sposób tworzenia kopii zapasowych nie ustrzegł się słabości, którymi są nieefektywne wykorzystanie nośników oraz nadmiarowość backupów – nawet te dane które nie uległy zmianie, są cały czas backupowane. Backup różnicowy ustępuje przyrostowemu pod względem czasu wykonywania operacji.

Spektrum technik tworzenia kopii bezpieczeństwa zamykają backup lokalny oraz backup sieciowy. Backup lokalny to rozwiązanie proste w instalacji i konfiguracji. Gwarantuje szybki transfer danych i krótki czas trwania procedury backupu. Niestety wymaga dość dużego udziału ze strony użytkownika, co zwiększa prawdopodobieństwo wystąpienia błędu ludzkiego. Jest ponadto stosunkowo drogi.

Backup sieciowy z kolei umożliwia centralne zarządzanie, intuicyjną automatyzację. Jest tańszy w administrowaniu aniżeli jego lokalny odpowiednik. Słabością backupu sieciowego jest duże obciążenie sieci, wolny transfer danych oraz długi czas potrzebny do przeprowadzenia backupu.

Ważne przy tworzeniu kopii bezpieczeństwa jest, by pierwotne dane i ich zbackupowane odpowiedniki nie znajdowały się w tym samym miejscu. Podczas ostatniej powodzi wiele firm na własnej skórze odczuło jak bolesnym błędem może być lokalizacja wszystkich danych w jednym miejscu. Podobnie w przypadku użytkowników domowych – przechowywanie kopii bezpieczeństwa danych razem lub w pobliżu danych pierwotnych odbiera sens backupowi.

Niestety nadal bardzo często – potwierdzając słowa przysłowia „mądry Polak po szkodzie" – o tym, jak istotny jest backup dowiadujemy się dopiero po utracie danych. Okazuje się wtedy, że przechowywane na komputerze dane mają w istocie większą wartość niż samo urządzenie.

Replikacja

Jednym z narzędzi backupu jest replikacja - proces powielania danych, tworzenia ich kopii na różnych nośnikach, np. serwerach. Wyróżniamy trzy podstawowe rodzaje replikacji danych: migawkową, transakcyjną oraz dwukierunkową. Replikacja migawkowa polega na powieleniu i rozprowadzeniu na różnych nośnikach danych zapisanych w określonym momencie. Ten typ replikacji znajduje swoje zastosowanie głównie w przypadku danych poddawanych rzadkim, ale znacznym modyfikacjom. Za niedogodność możemy uznać fakt, iż dane zapisane pomiędzy kolejnymi migawkami nie są replikowane. W przypadku replikacji transakcyjnej dane rozprowadzane są na podstawie logów transakcji tylko wtedy, gdy odbywa się synchronizacja. Replikacja dwukierunkowa polega zaś na dwukierunkowej wymianie danych – od serwera głównego, jak i od klientów. Słabością tego rodzaju replikacji jest niebezpieczeństwo zaistnienia konfliktu w czasie synchronizacji.

Podsumowanie

Prawie 500 miliardów gigabajtów informacji cyfrowych wygenerowanych w 2008 roku oraz prognoza mówiąca o 1,2 zettabajta danych, które mają powstać w 2010 roku, nie pozostawiają złudzeń. Tak dynamiczny przyrost ilości informacji elektronicznych wymusi na firmach i użytkownikach domowych zmianę podejścia do kwestii storage'u. Zamiast inwestować w kolejne nośniki, coraz więcej osób zacznie zadawać sobie pytanie, co zrobić, by lepiej wykorzystać te, które już posiadają.

STANISŁAW REJOWSKI

Ukończył Automatykę i Metrologię na Wydziale Elektrotechniki, Automatyki, Informatyki i Elektroniki Akademii Górniczo-Hutniczej w Krakowie. Od 18 lat związany z branżą IT. Specjalizuje się w problematyce serwerowej. Przez ponad 13 lat pracował w firmie OPTIMUS na stanowiskach: Technik w Dziale Sieci Komputerowych, pracownik w Dziale Integracji Systemów, Dyrektor Centrum Technologicznego Systemów Operacyjnych, Kierownik Serwisu Serwerów, Manager ds. Technologii Serwerowych w Biznes Unit Serwery, Kierownik Produkcji i Serwisu Serwerów oraz Dyrektor Działu Badań i Rozwoju. Od 2006 roku pełni obowiązki Dyrektora Działu Produkcji Serwerów Actina Solar w firmie ACTION S.A.

Neurony w komputerze

Wojciech Terlikowski

Sztuczne sieci neuronowe posiadają wiele zalet, które pozwoliły im stać się jedną z najpopularniejszych metod obliczeniowych sztucznej inteligencji. Znalazły zastosowanie w rozwiązywaniu zadań klasyfikacji, aproksymacji jak i predykcji. Artykuł przedstawia podstawowe zagadnienia związane z budową i uczeniem sieci. W dalszej części zaproponowano użycie sieci neuronowej jako systemu zwiększającego bezpieczeństwo sieci komputerowej.

Dowiesz się:

- Co to są sieci neuronowe, jak się je buduje i które ich właściwości decydują o szrokich możliwościach wykorzystania m.in. w zagadnieniach związanych z bezpieczeństwem.

Powinieneś wiedzieć:

- Znać pojęcia i ogólne zagadnienia z zakresu optymalizacji globalnej

Możliwości ludzkiego mózgu od niepamiętnych czasów fascynowały badaczy. Mimo, że pod względem szybkości i dokładności obliczeń arytmetycznych współczesne komputery uzyskują znakomite wyniki to możliwości układu nerwowego człowieka w zakresie uczenia, rozpoznawania i klasyfikacji pozostają wciąż niedoścignionym wzorem. Wzór ten z czasem stał się również inspiracją dla badaczy i inżynierów chcących tworzyć maszyny coraz lepsze i bardziej uniwersalne. W celu lepszego zrozumienia budowy i zasady działania sztucznych sieci neuronowych warto wiedzieć jak działa ich naturalny pierwowzór.

Układ nerwowy człowieka

Badania badania wykazują, że mózg człowieka składa się przede wszystkim z miliardów elementarnych komórek nerwowych – neuronów, połączonych w skomplikowaną sieć. Ocenia się, iż każdy z neuronów ma średnio kilka tysięcy połączeń z innymi. Daje to łącznie około stu trylionów połączeń w całym mózgu. Komórki nerwowe mogą przenosiśc i przetwarzać złożone sygnały elektrochemiczne. Impulsy te przyjmowane są za pośrednictwem wielu wejść informacyjnych tzw. dendrytów i scalane w ciele tzw. perikarionie. Informacja wyjściowa jest przekazywana do innych komórek za pośrednictwem pojedynczego włókna - aksonu i rozgałęzionej struktury wyjściowej - telodendronu. Akson jednego neuronu łączy się z dendrytami kolejnych za pomocą synapsy. Jeśli w wyniku pobudzenia przez synapsę komórka nerwowa przejdzie do stanu aktywnego, wysyła przez wyjście infomacyjne sygnał o charakterystycznych parametrach - kształcie, amplitudzie i czasie trwania. Za pośrednicwtem innych synaps impuls taki przekazywany jest do kolejnych neuronów i może spowodować ich aktywację. Przejście komórki nerwowej w stan aktywny następuje, gdy łączny sygnał, który do niej dotarł z wszystkich dendrytów przekroczy pewien poziom progowy. Należy zauważyć, że waga sygnałów z różnych wejść może być różna.

Taki, dość uproszczony model układu nerwowego człowieka, stał się inspiracją do zaproponowania modelu matematycznego zwanego sztuczną siecią neuronową (ang. *artificial neural network, ANN*).

Budowa i działanie sztucznej sieci neuronowej

Kiedy mówi się o sztucznych sieciach neuronowych, zwanych w uproszczeniu sieciami neuronowymi, należy pamiętać, że nie są one próbą wiernego odwzorowania procesów zachodzących w układzie nerwowym. Podobnie jak w przypadku innych metod wzorowanych na procesach zachodzących w przyrodzie np. algorytmów ewolucyjnych inspiracja biologiczna polega przede wszystkim na przyjęciu odpowiedniego nazewnictwa oraz ogólnego schematu działania dla zaawansowanej metody matematycznej.

Najprostszą sztuczną siecią neuronową jest perceptron. Może się on składać nawet z jednego neuronu,

posiadającego wiele wejść x_n i jedno wyjście *y*. Każde z wejść neuronu powiązane jest z wagą synaptyczną w_n, a dodatkowo istnieje również waga w_0 zwana wartością progową, nie powiązaną z żadnym z wejść. O kształcie i poziomie sygnału wejściowego decyduje funkcja aktywacji *f*. Stosuje się bardzo różne funkcje aktywacji, do najpopularniejszych należą: signum, tangens hiperboliczny, arcus sinus czy funkcja liniowa.

Pojedynczy neuron można opisać za pomocą równania $y=f(s)$ gdzie *s* jest sumą sygnałów ze wszystkich wejść z odpowienidmi wagami. $s=\Sigma_i(x_i w_i+w_0)$

Jak widać sposób działania prostej sieci neuronowej jest łatwy do zrozumienia. Okazuje się, że już tak nieskomplikowana sieć jak perceptron złożony z jednego neuronu posiada dobre właściwości w zakresie klasyfikacji.

Na podstawie powyższych równań widać, że o tym jak dobrze model realizuje swoje zadanie decyduje funkcja aktywacji oraz wartości wag. W bardziej złożonych przypadkach znaczenie będzie miała również liczba neuronów i ich rozmieszczenie w warstwach. Funkcja aktywacji jest najczęściej dobierana arbitralnie dla całej sieci, więc cała "pamięć" znajduje się w wagach. Dlatego proces uczenia perceptronu polega na wyznaczeniu odpowiednich wartości wag.

Uczenie perceptronu

Do uczenia perceptronu stosuje się uczenie z nadzorem (ang. *supervised learning*) zwane również uczeniem z nauczycielem. Algorytmy z tej grupy wymagają znajomości prawidłowej kategorii dla każdego z przykładów. Kolejne elementy zbioru uczącego podawane są na wejścia sieci. Jeśli na wyjściu przykład zostanie nieprawidłowo zaklasyfikowany to algorytm modyfikowany jest tab by zmniejszyć prawdopodobieństwo popełnienia kolejnych błędów.

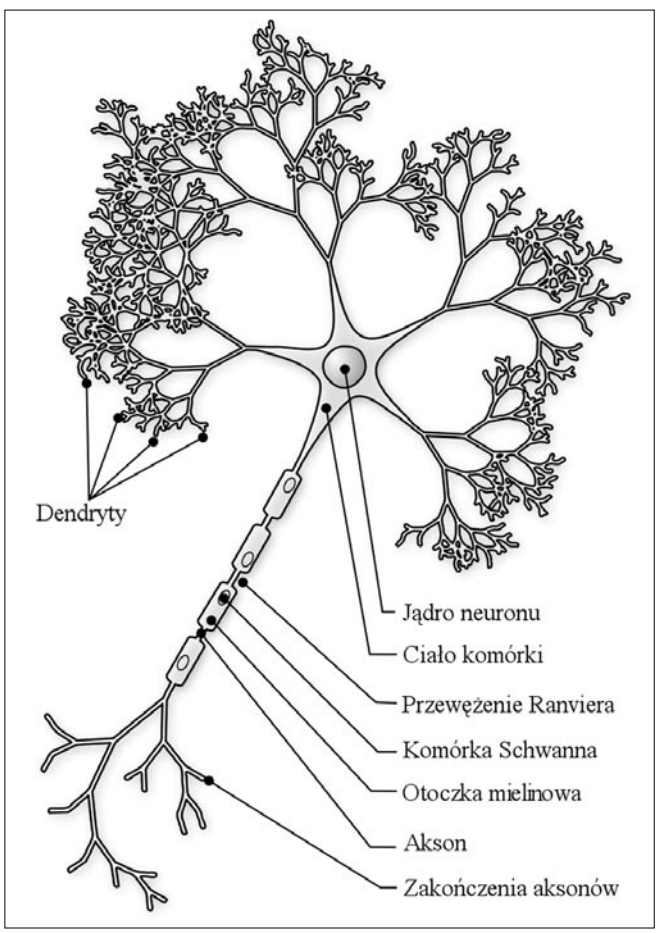
Podczas uczenia można natrafić na wiele trudności. Istotnym problemem, występującym podczas uczenia jest „nadmierne dopasowanie" (ang. *overfitting*) znane również jako przeuczenie lub przetrenowanie modelu. Ryzyko to dotyczy nie tylko sieci neuronowych, ale i innych metod uczenia się maszyn. Podstawowym sposobem zapobieżenia takiemu problemowi jest podział zbioru przykładów wykorzystywanych podczas uczenia na dwie części. Pierwsza z nich, zazwyczaj większa nazywana zbiorem uczącym, jest wykorzystywana bezpośrednio w procesie uczenia. Drugiej, tzw. zbioru testowego, używa się po zakończeniu procesu nauczania do sprawdzenia jakości wytrenowanego modelu. Popularną metodą określania błędu sieci jest błąd średniokwadratowy. Jeśli okaże się, że jest on znacznie większy dla przykładów ze zbioru testowego niż uzyskiwany podczas uczenia najpewniej oznacza to nadmierne dopasowanie. Należy wówczas podjąć odpowiednie działania naprawcze. Najczęściej wiąże się to ze zmianą struktury sieci neuronowej lub algorytmu uczenia i ponownym trenowaniem.

Do uczenia perceptronu stosuje się następujący algorytm:

1. Wybierz losowo wagi początkowe.
2. Podawaj na wejścia perceptronu kolejne przykłady ze zbioru uczącego.
3. Dla każdego przykładu oblicz wartość wyjściową.
4. Jeśli wartość wyjściowa różni się od wzorcowej oblicz nowe wartości wag według formuły $w+d(x_j)x_j$, gdzie *w* to wektor wag w poprzednim kroku uczenia, x_j to wektor wartości wejściowych dla j-tego przykładu natomiast $d(x_j)$ to wartość wzorcowa dla tego wektora.
5. Wróć do 2 i powtarzaj dopóki nie wyczerpie się zbiór uczący, lub nie zostanie uzyskany odpowiednio niski poziom błędu.

W uzasadnionych sytuacjach, gdy dostępna jest tylko niewielka liczba przykładów, dopuszczane jest użycie zbioru trenującego kilkukrotnie. Należy się wtedy jednak liczyć z większym ryzykiem nadmiernego dopasowania.

Jak się okazuje taki algorytm uczenia jest zbieżny i jeśli tylko istnieje wektor wag prawidłowo klasyfikujący



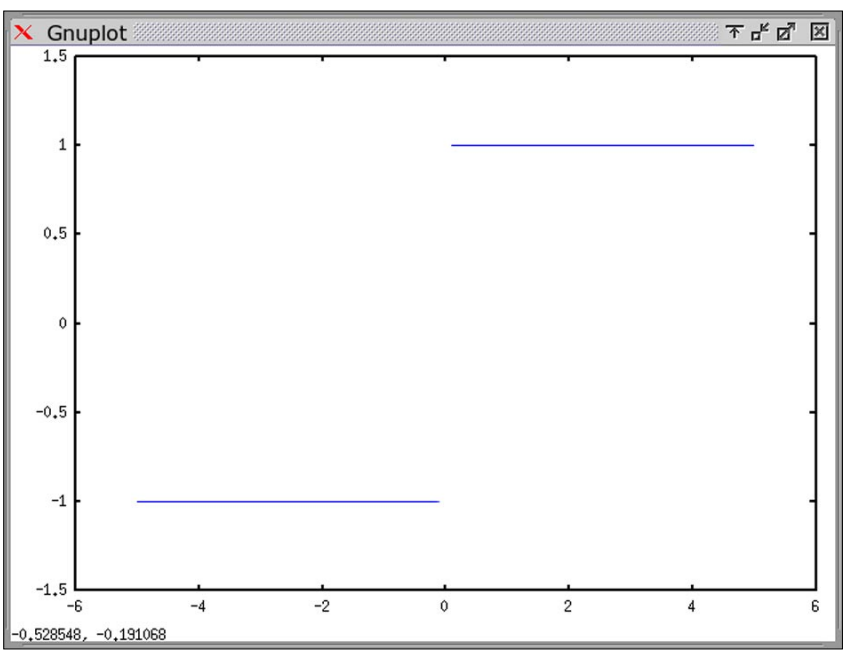
Rysunek 1. *Schemat komórki nerwowej. Źródło: Wikipedia.*

wszystkie przykłady to zostanie on znaleziony w skończonej liczbie kroków.

Bardziej złożone sieci neuronowe

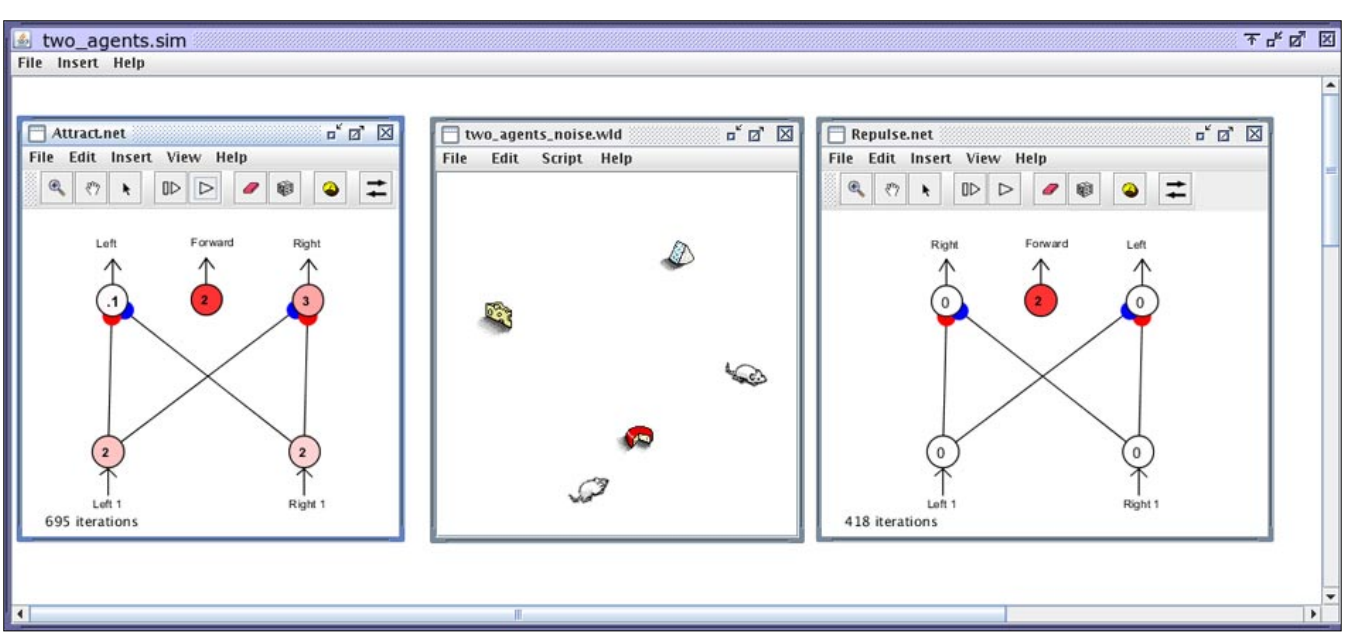
Sieć, składająca się z jednego neuronu nadaje się do zastosowania w stosunkowo niewielkiej klasie problemów, gdy kategorie przykładów są liniowo separowalne. Z tego powodu w praktycznych zastosowaniach używa się bardziej złożonych struktur. Sieci wielowarstwowe składają się z wielu neuronów pogrupowanych w tak zwane warstwy. Do pierwszej z warstw doprowadzone są wejścia, natomiast ostatnia ma wyjścia całego modelu. Najczęściej pomiędzy nimi znajdują się warstwy ukryte, czyli takie, które nie mają bezpośredniego połączenia z wejściem ani wyjściem sieci.

Wszystkie neurony mają połączenia z każdym z neuronów warstwy następnej. Tak zbudowana sieć jest bardziej elastyczna i może być zastosowana do rozwiązywania znacznie większej klasy problemów niż tylko klasyfikacja kategorii liniowo separowalnych. Pojawia się jednak ryzyko zbytniego skomplikowania modelu poprzez dodanie zbyt wielu neuronów warstw ukrytych. Nie istnieje uniwersalna metoda, pozwalająca na wyznaczenia optymalnej struktury sieci. W praktyce najczęściej trzeba polegać na własnym doświadczeniu, obserwując jednocześnie błąd na zbiorze uczącym i testowym. Jeśli ten drugi jest znacząco większy od pierwszego najczęściej oznacza to, że strukturę sieci trzeba uprościć, usuwając część neuronów lub jedną z warstw ukrytych. Jeśli natomiast zarówno błąd na zbiorze testowym jak i uczącym mimo długiego uczenia nie zmiejsza się do oczekiwanego poziomu, może to oznaczać, że zakładany model jest zbyt prosty. W takiej sytuacji pomóc może rozbudowanie sieci o dodatkowe neurony lub wartstwę ukrytą.



Rysunek 2. *Funkcja signum jest często wykorzystywana jako funkcja aktywacji neuronu.*

Uczenie sieci wielowartstowej, podobnie jak w przypadku jednowarstwowej, polega na minimalizacji błędu. Jest on najczęściej defioniowany jako błąd średniokwadratowy czyli suma kwadratów różnic pomiedzy uzyskanymi i wzorcowymi wartościami wyjściowymi. Do trenowania sieci można zatem zastosować algorytmy znane z optymalizacji: od metod gradientowych, na przykład metody największego spadku, poprzez



Rysunek 3. *Symulacja sterowania dwoma agentami — myszami za pomocą sieci neuronowych w programie Simbrain.*

metody drugiego rzędu takie jak algorytm Levenberga-Marquardta, aż do algorytmów ewolucyjnych.

Jednym z najpopularniejszych algorytmów uczenia sieci wielowarstwowych jest metoda wstecznej propagacji błędu (ang. *back propagation*). Mimo, że zwykle jest wolniejsza od nowoczesnych algorytmów, takich jak metoda Levenberga-Marquardta, wciąż ma wielu zwolenników. Przede wszystkim ze względu na prostotę oraz to, że nie nakłada na funkcję celu, opisującą błąd wielu, niekiedy trudnych do sprawdzenia ograniczeń. W efekcie pozwala rozwiązanie praktycznie każdego zadania, choć rzadko będzie to rozwiązanie najszybsze.

Algorytm wstecznej propagacji błędu:

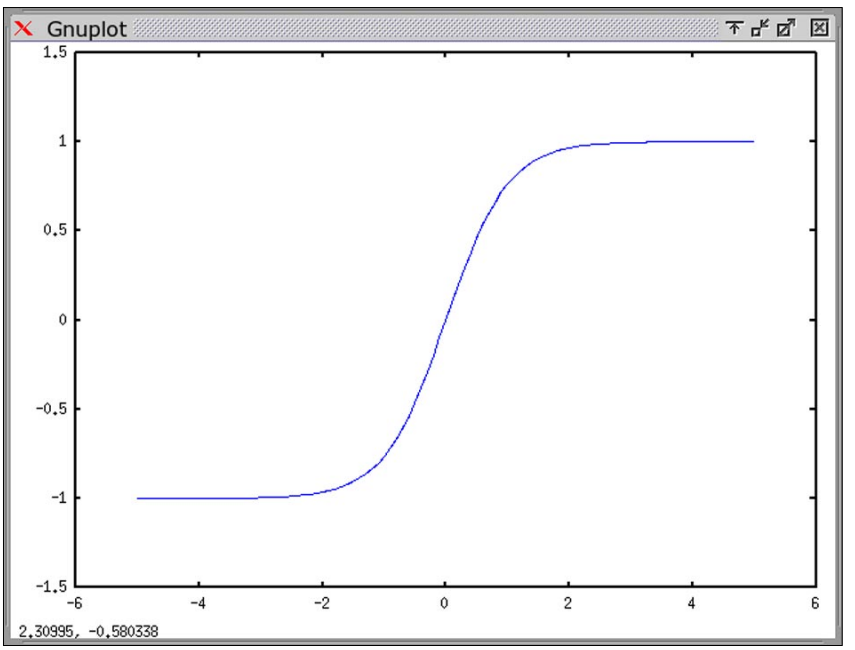
1. Wybierz losowo początkowe wartości wag. Najlepiej małe liczby z przedziału (0,1).
2. Podawaj na wejście sieci kolejne przykłady ze zbioru uczącego.
3. Dla każdego przykładu oblicz wartość wyjściową.
4. Oblicz wartości odchyleń w ostatniej warstwie.
5. Oblicz kolejno wartości odchyleń w poprzednich warstwach za pomocą propagacji wstecznej.
6. Na podstawie odchyleń wyznacz wartości poprawek dla wag we wszystkich warstwach.
7. Powtarzaj od 2, aż do uzyskania wymaganego poziomu błędu lub wyczerpania zbioru uczącego.

Opisany algorytm, podobnie jak inne metody gradientowe jest w niewielkim stopniu odporny na minima lokalne. Oznacza to, że znalezione ekstremum funkcji celu, czyli wartości wag sieci neuronowej, mogą wcale nie być optymalne, a jedynie najlepsze w najbliższym otoczeniu. Dlatego też zostało opracowanych kilka modyfikacji algorytmu wstecznej propagacji błędu przyspieszających jego zbieżność oraz uodparniających na minima lokalne. Do takich modyfikacji należy m.in. metoda Quickprop oraz Rprop.

Artykuł, ze względu na swoją niewielką obietość porusza tylko niewielki zakres tematów związanych z budową i zasadami działania i uczenia sieci neuronowych. Jednak podstawowa wiedza w nim zawarta powinna pozwolić na zrozumienie tych zasad i świadome a przez to skuteczniejsze i lepsze wykorzystywanie rozwiązań bazujących na tym modelu.

Propozycja zastosowania sieci neuronowej w systemie wykrywania włamań

Jako przykład praktycznego zastosowania sieci neuronowych zaproponowano wykorzystanie ich w systemie wykrywania włamań do systemów informatycznych. Wiele jest rozwiązań zabezpoieczających komputery i sieci komputerowe przed intruzami (ang. *Intrusion Detection systems,* IDS). Systemy tej klasy mają za zada-



Rysunek 4. *Funkcja tangens hiperboliczny jest często wykorzystywana jako funkcja aktywacji neuronu.*

nie rozpoznać czy aktywność podejmowana przez użytkowników jest próbą włamania. W tym celu najczęściej wykorzystywana jest baza reguł stworzoną przez ekspertów. Na podstawie reguł zdarzenia w systemie np. ruch sieciowy klasyfikowane są jako atak lub normalna praca. Takie podejście ma swoje dobre strony. Klasyfikator poprawnie rozpoznaje próby włamań i generuje niewiele fałszywych alarmów ponieważ wykorzystuje głęboką wiedzę i doświadczenie eksperta odpowiedzialnego za stworzenie reguł. Trudności pojawiają się, gdy zachodzi konieczność użycia IDS w środowisku dynamicznym. Intruz może zastosować nową, nieznaną wcześniej metodę włamania, lub modyfikację jednej ze znanych metod polegającą np. na przeprowadzeniu kolejnych faz ataku z różnych hostów. Klasyfikator wykorzystujący wiedzę eksperta zazwyczaj nie rozpozna prawidłowo takiej sytuacji. Można zwiększyć jego czułość, jednak wtedy będzie generował wiele fałszywych alarmów. W konsekwencji może zajść konieczność ingerencji eksperta i rozbudowania systemu o reguły obejmujące nowy rodzaj ataków.

Alternatywą do wykorzystania wiedzy eksperckiej jest zastosowanie modelu, który sam potrafi nauczyć się reguł klasyfikacji. W takim zastosowaniu dobrze sprawdzają się sztuczne sieci neuronowe. Nauczona sieć może mieć nieco gorsze wyniki w klasyfikacji znanych sposobów włamań niż reguły eksperta, ale jej największe zalety ujawnią się przy rozpoznawaniu nowych typów ataków. Jedną z najważniejszych cech, która decyduje o popularności sieci neuronowych jest ich bardzo dobra zdolność generalizacji. Nauczony model może uogólnić klasyfikację na przypadki, które nie znalazły się w zbiorze uczącym a jedynie są do nich podobne. Dzięki dobrze wytrenowana sieć prawidłowo rozpozna różne modyfikacje znanych form ataku. Jeśli odkryte zostaną nowe metody włamań w większości wystarczy tylko przeprowadzić uczenie modelu na podstawie nowych przykładów. Można również zastosować inne rodzaje sieci neuronowych tak zwane sieci Kohonena lub samo-organizujące się mapy nie opisane w artykule.

W praktycznych zastosowaniach rzadko IDS będzie działał w oparciu wyłącznie o sieć neuronową. Najczęściej będzie ona wykorzystana do wstępnej klasyfikacji zdarzeń a jej wyjście przekazywane jest do systemu eksperckiego. Innym sposobem połączenia obu rozwiązań może być bazowanie głównie na regułach eksperta i wykorzystywanie modelu neuronowego jedynie w przypadkach niepewnych, których nie pokrywają znane reguły.

Podsumowanie

Sztuczne sieci neuronowe uzyskują bardzo dobre wyniki podczas badań. Wiele z ich zalet zostało potwierdzonych nie tylko eksperymentalnie, ale również udowodnionych matematycznie. Mimo to, podobnie jak wiele inteligentnych metod obliczeniowych, z trudem zdobywają zaufanie zwyczajnych użytkowników. Najczęściej rozwijane i wykorzystywane są na uczelniach lub w instytutach naukowych. Jeśli pojawiają się w rozwiązaniach komercyjnych to stanowią najczęściej mało ważne, dodatkowe moduły. Artykuł, choć porusza bardzo niewielką część tematów związanych ze sztucznymi sieciami neuronowymi, stara się przybliżyć te zagadnienia i przekonać czytelnika, że metoda jest dobrze opracowana i sprawdza się zarówno w teorii jak i praktyce.

W Sieci

- *http://memaid.sourceforge.net/* - MemAid, programowe fiszki, wykorzystują sieć neuronową, by zaplanowac optymalny harmonogram nauki
- *http://moodss.sourceforge.net/* - moodss, graficzny program do monitorowania stanu usług
- *http://sourceforge.net/projects/rotator/* - Image Rotator, program, który wykorzystuje sieci neuronowe do rozpoznać orientacji zdjęć.

WOJCIECH TERLIKOWSKI

Autor jest z wykształcenia inżynierem elektroniki. Od ponad dwóch lat pracuje w jednej z największych polskich firm informatycznych, gdzie pełni rolę m.in. konsultanta do spraw bezpieczeństwa systemów informatycznych. Kontakt z autorem: w.terlikowski@terlikowski.eu.org.

Badanie pamięci flash urządzeń mobilnych

Salvatore Fiorillo

Niniejszy dokument stanowi wprowadzenie do badania pamięci flash ze szczególnym zwróceniem uwagi na kompletność dowodów uzyskanych z telefonów komórkowych.

Dowiesz się:
- jakie są badania pamięci flash
- o mobilnej kryminalistyce

Powinieneś wiedzieć:
- podstawowe informacje na temat urządzeń flash

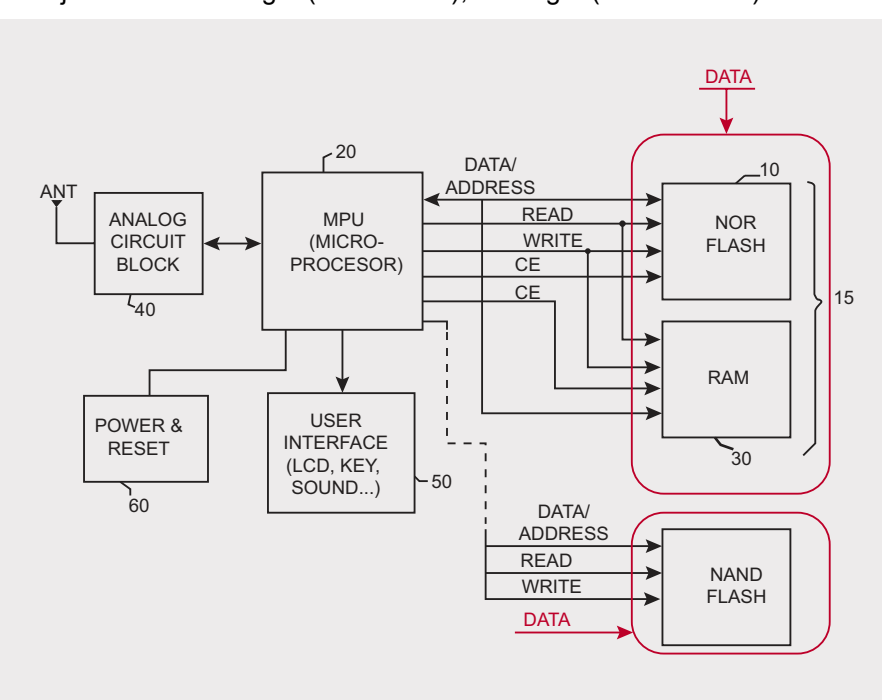
Przedstawiam szczególny charakter pamięci nielotnych, które są obecne w dzisiejszych telefonach komórkowych, jak naprawdę działają i jakie wyzwania stanowią dla badaczy sądowych. Następnie zostaną zaprezentowane zaawansowane badania, w których niektóre nowe pamięci flash są wykorzystywane do ukrywania danych w uszkodzonych przez człowieka blokach: celem jest sprawdzenie, czy oprogramowanie wykorzystywane do badania jest w stanie pobierać danych z takich bloków, oraz ocena możliwości ukrywania danych przed oczami analityków.

Środowisko mobilne

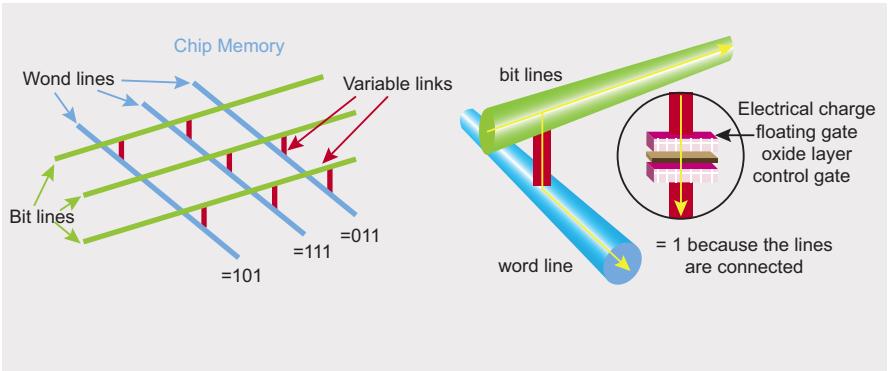
Mobile Equipment (ME) jest tu rozumiane jako słuchawka radiowa bardziej ogólnego telefonu komórkowego (Jansen i Ayers, 2007), wykonana przez różne elementy, z których najważniejsze zostały przedstawione na rysunku poniżej (Rys. 1).

W trakcie jego ewolucji telefon przeszedł z fazy PDA do dzisiejszych inteligentnych telefonów, które zmniejszają różnice w stosunku do komputerów osobistych. Pojemność przechowywania również gwałtownie wzrosła, od kilku Kilobitów na początku do kilku Gigabitów w obecnych telefonach komórkowych, zwiększając przestrzeń, w której dane mogą być przechowywane lub ukrywane i dodatkowo komplikując pracę stróżów prawa (Al-Zarouni, 2006): ten dokument ma przyczynić się do przesunięcia dziedziny badania pamięci flash z poznawalnych do znanych w modelu Cynefin (Kurtz i Snowden, 2003).

W dzisiejszych urządzeniach mobilnych istnieją generalnie dwie pamięci: jedna dla systemu operacyjnego (NOR flash), a druga (NAND flash) na dane



Rysunek 1. *Stary układ mobilnych urządzeń z opcjonalnym modułem NAND (Kwon, 2009)*



Rysunek 2. *Podstawowe projekt kości pamięci (po lewej) i połączeń pamięci flash (z prawej) (O'Kelly, 2007)*

użytkownika (i Kuo Chang, 2004). Zakres tego dokumentu jest ograniczony do danych przechowywanych w NAND: analiza ulotnej pamięci RAM i karta SIM jest odsunięta na bok.

NOR i NAND

Pamięć flash jest nieulotną pamięcią, która może być elektrycznie wymazana i ponownie zapisana w odpowiednim procesie: podobnie jak dysk twardy (mimo dużych różnic wynikających z braku mechanizmów fizycznych), pamięć flash nie potrzebuje zasilania do przechowywania danych zapisanych w mikroprocesorze (O'Kelly, 2007). Pochodzące z ewolucji EPROM, dwa główne rodzaje pamięci flash to NAND i NOR. NOR flash wymaga długiego czasu kasowania oraz zapisu, ale za to jest prawie odporna na korupcję i złe bloki, umożliwia swobodny dostęp do dowolnego miejsca pamięci i prawie wszystkie kontrolery w telefonach komórkowych posiadają interfejs NOR (Pon et al., 2007). NAND flash oferuje większą gęstość możliwości, jest tańszy niż NOR, jest mniej stabilna, potrzebuje wsparcia oddzielnej pamięci RAM do pracy i nie umożliwia sekwencyjnego trybu dostępu (Ga i Toledo, 2005). W mobilnych urządzeniach zwykle NOR przechowuje oprogramowanie wykonywalne (np. BIOS) a NAND przechowuje dane, takich jak obraz lub pliki mp3 (Peng, 2006, Raghavan et al., 2005). W dodatkach wskazana jest w tabeli zawierająca porównanie tych dwóch pamięci flash.

Model kodu

Istnieją dwie techniki uruchamiania kodu programu na urządzeniach flash (Numonyx, 2008a): Przechowaj i Pobierz (Store and Download - SND), wymagający zewnętrznej pamięci RAM, i Wykonaj w Miejscu (eXecute in Place - XIP) - szybsza niż SND, wymagająca swobodnego dostępu. NOR wykorzystuje XIP podczas gdy NAND używa SND.

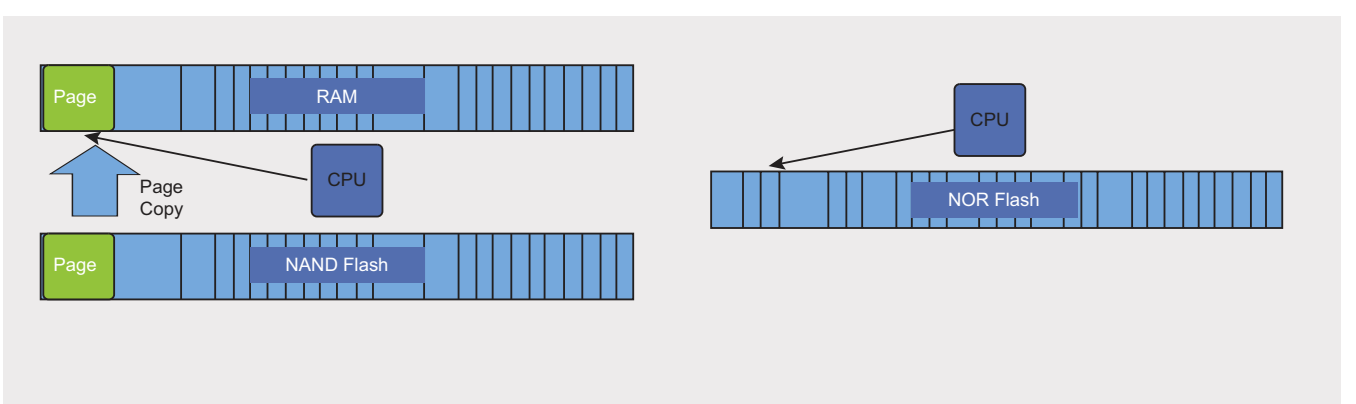
Programowania jednokierunkowe

Urządzenia flash są w stanie programować wartości jedynie z 1 na 0, ale nie z 0 na 1, więc gdy dane są aktualizowane, są one zapisywane w nowe miejsce a stare miejsce jest oznaczone jako nieprawidłowe (Numonyx, 2008a). Nieprawidłowe fragmenty są następnie kasowane - zwykle przez proces działający w tle - przed ponownym użyciem.

Zużywający cykl kasowania-zapisywania

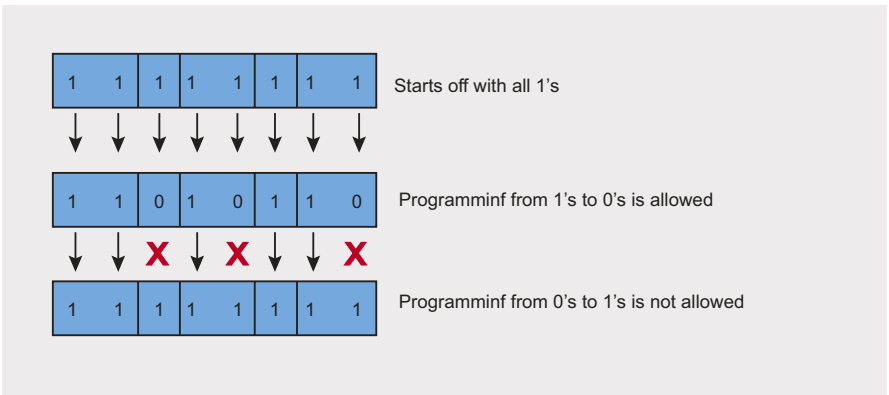
W przeciwieństwie do dysków twardych, cykl skasowania-zapisu w pamięciach flash jest fizycznie wyczerpującym zadaniem, więc czas życia pamięci flash jest odwrotnie proporcjonalna do jej użycia. Poszczególny fragment pamięci może być zaprogramowany i wykasowany niezawodnie do 100.000 razu i 10.000 razy respectively jako generalna zasada, następująca formuła może być stosowana do obliczenia spodziewanego okres eksploatacji pamięci flash NAND z systemem plików FAT (Numonyx, 2008a). Techniki służące obejścia problemu zużywania się pamięci flash zostaną omówione w dalszej części.

$$\text{Expected lifetime} = \frac{\text{Size of NAND flash x number of erase cycles x FAT overhead}}{\text{Bytes written per day}}$$

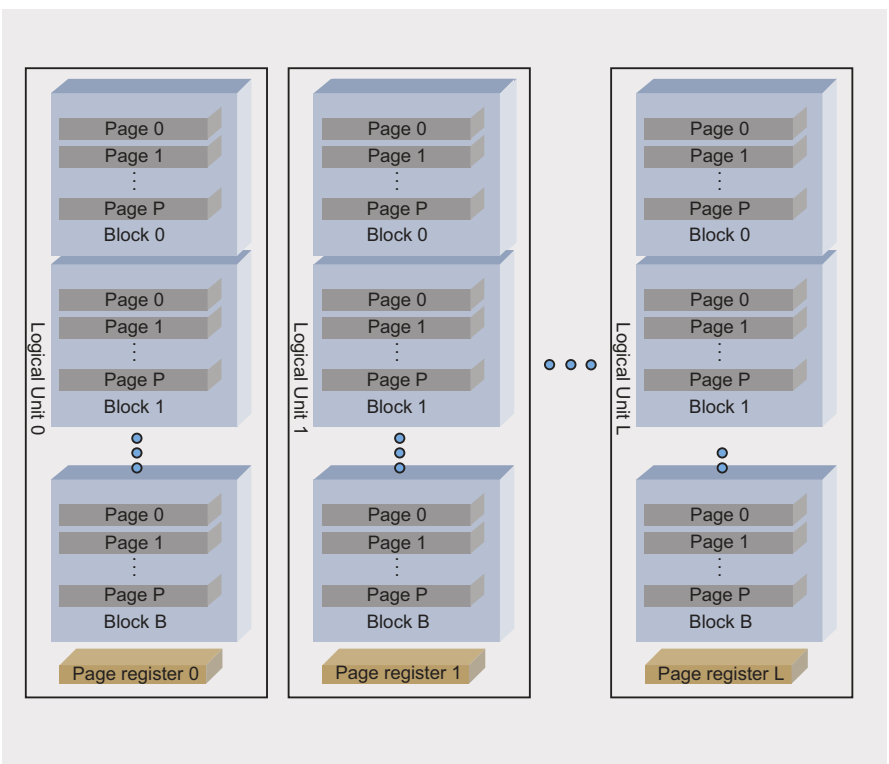


Rysunek 3. *Model SND (z lewej) i model XIP (z prawej) (Numonyx, 2008a)*

FAT zawiera wszystkie działania jakie system plików potrzebuje wykonać aby administrować plikami (Hendrikx, 1998)



Rysunek 4. *Ograniczenia programowania flash*



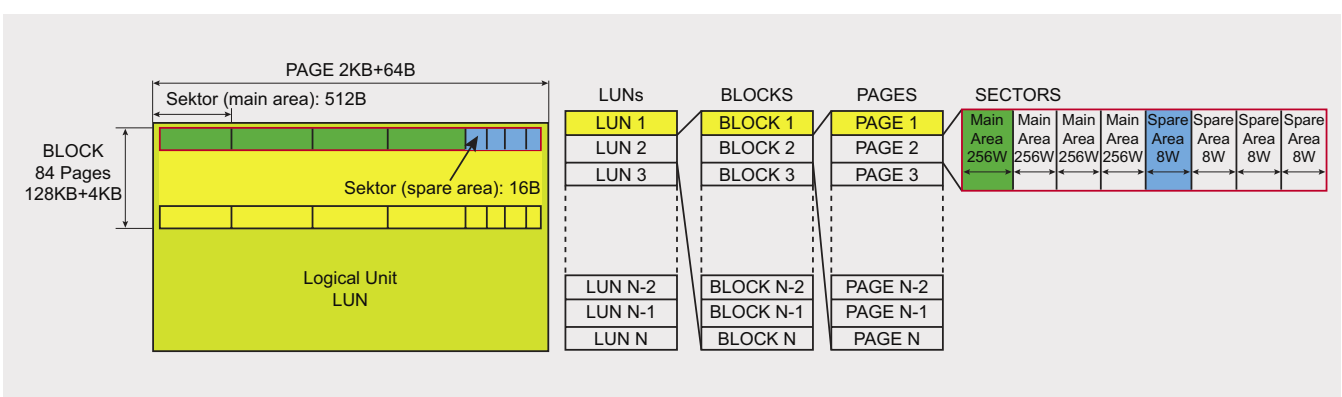
Rysunek 5. *Jednostka logiczna w pamięciach NAND*

Architektura systemu plików flash

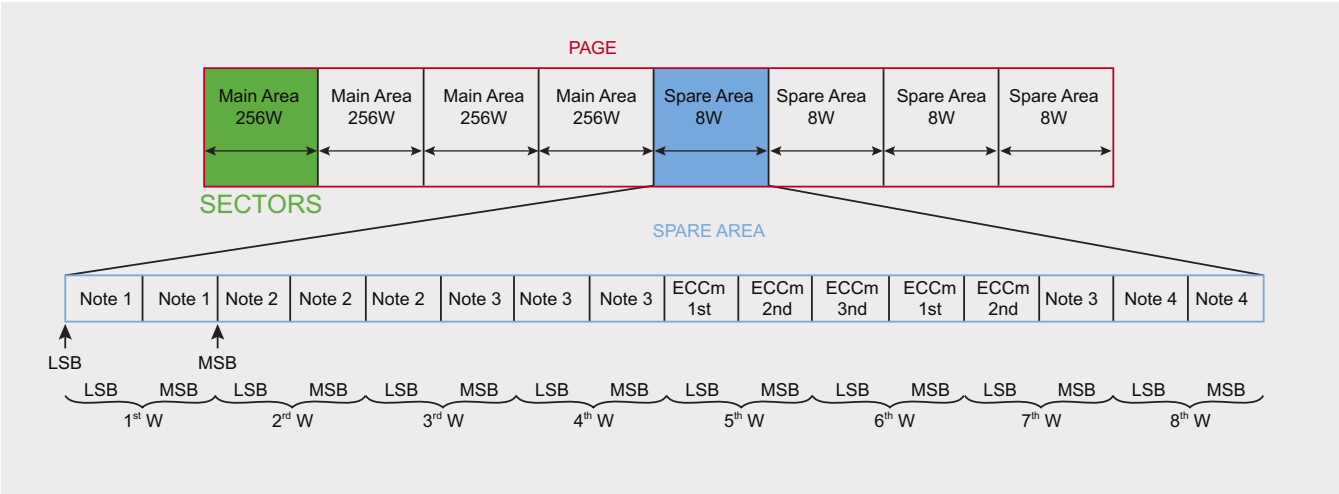
Architektura systemu plików flash jest oparta na jednostkach logicznych (logical unit - LUN), blokach, stronach i sektorach (Intel, 2006, Numonyx, 2008a, Samsung, 1999). LUN jest największą logiczną częścią obszaru pamięci.

LUN są następnie podzielone na bloki. Każdy blok może różnić się wielkością, gdzie najczęstszym jest 128KB. W większości urządzeń flash NAND każdy blok składa się z 64 stron w każdym 2KB. Strona jest podzielona na dwie części: obszar danych i dodatkowy obszar wykorzystywany do celów zarządzania pamięcią (więcej w dalszej części). Strony są podzielone na sektory (lub kawałki) 512-bajtowe aby naśladować popularny rozmiar sektora. Blok jest najmniejszą jednostką kasowalną podczas gdy strona jest najmniejszą programowalną jednostką.

Na początku, strona miała 528 bajtów w związku z tym że oryginalną intencją NAND Flash było zastąpienie magnetycznych dysków twardych, więc było wymagane aby strona była wystarczająco duża żeby przechowywać jeden sektor (512 bajtów) danych z dodatkowymi 16 bajtami w celu zarządzania (Inoue i Wong, 2004). Później, w miarę wzrostu pojemności flash, wzrósł domyślny rozmiar strony aby dostosować się do systemu FAT. W 1 GB pamięci flash, jest 128 MB adresowalnych przestrzeni: dla dysków twardych wielkości do 128 MB, domyślny rozmiar klastra w systemie



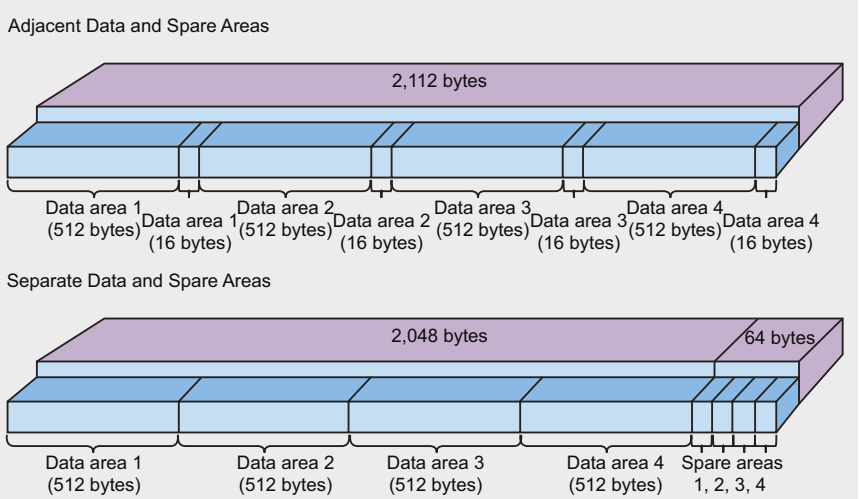
Rysunek 6. *Architektura pamięci flash*



Rysunek 7. *Architektura pamięci flash*

Word	Byte	Note	Description
1	LSB	1	Invalid Block information in 1st and 2nd page of an invalid block
	MSB		
2	LSB	2	Managed by intemal ECC logic for Logical Sector Number data
	MSB		
3	LSB		
	MSB	3	Reserved for future use
4	LSB		
	MSB		
5	LSB		Dedicated to intemal ECC logic. Read Only. ECCm 1st for main area data
	MSB		Dedicated to intemal ECC logic. Read Only. ECCm 2nd for main area data
6	LSB		Dedicated to intemal ECC logic. Read Only. ECCm 3rd for main area data
	MSB		Dedicated to intemal ECC logic. Read Only. ECCm 1st for 2nd word of spare area data
7	LSB		Dedicated to intemal ECC logic. Read Only. ECCm 2nd for 3rd word of spare area data
	MSB	3	Reserved for future use
8	LSB	4	Available to the user
	MSB		

Rysunek 8. *Przypisanie dodatkowej przestrzeni w wewnętrznej pamięci NAND w OneNANDTM ™ (źródło: Samsung)*



Rysunek 9. *Metody przechowywania dodatkowej przestrzeni*

FAT wynosi 2KB i złożony jest on z 4 sektorów, jak w pamięci flash, z wyjątkiem dodatkowych bajtów (64B) (Microsoft, 2009) .

Dodatkowa przestrzeń

Dodatkowa przestrzeń, zwana również poza zakresowymi danymi, jest regionem, zazwyczaj wykonanym z 16 bajtów i jest jedna dla każdego sektora lub kawałka (Ga i Toledo, 2005, Raghavan et al., 2005), a jej wielkość nie jest zawarta w pojemności urządzenia i nie jest bezpośrednio adresowalna (Elnec, 2009). Jednym z zastosowań tej części obszaru jest przechowywanie wyników weryfikacji danych: po tym jak strona została wykasowana, usunięta, zaprogramowana lub przeczytana, jej stan jest weryfikowany za pomocą szczególnego algorytmu (aka ECC - więcej w dalszej części) i później wyjście tego algorytmu jest używane do wykrywania błędów gdy dane są odczytywane ponownie (BPMicrosystems, 2008). Dodatkowa przestrzeń może przechowywać także informacje o stanie bloków i stron (Tsai et al., 2006), lub inne informacje podobne do metadanych w systemie plików NTFS (Carrier, 2005, Casey, 2004). Poniżej znajduje się przedstawienie dodatkowej przestrzeni w Samsung OneNANDTM™, więcej informacji na stronie (Samsung, 2005a).

Są dwie metody przechowywania dodatkowych obszarów: obszar przylegający do danych lub oddzielony od nich (Micron, 2006a). Patrząc na większość danych firmy Samsung wydaje się że używają głównie modelu drugiego.

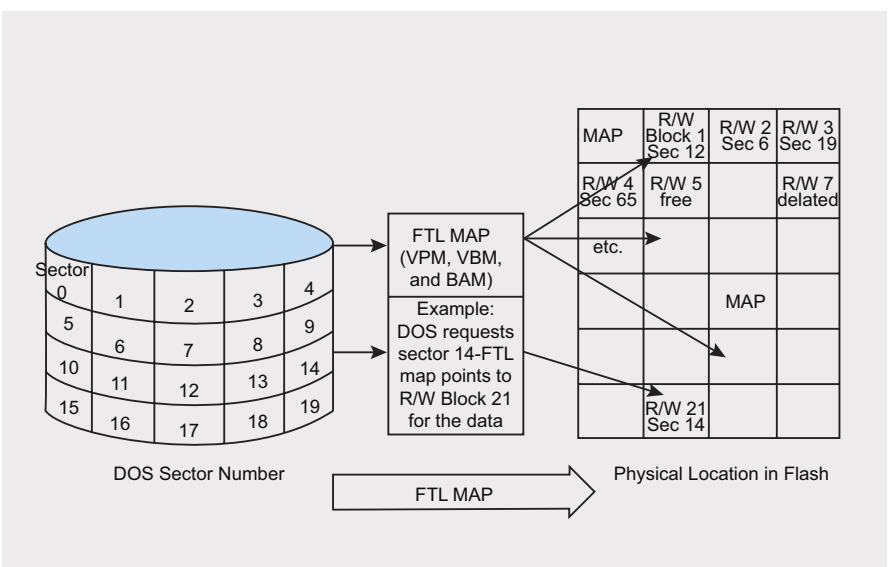
NAND vs. dysk twardy

Główne różnice pomiędzy urządzeniami flash i dyskami twardymi (Raghavan i in., 2005):

- standardowa wielkość sektorów (patrz rozmiar sektora flash na zdjęciu poniżej);
- w przeciwieństwie do dysków twardych, operacja pisania i usuwania w urządzeniu flash mogą być samodzielnymi działaniami i związane są z oprogramowania przyrządu flash;
- chipy flash mają ograniczoną trwałość ze względu na zużywanie podczas kasowania;
- urządzenia flash mogą być pozbawione zasilania bez właściwego zamknięcia i danych pozostaną spójne: nie jest to możliwe w przypadku dysku twardego ze standardowym systemem plików, więc systemy wbudowane potrzebują szczególnego zarządzania plikami ukierunkowanego na flash.

OneNAND			Density	NAND		
Sector	Page	Block		Block	Page	Sector
512 B + 16 B	1 KB (2 sectors)	64 KB (64 pages)	256 Mb	32 KB (64 pages)	512 B (1 sectors)	512 B + 16 B
	2 KB (4 sectors)	128 KB (64 pages)	512 Mb 1 Gb 2 Gb 4 Gb	128 KB (64 pages)	2 KB (4 sectors)	
	Not Available					

Rysunek 10. *Standardowy rozmiar bloku w sektorze urządzeń o gęstości poniżej 256 Mb i ponad 512 Mb (źródło: Samsung)*

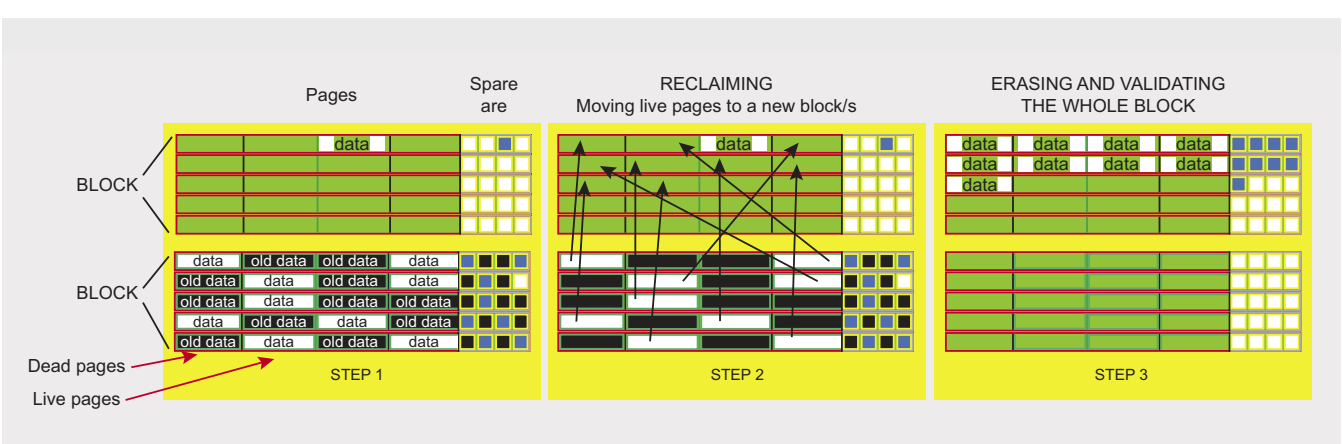


Rysunek 11. *FTL przeniesienie sektora*

Systemy plików flash i warstwa przenoszenia flash

"System plików jest strukturą danych, która reprezentuje zbiór zmiennych plików o dostępie swobodnym w hierarchicznej przestrzeni nazw" (Ga i Toledo, 2005). Do pracy z systemu plików hosta, pamięci NAND flash wymagały konkretnych systemów plików lub sterownik. Obecnie mamy zarówno konkretne systemy plików flash (jak YAFFS, JFFS, UBIFS i LogFS), jak również specjalny sterownik znany jako warstwa przenoszenia flash (Flash Translation Layer - FTL).

"FTL jest sterownikiem, który działa w połączeniu z istniejącym systemem operacyjnym (lub, w niektórych aplikacji wbudowanych, jako systemoperacyjny), aby liniowa pamięć flash zachowywała się jak dysk" (Intel, 2006).



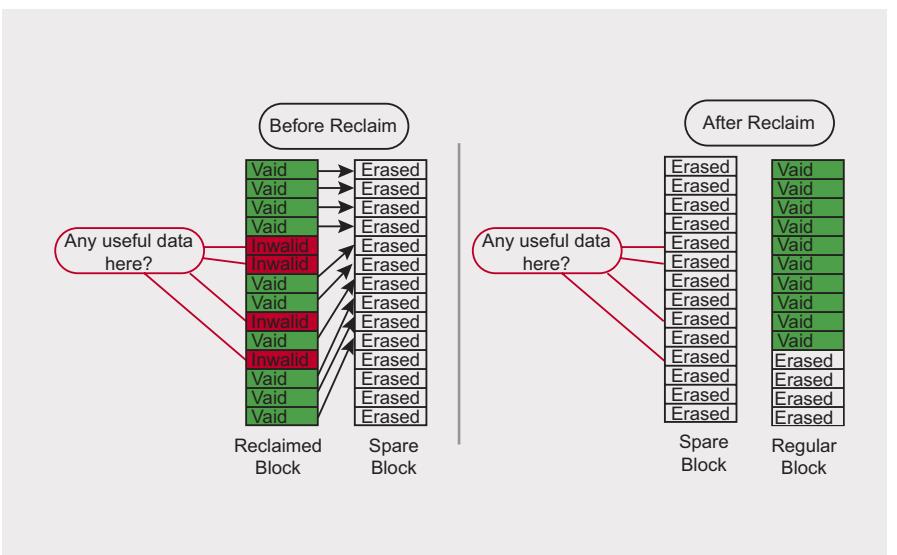
Rysunek 12. *Proces odzyskania w ramach polityki poziomowani zużycia*

Głównym zadaniem FTL jest wspieranie wszystkich zadań niezbędnych do zarządzania danymi w sposób przezroczysty dla systemu plików hosta, np.: system plików FAT będzie delegował na FTL wszystkich działań koniecznych do przechowywania i pobierania danych odpowiednio do / z urządzeń NAND flash (BPMicrosystems, 2008, Intel 1998, Morris, 2007). Główne zadania FTL to:

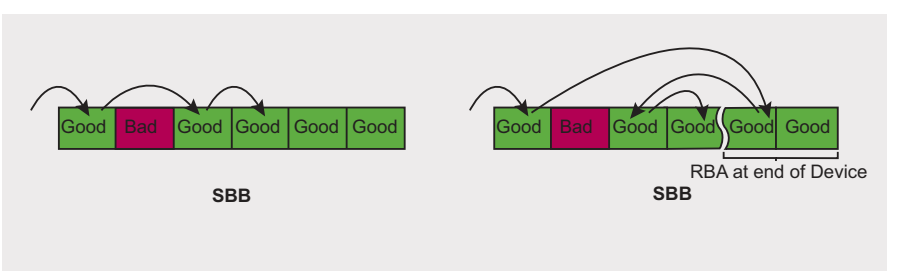
- Mapowanie powierzchni przechowywania danych w wirtualne małe sektory
- Zarządzanie danymi na flash tak aby wydawało się że są "zapisane w miejscu"
- Sprzątanie: jako że pamięci flash ulegają zużyciu, wymagane jest oprogramowanie, które będzie poziomowało wykorzystania obszarów pamięci.

FTL dla NAND może być dostarczone w różny sposób: może być stworzone przez producenta i wbudowane w urządzenie (np. Samsung), może być wbudowane w system operacyjny zorientowany na flash (np.: YAFFS) lub może być wykonane przez producenta jako port dla systemu operacyjnego jak Unistore II wyprodukowanych przez Samsung dla systemu operacyjnego Symbian (Morris, 2007, Samsung, 2006b). Aby uzyskać więcej informacji na temat algorytmów i struktur danych patrz (Ga i Toledo, 2005).

Wracając do UBIFS, jest to nowy system plików flash opracowane przez inżynierów firmy Nokia z pomocą Uniwersytetu w Szeged i może być traktowany jako kolejna generacja systemu plików JFFS2 (MTD_group, 2008).



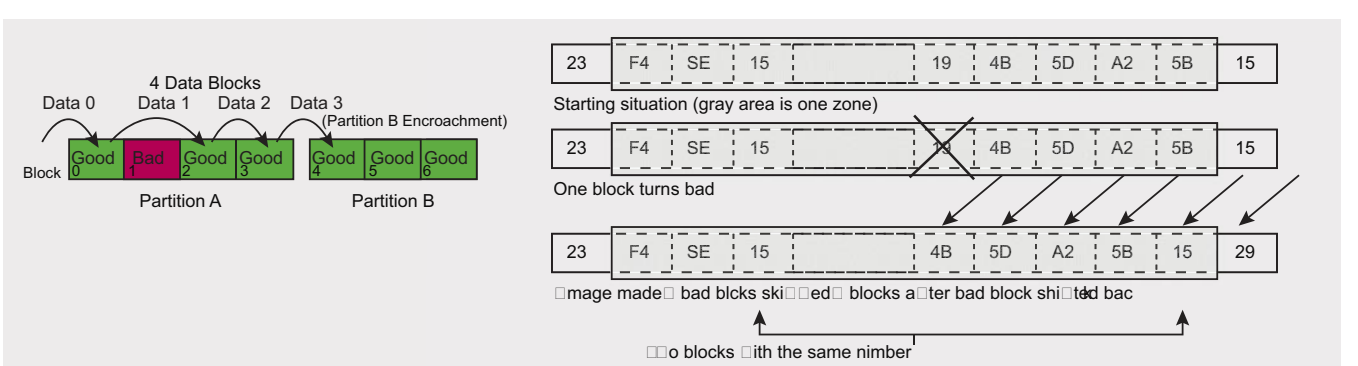
Rysunek 13. *Stan bloków przed i po procesie Reclaim (Intel, 2006)*



Rysunek 14. *Strategia SBB (z lewej) w porównaniu do RBA (z prawej) (BPMicrosystems, 2008)*

Zużycie poziomowania (Wear Levelling - WL) i Garbage Collection (GC)

Gdy dane w pamięci flash są aktualizowane, nie jest możliwe zaprogramowanie tej samej strony w jedno-kierunkowym programowaniu urządzeń flash, więc strona zawierająca dane do aktualizacji jest całkowicie przepisywana do nowej lokalizacji (może być w tym samym bloku lub nie). W dodatkowym obszarze strona z nowymi danymi jest oznaczone jako poprawna (żywa), a stara jest oznaczony jako nieprawidłowy (martwa). Gdy liczba martwych stron w bloku jest większa niż określają wytycz-



Rysunek 15. *Ingresja bloku (z lewej) i powielanie numer bloku (z prawej)*

ne wtedy żywe strony są przepisywane do nowych lokalizacji i blok zostaje usunięty, aby umożliwić jego przyszłe programowanie: jest to ukryty proces zwany Odzyskujący Garbage Collection (Reclaim Garbate Collection) i jest aktywowany bez udziału użytkownika i nie jest określony w czasie (Tsai et al., 2006).

Uwaga: w powyższym przykładzie, są używane tylko dwa bloki, ale w świecie rzeczywistym regeneracji może podlegać więcej bloków

W celu uniknięcia nadmiernego zużycia jednym obszarze na rzecz innych, proces zwany wyrównywaniem zużycia (Wear Levelling) zarządza blokami tak, aby były używane mądrze: istnieje statyczne i dynamiczne wyrównywanie zużycia oba próbują przedłużyć żywotność urządzenia flash (Numonyx, 2008c, Jones, 2008). Procedury wyrównywania zużycia możne być wbudowana w firmware pamięci flash lub pozostawiona pod opieką systemu plików hosta (Numonyx, 2008b, Numonyx, 2008c, Jones, 2008, JI et al., 2009).

Dane w nieprawidłowych blokach lub strony martwe mogą przechowywać informacje interesujące z punktu widzenia analizy sądowej i należy je pozyska zostanie przeprowadzony Reclaim Garbage Collection: analitycy są proszeni o nie zmienianie stanu materiału dowodowego, ale ponieważ Wear Levelling i Reclaim są procesami ukrytymi, wymóg ten może być trudny do osiągnięcia i trudny w zarządzaniu. W przyszłości pracach będą rozpatrywane wpływy procesu Reclaim na urządzenia wbudowane: wyniki będą podane.

Kod korekcji błędów (ECC)

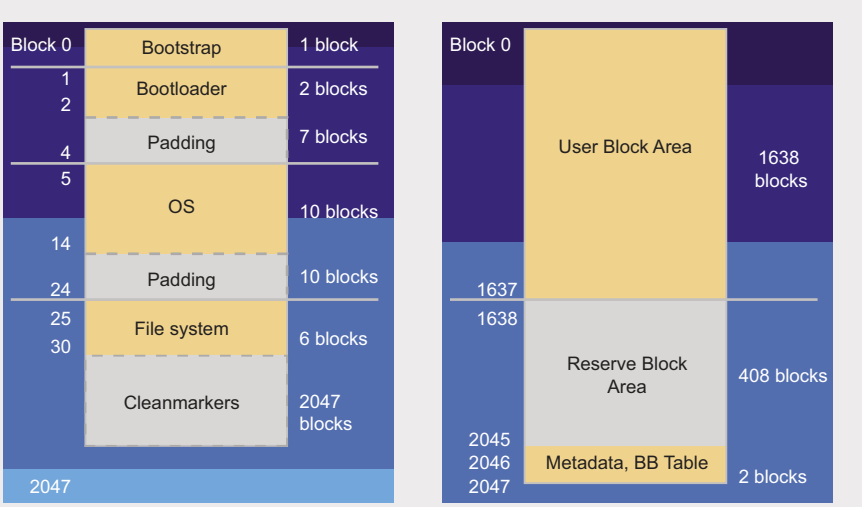
Strona może być zaprogramowana, wyczyszczona i czytana, po każdej operacji jest konieczna weryfikacja stanu strony. Do dokonania tej weryfikacji, urządzenia flash stosują algorytm weryfikacji który wytwarza coś w rodzaju wartości skrótu/CRC dla każdej dostępnej strony: wartość jest następnie zapisywane w dodatkowym obszarze (Numonyx, 2008d). Algorytm ten jest powszechnie zwany Error Correction Code. Jeżeli błąd zostaje wykryty po fazie czytania, może być odzyskane przez ECC, jeżeli błąd zostanie wykryty po cyklu programowania lub usuwanie następuje aktywacja polityki wymiany blok (Micron, 2006a, Samsung, 1999). Więcej informacji na temat ECC, zobacz (Samsung, 2004). W odróżnieniu od Wear Levelling, logika ECC jest na ogół wbudowane w firmware wszystkich pamięci flash.

Pomimo że algorytmy ECC są tajemnice handlowe, niektóre rozwiązania hakerskie są w stanie przerobić dane w urządzeniu flash rekonstruując ECC (jak kod obecny w Sony PlayStation 3 (NDT, 2008)). Jest to nowy zakres nielegalnej działalności, nie opisany tutaj.

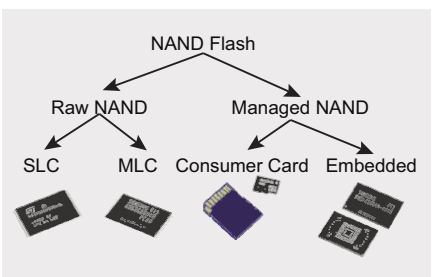
Zarządzanie uszkodzonymi blokami (Bad Management Block - BBM)

Jeśli ECC raportuje błąd nie do odzyskania, wymagane jest, aby obszar zaznaczono jako uszkodzony. Ponieważ najmniejszą jednostką kasowalną powierzchni jest blok, każdy nieodwracalny błąd pojawiający się na jakiejkolwiek stronie powoduje że cały blok do którego strona należy zostanie uznany za wymagający wymiany, nie będzie on więc ponownie dostępny (Samsung, 2006b). Uszkodzone bloki zidentyfikowane w trakcie cyklu życia NAND flash zostaną dodane do listy uszkodzonych bloków wygenerowanych w trakcie produkcji w fabryce i nie powinny przekraczać 2% ogólnej liczby bloków (Samsung, 2007, STMicroelectronics, 2004).

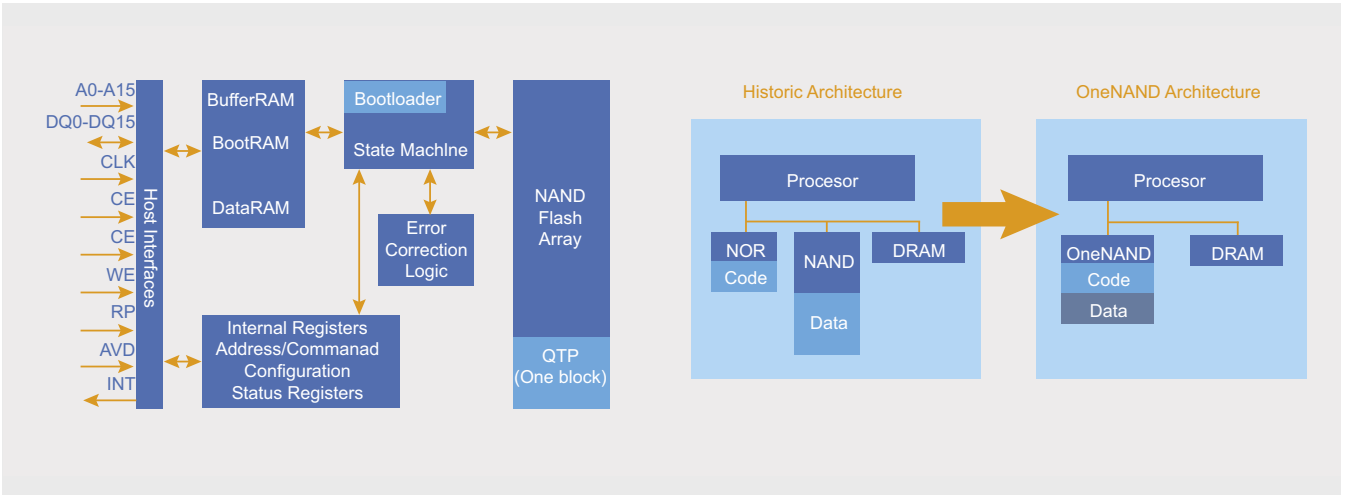
Do zarządzania nieważnymi blokami, producenci stosują unikalny zasady, ale odnoszą się do dwóch strategii: Pomiń Zły Blok (Skip Bad Block - SBB) i Rezerwuj Obszar Bloku (Reserve Block Area – RBA). W SBB, gdy zły blok zostanie wykryty system plików flash po prostu przeskakuje od razu do następnego dobrego bloku. W strategii RBA, wcześniej wydzie-



Rysunek 16. *Partycjonowanie dla SBB (z lewej) i RBA (z prawej) (White, 2008)*



Rysunek 17. *Kategorie NAND flash (BPM).*



Rysunek 18. *Układ OneNANDTM™ (z lewej) vs OneNANDTM ™ architektury (z prawej) (Samsung).*

lony jako rezerwa obszar, służy do dostarczania dobrych bloków w zamian za złe.

Strategia Pomiń Zły Blok (SBB) strategii i zagadnienia pokrewne

SBB mogą powodować przesunięcie pomiędzy fizycznym i logicznym rozmieszczeniem danych w pamięci urządzenia flash z więcej niż jednym LUN. SSB może również doprowadzić do ingresji blok, w przypadku gdy blok z partycji (B) jest przetwarzany przez usługi z poprzedniej i sąsiadującej partycji (A). Oznacza to, że będzie można mieć dwa bloki o tym samym numerze (BPMicrosystems, 2008, Breeuwsma et al., 2007).

Strategia Rezerwuj obszar bloku (RBA) i zagadnienia pokrewne

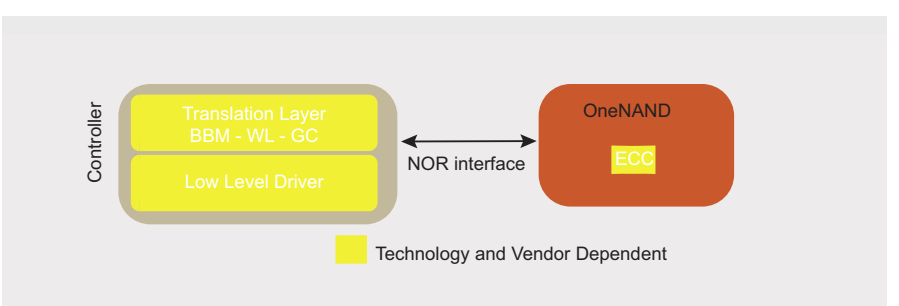
Przy wykorzystaniu RBA, partycjonowanie danych nie ma miejsca, urządzenie jest po prostu podzielony na obszar bloków przestrzeni użytkownika i obszar rezerwowy (BPMicrosystems, 2008, Samsung, 2006a). Odpowiednia tabela jest używana do odwzorowania uszkodzonych bloków na RBA. W przypadku utraty tej tabeli, powinno być możliwe odtworzenia jej przez odczytanie flag w dodatkowym obszarze wszystkich bloków - nawet jeśli niektórzy autorzy twierdzą że jest to niezwykle trudne (Inoue i Wong, 2004).

Surowy NAND i Zarządzany NAND

Gdy logika FTL i związane z nią funkcje są wbudowanych w NAND, to flash jest wtedy skategoryzowany jako zarządzany NAND, a gdy FTL jest pod opieką systemu plików hosta (logika jest zewnętrzna dla NAND), to mówi się o surowym NAND. Surowy NAND zawiera tylko tablice pamięci flash i kontroler P/E/R (Program/Erase/Read) (Pon et al., 2007). Do analizy sądowej, konieczne jest branie pod uwagę różnic między surowym i zarządzanym NAND, ze szczególnym uwzględnieniem skutków odzyskiwania obszaru i zarządzanie złymi blokami.

Ewolucja pamięci flash: Samsung OneNANDTM ™

W 2003 r. Samsung opracował nowy jednolity moduł pamięci flash do przechowywania kodu i danych: OneNANDTM ™. Urządzenie to posiada zarówno dużą prędkości odczytu danych z NOR Flash oraz wysoką prędkość zapisu i możliwości NAND Flash. W dniu pisania tego artykułu można przechowywać dane o pojemności do 16Gb w obszarze NAND. OneNAND posiada interfejs NOR, więc chipset wykryje OneNANDTM ™ jak NOR, podczas gdy dane mogą być przechowywane bezpośrednio w obszarze NAND za pomocą multipleksowane linii dostępu. OneNANDTM ™ jest sklasyfikowany jako surowy NAND z funkcją wewnętrznego ECC (Samsung, 2005b).



Rysunek 19. *Surowy OneNAND™*

SALVATORE FIORILLO
Konsultant ds. bezpieczeństwa
salvatore.fiorillo @ tesoro.it

Następny numer dostępny on-line ostatniego dnia sierpnia 2010