Internet do poprawki! Czekamy na Internet 2.0

1 grudnia 2011

Jesteśmy przyzwyczajeni do tego, że Internet jest i działa. Ale czy działa dobrze? Moje pokolenie wychowało się w czasie pierwszych komputerów, które nie były podłączone do sieci. Wraz z dostępem do Internetu pojawiła się możliwość oglądania prostych stron internetowych, korzystania z poczty czy grup dyskusyjnych. Wszystko działało w miarę dobrze do czasu upowszechnienia się Internetu w postaci, jaką znamy od niecałej dekady.

Dziś Internet wykorzystuje się do wszystkiego. Niestety, pierwotne standardy i protokoły sieciowe, stworzone do prostych celów, obrosły w rozszerzenia, dostosowujące ich działanie do potrzeb współczesnego użytkownika. Internet, jaki znamy, jest więc z pozoru bardzo nowoczesny, ale u podstaw oparty o przestarzałe technologie. To dlatego mamy problemy ze spamem, wirusami, przeciążonymi serwerami czy dostępem do wybranych stron. I chociaż nie każdy z nas te problemy zauważa, to i tak wszyscy płacimy za ten stan rzeczy. Ponosimy przecież koszty utrzymania całej infrastruktury i godzimy się na ograniczenia, wynikające z konstrukcji sieci.

Strony internetowe to dziś mieszanka różnych technologii, starych i nowych. Wszystkie mają wspólny mianownik: są niedopracowane i nie w pełni wdrożone. Podmioty tworzące przeglądarki internetowe zrzeszone są w organizacji W3C, zajmującej się tworzeniem standardów internetowych. Te same firmy i organizacje tak skutecznie wprowadzają standardy do swoich programów, że jeszcze żaden nie został w pełni zaimplementowany.

Zanim przeglądarki internetowe otrzymały pełną obsługę HTML4, wprowadzono XHTML. Ale i XHTML-a nie w pełni zaimplementowano,

ponieważ po kilku latach zarzucono pracę nad rozwojem tego języka, tworząc nowy język HTML5. W międzyczasie zaniechano prac nad wdrażaniem CSS2, zajmując się CSS3. W rezultacie współczesne przeglądarki obsługują tylko część rozwiązań, zawartych w wymienionych standardach i są między sobą niekompatybilne.

Twórcy stron internetowych muszą się mocno nagimnastykować, aby tworzona przez nich strona wyglądała i działała tak samo pod każdą z przeglądarek. Wiele rozwiązań jest zarzucanych z powodu fragmentarycznej obsługi po stronie użytkownika. Przykład? Praktycznie wszystkie współczesne przeglądarki przechodzą test Acid2 i Acid3 – zestaw skryptów, sprawdzający kompatybilność ze standardami W3C. Mimo to żadna z przeglądarek nie obsłuży tak prostej rzeczy, jak zmiana koloru czy wielkości elementu checkbox – podstawowego pola formularza HTML. Na cztery sprawdzone przeze mnie przeglądarki żadna tego nie obsługuje. Firefox i Chrome mają całkowicie pominiętą obsługę parametryzowania tego elementu, a Internet Explorer oraz Opera pozwalają tylko na dodanie obramowania (parametr „border”), przy czym nie działa to poprawnie, ponieważ w Operze obramowanie wchodzi do środka pola formularza, a w IE grubsza ramka wypełnia całe wnętrze elementu. Jeśli ktoś nie wierzy, może sprawdzić sam. Powstała nawet strona pokazująca, na jakim etapie są przeglądarki internetowe, jeśli chodzi o obsługę pola checkbox (link), wersje sprawdzonych tam programów są nieco starsze niż współczesne, ale zapewniam, że do dziś niewiele się zmieniło w tym temacie.

Mnogość standardów, które wciąż są w użyciu, powoduje, że przeglądarki pracują ociężale, ponieważ muszą obsługiwać wiele standardów jednocześnie. Współczesna przeglądarka musi być przygotowana na obsługę pięciu kolejnych wersji HTML plus XHTML oraz obsługiwać tzw. quirks mode (z ang. tryb dziwactw), czyli niepisany standard interpretowania kodu HTML, pochodzący z czasów, gdy strony tworzyło się bez zwracania uwagi na poprawność składni. W efekcie współczesne strony, te

największe, działają ociężale na wolniejszych komputerach typu netbook. Na moim starszym laptopie, posiadającym procesor 1.8 GHz, nasze portale oraz serwisy społecznościowe wyświetlają się i przewijają bardzo wolno. Sama obsługa CSS, przy mocniejszym jego wykorzystaniu w kodzie strony, powoduje niepłynne przewijanie zawartości i powolne odrysowywanie się układu strony. Widać to wyraźnie na stronach z poradnikami CSS. Gdy dołączymy do tego obsługę JavaScriptu i Flasha, który potrafi zająć większość pracy procesora, okazuje się, że niektóre komputery stają się po prostu za wolne do korzystania ze współczesnego Internetu. Same przeglądarki z wersji na wersję wydają się być coraz gorzej zoptymalizowane, działają wolniej, potrzebują więcej pamięci niż kiedyś. Gdzie te czasy, gdy przeglądarka uruchamiała się w ułamek sekundy, a wciśnięcie „wstecz” powodowało błyskawiczne załadowanie poprzedniej strony?

AJAX to technologia, która w teorii odciąża Internet od ciągłego przeładowywania stron, poprzez doładowywanie treści na bieżąco. W praktyce powoduje, że obsługa stron działa wolniej. Można to sprawdzić samemu. Każdy z liczących się serwisów posiada wersję statyczną, uruchamianą przy detekcji przeglądarki, która ma wyłączoną obsługę JavaScriptu. Proponuję wyłączyć JS i wejść na taką stronę. Efekt – podstrony przeładowują się błyskawicznie. Przykład: statyczna wersja poczty Gmail. Przełączanie się między mailami trwa ułamek sekundy, podczas gdy w pełnej, AJAX-owej wersji Gmaila trwa to zauważalnie dłużej. Inny przykład to NK.pl. Tu również przy wyłączonym JS wszystko otwiera się błyskawicznie. Nie wszystko będzie działało, ale przeglądanie profilów i oglądanie zdjęć działa tak szybko, jak to tylko możliwe. Tam, gdzie AJAX powoduje przystopowanie ładowania treści, wersja statyczna daje nam natychmiastowy wynik.

Mimo wciąż rosnącej przepustowości łącz, same strony objętościowo ładują się coraz dłużej. Dlaczego? Dziś każda większa strona, oprócz prezentowania treści, łączy się z

wieloma innymi serwerami, które doładowują skrypty do statystyk czy reklamy bannerowe, wydłużając czas ładowania strony. Pogoń za implementacją standardów spowodował, że typowa strona posiada więcej kodu HTML, CSS i JavaScript, niż treści, do prezentacji której została stworzona.

Aby powiązać domenę internetową ze stroną, należy przygotować serwer pod przyjęcie domeny i skonfigurować adresy DNS. Następuje wtedy propagacja domeny, proces trwający czterdzieści osiem godzin, w trakcie których serwery DNS z całego świata otrzymują informację o adresie nowej domeny. Gdy informacja dotrze do wszystkich serwerów, dopiero wtedy każdy użytkownik Internetu będzie mógł dostać się na naszą stronę. Proces ten jest technicznie skomplikowany, trwa bardzo długo i bywa awaryjny. Kilka tygodni temu propagowanie jednej z moich domen trwało tak długo, że użytkownicy niektórych sieci nie mieli dostępu do mojej strony przez ponad tydzień. Opóźnienie odnotowałem m.in. przy próbach połączenia się z Neostrady i Internet DSL od TP SA, a także z Play Online. Dlaczego coś, co powinno trwać do 48 godzin (dla świata, bo dla Polski do 12 godzin) trwało ponad 200 godzin – nie wiadomo. Należy przyjąć, że była to po prostu jedna z wielu niezidentyfikowanych awarii Internetu, na które po prostu nie ma rady.

W tym tygodniu wystąpił problem z wymianą poczty na styku Google i Allegro. Sprzedawcy i kupujący, którzy posiadali skrzynki pocztowe w usłudze Gmail, nie dostawali korespondencji z serwisu aukcyjnego. Problem dość duży, ponieważ serwis mailowo informuje obie strony transakcji o każdy etapie sprzedaży, od informacji o zakupie do powiadomienia o wpłacie pieniędzy. Po prostu w pewnym momencie serwery Google’a bez powodu zaczęły odrzucać maile wysyłane z Allegro. Po kilku dniach sytuacja została opanowana, ale do tego czasu proces komunikacji między serwisem a klientami był wstrzymany.

Z pocztą elektroniczną wiąże się jednak kilka większych problemów. Przede wszystkim nie ma jakiejkolwiek autoryzacji

wysyłanej korespondencji, przez co większość krążących w sieci maili to spam. Każdy może wysłać maila z dowolnym adresem nadawcy, przy czym większość użytkowników Internetu nie ma o tym pojęcia, więc przestępcom łatwo jest podszyć się pod inną osobę czy instytucję i dokonać wyłudzenia. Spamerzy, wykorzystując miliony zainfekowanych komputerów, dysponują ogromną siecią spambotów, które mogą zablokować każdą skrzynkę pocztową. Nazywa się to atakiem „joe-job", który polega na tym, że jako nadawcę maila podaje się adres konkretnego użytkownika. Następnie wysyłane są masowo maile na losowe adresy, a ponieważ większość z nich nie istnieje, na adres bogu ducha winnego „nadawcy" przychodzą dziesiątki tysięcy zwrotów. Efekt – mimo że właściciel konta nie wysłał żadnego maila, jego skrzynka zostaje zapchana i zablokowana. I nie ma przed tym obrony.

Nadawcy nie mają również żadnej pewności, że ich mail dotrze do adresata. Zdarza się to rzadko, ale wiele razy miałem do czynienia z sytuacją, że mail po prostu nie dotarł, przy czym nie otrzymałem żadnego zwrotu. Problemem są również filtry antyspamowe, które powodują, że owszem, niechcianej korespondencji dostajemy nieco mniej, ale kosztem tego, że od czasu do czasu nie dostaniemy ważnego maila, ponieważ zostanie on zakwalifikowany jako spam. W pewnych sytuacjach, gdy serwis korzysta z potwierdzania transakcji poprzez wysyłanie linka weryfikacyjnego, uniemożliwia to dostęp do usług. W takiej sytuacji znalazły się osoby, które swego czasu zamawiały online bilety na mecze siatkówki – do spamu w ich skrzynkach pocztowych trafiały maile z linkiem weryfikacyjnym, potrzebnym do aktywowania dostępu do platformy sprzedażowej.

Ściągnięcie większej ilości materiałów z FTP-a nie może się obyć bez kilkakrotnego zerwania połączenia. Pobieranie plików ze stron internetowych również bywa uciążliwe, gdy w połowie ściągania nastąpi zerwanie połączenia i często nie ma możliwości wznowienia operacji, pozostaje jedynie zainicjowanie ściągania od nowa. Połączenia z protokołem FTP

we współczesnych sieciach to również przykład na to, jak nierówno traktowani są użytkownicy Internetu. Podczas gdy transfer przez protokół HTTP, czyli zwykłe połączenie ze stroną internetową, odbywa się z dużą prędkością, połączenia z FTP mają zazwyczaj przycięte pasmo przesyłu danych. Dlaczego? Nie wiadomo.

Struktura internetu umożliwia również prostą zabawę z blokowaniem serwisów internetowych. Każdy serwer posiada ograniczoną przepustowość, wystarczy więc większa aktywność na stronie i dostęp do witryny zostanie wyłączony. W złośliwych akcjach jest to nazywane atakiem DDoS. Ale serwisy padają również ofiarą zwykłej popularności. Większość serwerów, na których utrzymywane są współczesne strony, nie wytrzyma ruchu, jaki nastąpi po podaniu adresu strony w popularnym programie telewizyjnym. Duża ilość osób, które w tym samym momencie chcą dostać się na stronę, powoduje, że serwis na długie godziny przestaje być dostępny. Rozwiązaniem jest jedynie przewidywanie takich sytuacji i wykupowanie dziesięć razy mocniejszych serwerów po to, aby kilka dni w roku wytrzymały większy ruch niż zazwyczaj.

Współczesny Internet to również niebezpieczeństwo zarażenia komputera wirusem. Kilka wersji przeglądarek wstecz wystarczyło wejść na stronę ze złośliwym kodem, by bez naszej wiedzy został zainstalowany wirus na naszym komputerze. Luki w systemie Windows XP w czasach, gdy nie było jeszcze dodatku SP1 do tego systemu, powodowały, że komputer z nowo zainstalowanym system operacyjnym tuż po podłączeniu go do sieci od razu otrzymywał porcję wirusów, które blokowały dostęp do Internetu.

Twórcy oprogramowania są zawsze kilka kroków wstecz względem zagrożeń płynących z Internetu. Popularny język skryptowy PHP został przygotowany tak, że pierwotnie nie przewidziano w nim ochrony przez atakami SQL Injection, a wewnętrzna instrukcja „mail”, służąca do wysyłania poczty z poziomu skryptu, została tak opracowana, że jej użycie zgodnie z podręcznikiem oznacza

otworzenie furtki dla robotów, które wykorzystają ją do rozsyłania spamu. Typowy skrypt do forum, po umieszczeniu na popularnej stronie, z miejsca potrzebuje albo programisty, który będzie systematycznie łatał skrypty, albo moderatora, który będzie poświęcał czas na usuwanie pojawiającego się tam spamu.

Można mnożyć zarzuty pod adresem konstrukcji Internetu. Jego funkcjonowanie jest genialne i jednocześnie oparte na najgorszych z możliwych rozwiązań. Dziś, gdyby ktoś chciał opracować Internet 2.0, na pewno sieć działałaby inaczej. Wiele rzeczy można byłoby opracować lepiej, na przykład usprawnić propagację domen czy zapewnić nieskończoną pulę adresów IP, tak aby fizyczne adresy były dobrem dostępnym dla każdego, a nie towarem reglamentowanym za opłatą. Również język do tworzenia stron mógłby zostać stworzony raz a porządnie. Ale dziś wielu błędów nie da się naprawić, ponieważ rozsiane są one w milionach miejsc na świecie. Stworzenie Internetu 2.0 i wprowadzenie go do powszechnego użytku oznaczałoby albo odcięcie się od obecnej infrastruktury, co spowodowałoby technologiczny chaos, albo drogie utrzymywanie dwóch struktur jednocześnie i powolne zastępowanie jednego systemu drugim.

Czy doczekamy się Internetu 2.0? Raczej nic na to nie wskazuje. Pozostaje pogodzić się z faktem, że musimy utrzymywać dużą liczbę informatyków, serwisantów i moderatorów oraz dopłacać do nadmiarowego sprzętu i dodatkowego transferu, aby obsłużyć ruch generowany przez spam, błąkające się po łączach pakiety oraz zrywające połączenia.

Autor: Sławomir Wilk
Źródło: Dziennik Internautów