MONITOROWANIE SIECI I FILTROWANIE TREŚCI

WPROWADZENIE.

Internet nazywany jest szambem, czasami w odniesieniu do liczby zainfekowanych wirusami i kontrolowanych przez hakerów maszyn, ale częściej w odniesieniu do ilości budzących zastrzeżenia treści dostępnych za jednym kliknięciem myszy. W tym rozdziale omówiono działania mające na celu monitorowanie i kontrolowanie dostępu do niektórych treści. Aplikacje, które wykonują tego rodzaju działania, budzą kontrowersje: zwolennicy prywatności i wolności słowa regularnie odwołują się do oprogramowania cenzurującego, podczas gdy autorzy takiego oprogramowania zwykle używają terminu filtrowanie treści. W tym rozdziale zastosowano filtrowanie treści, bez potrzeby zajmowania się tym argumentem. Więcej informacji na temat zasad i kwestii prawnych. W tym rozdziale omówiono pokrótce możliwe motywacje, które doprowadziły do podjęcia decyzji o filtrowaniu treści, bez debaty na temat zasadności tych motywów. Biorąc pod uwagę różnorodność dobrych i złych powodów, dla których należy monitorować i filtrować zawartość sieci Web, w tym rozdziale omówiono różne techniki filtrowania, a także niektóre sposoby pokonania monitorowania i filtrowania.

NIEKTÓRE TERMINOLOGIE

Proxy - komputer, który uczestniczy w komunikacji w imieniu klienta. Serwer proxy odbiera żądanie klienta, a następnie generuje je ponownie z własnym adresem serwera proxy jako adresem źródłowym. W związku z tym serwer widzi tylko informacje identyfikacyjne z serwera proxy, a tożsamość klienta pozostaje ukryta (przynajmniej z punktu widzenia adresowania sieciowego). Serwery proxy są szeroko stosowane przez organizacje do kontrolowania wychodzącego ruchu (głównie ruchu internetowego) i do ochrony użytkowników przed bezpośrednim połączeniem z potencjalnie szkodliwymi witrynami internetowymi.

Anonimizacja proxy-aproxy, która umożliwia użytkownikom ukrywanie ich aktywności w Internecie. Zazwyczaj taki serwer proxy znajduje się poza granicami organizacji i często jest używany do obchodzenia reguł filtrowania.

Technologie zwiększające prywatność (PET) - klasa technologii, która pomaga użytkownikom zachować prywatność korzystania z sieci. Obejmują one szyfrowanie, anonimizację serwerów proxy, mieszanie sieci i routing cebulowy.

Szyfrowanie-odwracalne zniekształcanie tekstu (zwykłego tekstu) na losowo wyglądające dane (tekst zaszyfrowany), których nie można odwrócić bez użycia tajnych informacji (kluczy).

Użytkownicy sieci mieszanych szyfrują swoje wiadomości wychodzące kluczem publicznym odbiorcy (lub adresatów), a następnie także kluczem publicznym „serwera mieszania”. Zaszyfrowana wiadomość jest wysyłana do serwera Mix, który działa jako broker wiadomości i odszyfrowuje kryptogram w celu ustalenia prawdziwego adresu odbiorcy i przekazania wiadomości.

Onion routing - anonimowy i bezpieczny protokół routingu opracowany przez Naval Research Laboratory Center for High Assurance Computer Systems w latach 90. Wiadomości są wysyłane przez sieć (chmurę) routerów cebulowych, które komunikują się ze sobą za pomocą kryptografii klucza publicznego. Po ustanowieniu obwodu tymczasowego nadawca szyfruje wiadomość wychodzącą za pomocą każdego z kluczy publicznych wszystkich routerów cebulowych w obwodzie. W szczególności architektura zapewnia dwukierunkową komunikację, mimo że nikt oprócz serwera proxy inicjatora nie wie nic poza poprzednimi i następnymi przeskokami w łańcuchu komunikacyjnym. Oznacza to, że ani respondent, ani jego serwer proxy, ani żaden zewnętrzny obserwator nie muszą znać tożsamości inicjatora lub jego serwera proxy.

Implementacją routingu cebulowego trzeciej generacji jest Tor (router cebulowy).

MOTYWACJA.

Zgodnie z ogólną koncepcją, osoba lub grupa decyduje się na monitorowanie aktywności sieciowej i filtrowanie treści na podstawie relacji między władzami. Najczęstsze relacje prowadzące do monitorowania i filtrowania to:

* Rządy kontrolujące swoich obywateli

* Rodzice i szkoły chroniące swoje dzieci

* Organizacje egzekwujące swoje zasady

* Rządy egzekwujące swoje prawa

* Rządy monitorujące terrorystów i inne państwa narodowe

Zapobieganie niezgodzie

Prawdopodobnie najczęstszym powodem sprzeciwu wobec filtrowania treści jest fakt, że wiele represyjnych rządów filtruje treści, aby zapobiec sprzeciwowi. Jeśli obywatele (lub podmioty) nie mogą uzyskać dostępu do informacji, które albo źle odbijają się na rządzie, albo kwestionują filozoficzne lub religijne doktryny kraju, wówczas obywatele prawdopodobnie nie zdadzą sobie sprawy, w jakim stopniu są represjonowani. W krajach takich jak Stany Zjednoczone, gdzie Konstytucja gwarantuje wolność słowa i prasy, wiele osób uważa, że tylko ci, którzy mają coś do ukrycia w swojej działalności, będą walczyć z cenzurą. To pojęcie, w połączeniu z budzącymi zastrzeżenia treściami, tak łatwo dostępnymi w Internecie, może być powodem, dla którego próby zakwestionowania stosowania oprogramowania filtrującego w bibliotekach i innych miejscach publicznych spotkały się z oporem lub obojętnością. Te przykłady powinny być przestrogą dla czytelników w krajach, które są obecnie bardziej liberalne w polityce informacyjnej.

Ochrona dzieci.

Z tych samych powodów, dla których sklepy ogólnospożywcze, przynajmniej częściowo, zakrywają czasopisma dla dorosłych, rząd zażądał, aby niektóre rodzaje informacji w Internecie były niedostępne dla dzieci w szkołach publicznych. Wynika to z postrzegania roli szkoły jako zastępczego rodzica i odpowiedzialności rządu za ewentualne niepowodzenie w tej roli. Różnorodne przepisy wymagały stosowania monitorowania i filtrowania treści w szkołach i bibliotekach, zgodnie z teorią, że takie publiczne lub publicznie wspierane terminale nie powinny wykorzystywać pieniędzy podatników do dostarczania dzieciom treści budzących zastrzeżenia. Sąd Najwyższy Stanów Zjednoczonych orzekł, że niektóre z tych wysiłków są niezgodne z konstytucją po skrajnych protestach bibliotek. Ostatnio ustawa o ochronie dzieci w Internecie (CIPA) 4 wymagała od okręgów szkolnych, które otrzymują określone rodzaje funduszy rządowych, stosowania technologii filtrujących. Większość szkół wdrożyła filtrowanie sieci, agresywnie ograniczając zawartość sieci Web dostępną dla uczniów. Szacunki dotyczące ilości zablokowanych treści są bardzo zróżnicowane, ale jeden okręg szkolny twierdzi, że filtruje około 10 procent całego ruchu internetowego ze względu na wątpliwe treści. Oczywistym celem technologii filtrującej w szkołach jest uniemożliwienie uczniom oglądania materiałów uważanych za szkodliwe dla nieletnich. Innym, mniej rozpowszechnionym powodem filtrowania zawartości sieci Web jest uniemożliwienie uczniom robienia w sieci rzeczy, których mogliby nie robić, gdyby wiedzieli, że ktoś ją obserwuje. Nadzieja polega na tym, że uczniowie nie będą karani, zanim ukończą szkołę średnią, chroniąc ich przed złym osądem, gdy uczą się, jak dobrze oceniać i poznawać zasady obowiązujące w społeczeństwie. Zapewne jest to praca rodzica, a nie szkoła, ale szkoły zapewniające dostęp do

Internetu muszą jednak zapewniać tego rodzaju ochronę. Uczniowie szkół średnich i przedszkolaki na ogół mają różne poziomy dojrzałości, a rozsądna polityka filtrowania obejmowałaby elastyczny, stopniowy stopień filtrowania w zależności od wieku. Szkoły mogłyby również stosować mniej inwazyjne metody kontrolowania dostępu do treści internetowych, takie jak bezpośredni nadzór nad uczniami korzystającymi z komputerów. Rodzice mają również możliwość filtrowania treści, które ich dzieci oglądają w domu.

Wspieranie polityki organizacyjnej w zakresie zasobów ludzkich.

Organizacje mają różne powody, dla których monitorują i filtrują zawartość sieci Web, do której mają dostęp ich pracownicy. Najprostsza to chęć, aby podczas pracy pracownicy wykonywali czynności związane z pracą. Pomimo badań wskazujących, że elastyczność w prowadzeniu ograniczonej działalności osobistej, takiej jak bankowość lub poczta elektroniczna, z pracy skutkuje szczęśliwszymi, bardziej produktywnymi pracownikami, menedżerowie czasami postrzegają osobiste korzystanie z komputerów służbowych jako kradzież czasu dla firmy. Bardziej pragmatycznym powodem filtrowania niektórych treści w sieci jest zapobieganie odpowiedzialności „wrogiego miejsca pracy" na mocy przepisów dotyczących równych szans w zatrudnieniu. Problem z tym podejściem polega na tym, że wiele rodzajów treści mogłoby być potencjalnie obraźliwych dla współpracowników, dlatego trudno jest zastosować technologię filtrowania, aby zagwarantować, że nikt nie zostanie obrażony, a jednocześnie umożliwia rozsądne korzystanie z Internetu w celach zawodowych. Niektóre organizacje wolą monitorować ruch sieciowy za pomocą zautomatyzowanych systemów zamiast blokować cokolwiek i powiadamiać użytkowników o monitorowaniu. Powiadomienie może znajdować się w ogólnych dokumentach dotyczących zasad lub w postaci wyskakującego okienka informującego, że użytkownik ma zamiar przeglądać treści, które mogą naruszać zasady. Tak czy inaczej, chodzi o to, że organizacja może uniknąć odpowiedzialności, ostrzegając użytkownika, ale może również uniknąć skarg dotyczących prywatności i wolności słowa. Podejście monitoruj i powiadamiaj również wysyła wiadomość, że organizacja ufa swoim pracownikom, ale chce utrzymać pozytywne i produktywne środowisko pracy.

Egzekwowanie prawa.

Organy ścigania rzadko angażują się w filtrowanie treści, ale monitorowanie ruchu jest często wykorzystywanym narzędziem do badania przestępstw komputerowych. Gromadzenie dowodów na potrzeby tych dochodzeń często wiąże się z przechwytywaniem ruchu, który faktycznie dociera do celu, więc filtrowanie przyniosłoby efekt przeciwny do zamierzonego. W tym przypadku udowodnienie tożsamości jest kluczem do uzyskania użytecznych dowodów, więc technologie zwiększające prywatność są prawdziwymi problemami. W niektórych przypadkach dzienniki serwerów proxy sieci Web, które identyfikują rzeczywisty adres źródłowy komputera klienckiego, są dostępne wraz z wezwaniem do sądu. Dochodzenia tego rodzaju często dotyczą pornografii dziecięcej, produkcji narkotyków lub kradzieży sprzętu komputerowego chronionego oprogramowaniem do odzyskiwania zasobów.

Nadzór bezpieczeństwa narodowego.

Podczas gdy rządowe organy ścigania pracują nad kontrolą działań, które naruszają krajowe prawo, inne agencje rządowe są zaniepokojone rozwojem technologii informacyjnych i komunikacyjnych jako narzędzi terroru i konfliktów międzynarodowych. Od groźby ewentualnej wojny cybernetycznej po bardziej bezpośrednie obawy związane z terroryzmem, rządy badają sposoby wykrywania wykorzystania poczty elektronicznej, telefonów komórkowych, platform mediów społecznościowych, a nawet gier online do komunikowania się i koordynowania planów wrogich działań. Monitorowanie tych mediów stało się pilnym, kosztownym i ryzykownym priorytetem dla rządów krajowych. W czasie,

gdy to wydanie miało się ukazać, wyciekły tajne informacje o działaniach Agencji Bezpieczeństwa Narodowego rządu USA w zakresie monitorowania komunikacji, ożywiając odwieczną debatę na temat równowagi między prywatnością a bezpieczeństwem w republice demokratycznej. Pomimo wieloletniego precedensu i zapewnień o konstytucyjnej ochronie przed nierozsądnymi poszukiwaniami, rewelacje dotyczące PRISM NSA i powiązanych programów dały jasno do zrozumienia, że komunikacja sieciowa jest uczciwą grą w wysiłkach narodu, aby pozostać o krok przed swoimi przeciwnikami (niezależnie od tego, czy są one prawdziwe , potencjalne, a nawet wyimaginowane) .6 Ponadto, jak odkryli obywatele USA, rządowi bardzo trudno jest skutecznie podsłuchiwać wrogie zagraniczne komunikaty elektroniczne bez gromadzenia wszystkich komunikatów, narażając się na naruszenie oczekiwań obywateli i sojuszników w zakresie prywatności. zarówno.

TECHNIKI OGÓLNE.

Filtrowanie komunikacji sieciowej może obejmować dwie podstawowe taktyki: badanie metadanych i badanie faktycznej treści wiadomości. Amerykańskie media ponownie odkryły koncepcję metadanych w następstwie wycieku PRISM, ponieważ przedstawiciele NSA nalegali, że rzeczywista treść wiadomości zostanie zbadana tylko po odfiltrowaniu nieistotnej komunikacji poprzez sprawdzenie numerów telefonów, czasu trwania połączeń, adresów IP i innych informacji transakcyjnych który ujawnia źródło, cel i czasowy kontekst rozmowy. Oparty na przytłaczającym natężeniu ruchu komunikacyjnego przechodzącego przez kable miedziane, kable światłowodowe i częstotliwości bezprzewodowe, podejmowanie decyzji filtrujących (prawnych lub wojskowych) w oparciu o badanie każdego słowa każdej wiadomości jest rzeczywiście zniechęcającym zadaniem. Dla większości celów znacznie skuteczniejsze jest przynajmniej rozpoczęcie od tych metadanych i przyjęcie założeń dotyczących znaczenia lub niebezpieczeństwa określonego komputera, człowieka lub partnera komunikacyjnego w organizacji. Bezpośrednie badanie pełnej treści wiadomości wymaga znacznie więcej zasobów, chociaż poczyniono postępy w skutecznym i wydajnym automatycznym przetwarzaniu tekstu i obrazów.

Dopasowanie żądania.

Najprostszą techniką stosowaną w technologiach filtrujących jest dopasowywanie ciągów znaków do list słów kluczowych. Każde żądanie sieci Web korzysta z jednolitego lokalizatora zasobów (adresu URL) w ogólnej postaci:

protokół: //server.organization.top-level-domain/path-to-file.file-format

Filtrowanie żądania adresu URL może sprawdzić dowolną część ciągu pod kątem dopasowania do zabronionych ciągów:

* Filtry mogą pasować do pola protokołu, aby wymusić zasady korzystania z zaszyfrowanego ruchu internetowego (HTTP i HTTPS). Jest to bardziej ogólny problem z bezpieczeństwem niż problem z filtrowaniem sieci Web, chociaż organizacja martwiąca się koniecznością analizowania całego ruchu może uniemożliwić korzystanie z zaszyfrowanego ruchu internetowego poprzez blokowanie wszystkich żądań HTTPS.

* Pola serwera i organizacji opisują, kto hostuje zawartość. Filtrowanie oparte na tych ciągach jest podejściem szerokim, ponieważ prowadzi do blokowania całego serwera WWW lub ruchu sieciowego całej organizacji.

* Pole domeny najwyższego poziomu może służyć do filtrowania treści

* Pole ścieżka do pliku zawiera rzeczywisty tytuł żądanej strony internetowej, więc różni się on najbardziej w poszczególnych żądaniach. To, czy istnieje większe prawdopodobieństwo niż inne pola zawierające informacje przydatne w filtrowaniu, zależy od konwencji nazewnictwa serwera. To pole (i pole formatu pliku) jest opcjonalne, jak pokazano w żądaniu witryny www.wiley.com, które kieruje serwer do wyświetlenia strony domyślnej.

* Pole formatu pliku informuje przeglądarkę internetową, jak postępować z tekstem wyświetlanym na stronie, czy to w prostym formacie html, czy zakodowanym w innym formacie pliku (np. Doc, pdf), lub czy zezwolić na dynamiczne generowanie treści (np. , asp). Niewiele produktów filtrujących używa tego pola do filtrowania tradycyjnych rodzajów treści budzących zastrzeżenia, chociaż wymusza inne zasady dotyczące kodu dynamicznego w przypadku wysokiego poziomu bezpieczeństwa środowiska mogą wymagać dopasowania tego pola.

Dopasowywanie hosta.

Niektóre systemy filtrujące próbują odróżnić akceptowalne i niedopuszczalne źródła w sieci WWW, sprawdzając określone serwery lub ogólne portale informacyjne, takie jak wyszukiwarki.

Listy blokowe serwerów.

Kontrowersyjne treści zwykle koncentrują się na pojedynczych serwerach. To naturalnie prowadzi niektóre organizacje do blokowania dostępu do tych serwerów. Dwie metody blokowania serwerów to adres protokołu internetowego (IP) i nazwa. Blokowanie na podstawie adresu IP po prostu odrzuca cały ruch (lub cały ruch HTTP) do iz określonych adresów. Ta taktyka wiąże się z kilkoma trudnościami w zależności od tego, jakie są adresy używane. Po pierwsze, adresy IP nie są trwałe. Chociaż adresy numeryczne większości dużych serwerów komercyjnych zwykle pozostają niezmienione w czasie, wiele mniejszych serwerów ma dynamicznie przypisywane adresy, które mogą się okresowo zmieniać. W związku z tym blokowanie adresu IP może uniemożliwić dostęp do treści hostowanych na zupełnie innym serwerze niż zamierzono. Po drugie, serwery komercyjne często udostępniają treści dla wielu różnych klientów, a blokowanie całego serwera spowoduje zablokowanie całej zawartości, a nie tylko tych, które budzą zastrzeżenia. Jest to szczególny problem w przypadku bardzo dużych dostawców usług, takich jak AOL, który zapewnia każdemu użytkownikowi możliwość hostowania witryny internetowej. Blokowanie serwerów AOL z powodu nieodpowiedniej zawartości mogłoby spowodować nadmierne zablokowanie dużej liczby osobistych stron internetowych. Po trzecie, blokowanie na podstawie adresów stwarza możliwość złośliwego wpisu, praktyki, w której osoba atakująca (lub konkurent) fałszuje adres serwera internetowego i dostarcza treści budzące zastrzeżenia, które mogą wylądować serwer na listach blokujących. Niektóre produkty filtrujące umożliwiają użytkownikom przesyłanie „złych" witryn, co stanowi kolejną okazję do umieszczenia złośliwego wykazu. Blokowanie według nazwy obejmuje system nazw domen (DNS), w którym nazwa czytelna dla człowieka (np. Www.wiley.com) jest odwzorowywana na czytelne dla komputera adresy IP (w tym przypadku 208.215.179.146). DNS umożliwia organizacji zmianę adresu fizycznego jej serwera internetowego poprzez aktualizację listy DNS tak, aby wskazywała na nowy adres IP, chociaż zależy to od systemu jako całości propagującego zmianę w rozsądnym czasie. Urządzenie, które monitoruje lub filtruje ruch na podstawie nazwy domeny, musi mieć możliwość okresowego odświeżania listy mapowań nazwa-adres, aby uniknąć blokowania witryn, których adresy okresowo się zmieniają. Podobnie jak w przypadku blokowania opartego na adresach, blokowanie oparte na nazwach również grozi złośliwym wyświetlaniem, a także blokowaniem zarówno zbyt niskim, jak i nadmiernym. Wiele organizacji rejestruje wiele nazw swoich serwerów z różnych powodów. Na przykład organizacja może zarejestrować swoją nazwę w domenach najwyższego poziomu .net, .com, .org i .biz, aby zapobiec błędom opisanym w sekcji 31.44 (whitehouse.com). Inne powody

rejestrowania nazwy w wielu domenach obejmują zapobieganie wykorzystywaniu przez konkurentów rejestracji nazwy do kradzieży klientów oraz zapobieganie spekulacyjnemu kupowaniu podobnych nazw przez osoby, które chcą je sprzedać z dużym zyskiem. Firmy hostingowe świadczą również usługi dla wielu różnych organizacji, więc różne adresy URL wskazywałyby na adres IP tego samego serwera hostingowego. W związku z tym blokowanie według nazwy serwera może powodować niedblokowanie, nieuwzględniając wszystkich możliwych zarejestrowanych nazw wskazujących na ten sam serwer, i nadmierne blokowanie przez dopasowanie nazwy serwera dostawcy usług, który obsługuje witryny dla wielu różnych klientów, którzy używają nazwy serwera dostawcy w adresie URL ich strony internetowe.

Blokuj / modyfikuj pośredników.

Sieć stała się olbrzymim repozytorium informacji, wymagającym opracowania potężnych narzędzi wyszukiwania, aby znaleźć informacje. Wczesne wyszukiwarki ustąpiły miejsca bardziej wyrafinowanym portalom informacyjnym, takim jak Yahoo !, Google, AOL i MSN. Umożliwiając zaawansowane wyszukiwanie i spersonalizowane wyniki, portale te zapewniają użytkownikom łatwy dostęp do informacji, które byłyby trudne lub niemożliwe do znalezienia przy użyciu technik wyszukiwania ręcznego. Portale stały się niezwykle popularnymi narzędziami dostępu do Internetu jako całości; Pod koniec 2000 r. Google odnotowywał 100 milionów zapytań dziennie; Szacuje się, że w 2013 r. indeks jego witryn zawierał ponad 48 miliardów stron8. Dostęp do informacji w tej skali sprawia, że portale są naturalnymi celami monitorowania i filtrowania. Niewiele organizacji komercyjnych uniemożliwia swoim pracownikom korzystanie z popularnych portali, ponieważ stali się oni częścią sposobu, w jaki ludzie korzystają z Internetu. Jednak niektóre kraje zablokowały dostęp do niektórych portali swoim obywatelom, mając nadzieję na kontrolowanie dostępu do informacji, które mogą naruszać przepisy krajowe (np. Dostęp do nazistowskich pamiątek we Francji) lub wzbudzić sprzeciw obywateli wobec rządu (np. Dostęp do informacje o masakrze na placu Tiananmen w Chinach).

Dopasowanie domeny.

Od 2004 r. do ostatecznej decyzji w 2011 r. Internet Corporation for Assigned Names and Numbers (ICANN) rozpatrzyła i odrzuciła wnioski o nową domenę najwyższego poziomu, .xxx, która umożliwiłaby dostawcom treści o charakterze seksualnym dobrowolną ponowną rejestrację swoich witryn. Twierdzono, że taka domena byłaby łatwa do przefiltrowania w adresie URL żądania, co prawdopodobnie spodobałoby się zwolennikom filtrowania w sieci. Umożliwiłoby to również dostawcom treści wykazanie, że przestrzegają przepisów uniemożliwiających dzieciom dostęp do nieodpowiednich materiałów, umożliwiając skuteczniejsze filtrowanie przez rodziców. Niemniej jednak posunięcie napotkało opór na obu frontach. Konserwatywne grupy religijne obawiały się, że ustanowienie domeny .xxx usankcjonowałoby pornografię, podczas gdy nie wszyscy dostawcy treści seksualnych zgodzili się, że postrzegane korzyści przeważą albo zwiększone filtrowanie ich witryn, albo łatwiejsze monitorowanie ruchu klientów. ICANN odrzucił kilka zmian propozycji domeny .xxx na przestrzeni lat, powołując się na brak jednomyślności w społeczności dostawców treści erotycznych, a także obawę, że ICANN może zostać postawiony w pozycji regulującej zawartość, która jest poza statutem organizacji. Ostatecznie jednak w 2011 roku domena .xxx została zatwierdzona.

Dopasowywanie treści.

Dopasowywanie łańcuchów jest proste, ale trudne do wykonania bez nad- i podblokowania. Na przykład jedną z najpopularniejszych kategorii filtrowania treści (szczególnie w Stanach Zjednoczonych) jest seks. Zablokowanie wszystkich treści dokładnie pasujących do słowa „seks" nie pozwoliłoby dopasować słów „seksowny" i „seksualny". Aby uniknąć tego rodzaju podblokowania, listy

słów muszą być bardzo długie, aby uwzględnić wszystkie permutacje. Nieco skuteczniejszą taktyką jest blokowanie wszystkich prac zawierających ciąg „seks", ale spowodowałoby to zablokowanie słów „Essex", „Sussex" i „aseksualny". Poszukiwanie wszystkich ciągów zaczynających się od kombinacji „płeć" spowodowałoby przesadzenie „sexton", „sextet" i „sextant". Proste dopasowanie ciągów również ignoruje kontekst, więc blokowanie „płci" będzie pasowało w przypadkach, gdy strona ankiety poprosiła respondenta o określenie płci za pomocą słowa „płeć" lub na stronach opisujących odziedziczone cechy płciowe lub role płciowe lub procesy sądowe o dyskryminację seksualną. Inne trudności w dopasowywaniu strun dotyczą kaprysów języka. Adresy URL mogą być wyświetlane w dowolnym języku, którego zestaw znaków rozpoznaje komputer, więc filtr podbije żądania w języku, dla którego nie ma list słów. Mówiąc bardziej ogólnie, w dowolnym języku można zaciemnić zawartość witryny za pomocą pozornie niegroźnego adresu URL, aby uniknąć filtrowania. Klasycznym tego przykładem jest witryna pornograficzna www.whitehouse.com, założona prawdopodobnie w celu wyłapywania odwiedzających, którzy omyłkowo wpisali „com", próbując wejść na stronę internetową Białego Domu Stanów Zjednoczonych (www.whitehouse.gov). Niedawno w kampaniach marketingowych spamu zostały skonfigurowane strony internetowe połączone w wiadomości e-mail, z bezsensownymi ciągami cyfr i znaków w adresie URL (np. Http://2sfh.com/7hioh), co utrudnia filtrowanie witryn. W badaniu z 2006 roku firma Veritest porównała trzy wiodące w branży produkty filtrujące w sieci Web (WebSense, SmartFilter i SurfControl). Zwycięski produkt zablokował 7 witryn, przesadził 8 witryn i błędnie skategoryzował 10 witryn z wstępnie wybranej listy 600 adresów URL. Dwa konkurujące ze sobą produkty wypadły gorzej, podblokowując 23 i 14 witryn oraz nadmiernie blokując 9 i 12 z 600.10. Meta-badanie badań skuteczności filtrowania przeprowadzone w 2010 r. Wykazało, że od 2001 do 2008 r. Średnia dokładność produktów filtrujących wynosiła 78%, przy czym pewien rosnący sukces: w latach 2007 i 2008 ankiety wykazały, że wskaźnik sukcesu wzrósł do 83 procent11. Jeśli są to wyniki wiodącej branży w branży, to niewątpliwie technologia wciąż się rozwija. Biorąc pod uwagę trudności z dokładnym dopasowaniem na podstawie tekstu lub adresu związanego z żądaniem strony internetowej, naturalną alternatywą jest zbadanie samej zawartości strony. Oczywiście dopasowywanie treści musi mieć dostęp do niezaszyfrowanych danych podczas przesyłania, więc zaszyfrowane sesje WWW stanowią prawdziwy problem dla tej taktyki. Niektóre organizacje zezwalają (lub wymagają), aby sesje HTTPS były przerywane na własnym serwerze proxy organizacji, potencjalnie umożliwiając serwerowi proxy odszyfrowanie danych i przeprowadzenie analizy zawartości.

Tekst.

Możliwe jest, chociaż wymaga dużych zasobów, obserwowanie strumienia ruchu sieciowego i wyszukiwanie tekstu, który pasuje do listy niepożądanych treści. Ten rodzaj dopasowywania zazwyczaj przeprowadza niewielką analizę kontekstu i dlatego jest podatny na ten sam rodzaj fałszywych alarmów (nadmierne blokowanie) i fałszywie negatywne (niedblokowanie), co opisano w sekcji 31.4.2. Ponadto, ponieważ coraz większa ilość treści w sieci Web obejmuje obrazy i dźwięki, dopasowywanie tekstu staje się mniej skuteczne.

Grafika.

Obiecująca nowa technika, z zastosowaniami do wyszukiwania wizualnego, a także blokowania treści wizualnych, dzieli obraz graficzny na mniejsze obiekty według koloru lub wzoru. Technika następnie ocenia każdy obiekt w bazie danych obrazów referencyjnych pod kątem dopasowania do pożądanych kryteriów. W przypadku blokowania budzących zastrzeżenia treści erotycznych obiekty mogą zostać ocenione pod kątem odcienia skóry i albo od razu zablokowane, albo skierowane do administratorów w celu ręcznego sprawdzenia, jeśli dopasowanie nie jest rozstrzygające. Chociaż filtrowanie oparte na treści nie rozwinęło się jeszcze w produkt komercyjny, narzędzia istnieją, a technologia wydaje się mieć

zastosowanie nie tylko do nieruchomych obrazów, ale także do treści wideo, a nawet audio. W 2006 roku biuro Inspektora Generalnego NASA wykorzystało program do wyszukiwania obrazów o nazwie Web ContExt, aby złapać pracownika, który zajmował się handlem dziecięcą pornografią.

WYKONANIE.

Z wyjątkiem dopasowywania na podstawie treści, które jeszcze nie dotarło na rynek w żaden znaczący sposób, większość filtrów - czy to adresów, domen czy słów kluczowych - obejmuje dopasowywanie list tekstowych.

Ręczne listy „złych adresów URL".

Wiele zapór zapewnia administratorom możliwość blokowania poszczególnych adresów URL w konfiguracji zapory. Wprowadzane ręcznie reguły te nadają się do jednorazowego blokowania, gdy alert zabezpieczeń lub dochodzenie zidentyfikuje witryny zawierające wirusy lub inne złośliwe oprogramowanie. Podejście to jest również przydatne do demonstrowania ogólnych możliwości filtrowania zapory oraz do testowania innych technologii Webblocking. Na przykład w organizacji korzystającej z komercyjnego rozwiązania blokującego na serwerze proxy sieci Web prosta reguła blokowania adresów URL na granicznym firewallu organizacji zapewniłaby łatwe wyrywkowe testowanie skuteczności komercyjnego rozwiązania. Biorąc jednak pod uwagę ekstremalne rozmiary i ciągły rozwój sieci, ręczne podejście nie jest dobrze skalowane, aby chronić przed wszystkimi możliwymi źródłami niepożądanego materiału.

Listy zablokowane innych firm.

Przy ogromnym rozmiarze sieci Web bardziej typowym podejściem jest użycie listy blokowanych stron trzecich. Większość z nich to produkty komercyjne z zastrzeżonymi bazami danych, opracowane przez połączenie zautomatyzowanych „robotów internetowych" i personelu technicznego oceniającego witryny internetowe. Niektóre firmy próbowały uniemożliwić naukowcom próbę poznania list blokujących i strategii, ale Urząd Praw Autorskich Stanów Zjednoczonych przyznał w 2003 r. Zwolnienie z ustawy Digital Millennium Copyright Act (DMCA) w celu dozwolonego użytku przez naukowców badających te listy. Istnieją również dwie alternatywy filtrowania typu open source, z publicznie wyświetlanymi (i konfigurowalnymi) listami bloków, które działają na buforujących serwerach proxy: SquidGuard15 i DansGuardian.

EGZEKWOWANIE.

Filtrowanie ruchu internetowego zazwyczaj odbywa się w wąskim punkcie sieci, takim jak zapora lub serwer proxy sieci Web, albo na indywidualnym komputerze klienckim. Korzyści skali skłaniają organizacje do filtrowania na urządzeniach sieciowych, podczas gdy produkty zaprojektowane do kontroli rodzicielskiej korzystania z Internetu przez dzieci zwykle znajdują się na indywidualnych komputerach domowych.

Pełnomocnicy.

Serwer proxy to urządzenie, które akceptuje żądanie z komputera klienckiego, a następnie przekierowuje je do ostatecznego miejsca docelowego. Serwery proxy służą organizacji do różnych celów, w tym do redukcji ruchu przez drogie łącza sieci rozległej i połączeń internetowych, zwiększonej wydajności dzięki buforowaniu często odwiedzanych stron internetowych oraz ochronie użytkowników wewnętrznych poprzez ukrywanie ich rzeczywistych adresów IP przed docelowymi serwerami sieci Web . Serwery proxy stanowią również naturalne miejsce kontroli dla organizacji, umożliwiając uwierzytelnianie i śledzenie żądań sieci Web, które przechodzą przez to pojedyncze

urządzenie. Większość przeglądarek obsługuje ręczną konfigurację serwera proxy dla całego ruchu internetowego, a także automatyczne wykrywanie serwerów proxy działających w sieci organizacji. Organizacje korzystające z serwerów proxy sieci Web zazwyczaj zezwalają na wychodzący ruch sieciowy tylko z adresu IP serwera proxy, zmuszając cały ruch HTTP do korzystania z serwera proxy. Korzystanie z zaszyfrowanej sesji internetowej (HTTPS) jest możliwe za pośrednictwem serwera proxy, chociaż albo kosztem możliwości monitorowania treści (jeśli proxy tylko przepuszcza ruch), albo kosztem kompleksowej prywatności zaszyfrowane łącze (jeśli proxy odszyfrowuje i ponownie szyfruje sesję). Osoby fizyczne używają również serwerów proxy, aby zachować prywatność swoich działań w sieci, jak opisano w sekcji 31.7.4. W związku z tym, oprócz pełnienia funkcji naturalnego nośnika dla aplikacji filtrujących zawartość, serwery proxy stanowią również poważne zagrożenie dla tych samych aplikacji.

Zapory ogniowe.

Zadaniem zapory jest analizowanie informacji o ruchu przez nią przechodzącym i stosowanie reguł opartych na tych informacjach.

Utrzymanie akceptowalnego czasu reakcji i przepustowości wymaga, aby zapora działała szybko i skutecznie. W tym celu większość zapór sprawdza jedynie informacje warstwy sieci, takie jak adresy źródłowe i docelowe oraz porty. Niedawno dostawcy zapór ogniowych dodawali więcej funkcji w celu zwiększenia bezpieczeństwa i atrakcyjności produktu. Wiele firm nazywa obecnie swoje bardziej zaawansowane zapory „bramami usługowymi" lub „bramami bezpieczeństwa", ponieważ pojęcie Unified Threat Management (UTM) staje się coraz bardziej popularne. Te urządzenia UTM łączą wiele funkcji, które wcześniej wymagały pojedynczych urządzeń, takich jak oprogramowanie antywirusowe, wykrywanie włamań i filtrowanie zarówno wiadomości-śmieci, jak i treści internetowych. Zaawansowane badanie ruchu zwiększa zapotrzebowanie na sprzęt firewall. Aby zmniejszyć spadek wydajności spowodowany zwiększoną inspekcją pakietów, wiele zapór umożliwia administratorowi zdefiniowanie określonych reguł lub protokołów do zaawansowanego sprawdzania. Na przykład, ponieważ wirusy są najbardziej rozpowszechnione w połączeniach e-mail, sieci Web i peer-to-peer, administrator zapory może potrzebować skonfigurować tylko kontrolę antywirusową na regułach mających zastosowanie do tych protokołów. Podobnie, jeśli firewall musi monitorować tylko wychodzące żądania HTTP z jednego adresu IP, serwera proxy sieci Web, dodatkowe obciążenie funkcji monitorowania może być ograniczone do tego profilu ruchu. Decyzja między filtrowaniem ruchu internetowego na serwerze proxy (zezwolenie firewallowi po prostu na przepuszczanie ruchu z tego adresu) a filtrowaniem ruchu internetowego na firewallu (posiadanie firewalla do inspekcji adresu URL) zależy od ilości ruchu internetowego i budżetu (jedno urządzenie lub dwa). Decyzja wpływa również na siłę twierdzenia, że organizacja skutecznie filtruje treści budzące zastrzeżenia. Jeśli graniczna zapora sieciowa organizacji przeprowadza filtrowanie, to jest to zależne od tego, czy zapora sieciowa jest jedynym sposobem, w jaki ruch opuszcza sieć organizacji. W grę mogą wchodzić inne wektory ruchu, w tym sieci bezprzewodowe, tunelowanie protokołów i anonimowe serwery proxy. Jeśli organizacja polega na komputerach klienckich do korzystania z serwera proxy sieci Web według zasad, należy również wziąć pod uwagę stopień, w jakim użytkownicy mogą obchodzić tę zasadę.

Narzędzia rodzicielskie.

Chociaż filtrowanie sieci Web oparte na kliencie nie jest powszechne w dużych organizacjach ze względu na koszty i zarządzanie takimi usługami na dużą skalę, produkty umożliwiające rodzicom blokowanie treści dla ich dzieci w domu stały się wielkim biznesem. Wielu dużych dostawców usług internetowych, takich jak AOL i MSN, oferuje narzędzia do blokowania treści przez rodziców jako bezpłatną funkcję swoich usług. Inne firmy sprzedają samodzielne produkty, które można zainstalować

na komputerze domowym, z zabezpieczonym hasłem rodzicielskim dostępem administracyjnym do funkcji blokowania treści. Net Nanny, CYBERsitter i CyberPatrol to jedne z bardziej popularnych ofert. Te produkty zwykle znajdują się na pojedynczych komputerach, a nie na urządzeniu sieciowym, chociaż jeśli komputer domowy jest skonfigurowany jako koncentrator łączności sieciowej (na przykład w przypadku udostępniania połączenia internetowego firmy Microsoft), elementy sterujące mogą filtrować ruch w taki sam sposób, jak organizacyjny serwer proxy. Wiele z tych produktów filtruje również inny ruch, w tym pocztę e-mail, udostępnianie plików peer-to-peer i komunikatory internetowe, a także oferuje filtrowanie w językach obcych, blokowanie adresów docelowych i reguły dostępu według pory dnia.

PODATNOŚCI.

Żaden schemat zabezpieczeń, zarówno fizyczny, jak i logiczny, nie jest całkowicie wolny od luk, a filtrowanie sieci z pewnością nie jest wyjątkiem od tej reguły. Użytkownicy, którzy chcą uzyskać dostęp do zablokowanych treści, mają do dyspozycji różne taktyki, chociaż rozwiązania różnią się pod względem łatwości użycia. Spoofing IP, tunelowanie protokołów i niektóre formy szyfrowania nie są trywialnymi praktykami, a zatem są narzędziami technicznie biegłych użytkowników w stosunkowo niewielkiej liczbie. Jednak inne technologie, takie jak anonimowe serwery proxy, witryny tłumaczeniowe i usługi pamięci podręcznej, są łatwym sposobem na pokonanie filtrowania przez przeciętnego użytkownika. Dostawcy filtrów internetowych nieustannie dążą do zwiększenia skuteczności swoich produktów, a zwolennicy ochrony prywatności i wolności słowa wspierają ciągłe wysiłki mające na celu pokonanie tego, co nazywają oprogramowaniem cenzurującym.

Podszywanie się.

W organizacji, która przeprowadza filtrowanie sieci Web na serwerze proxy, graniczna zapora sieciowa organizacji musi zezwalać na wychodzące żądania HTTP z adresu IP tego serwera proxy. Użytkownik, który może skonfigurować ruch tak, aby wyglądał tak, jakby pochodził z adresu IP serwera proxy, może być w stanie przechodzić przez zaporę sieciową bez faktycznego przechodzenia przez serwer proxy. Ta taktyka, znana jako fałszowanie adresów, wykorzystuje luźne zasady routingu na routerach, które przekazują cały nieznany ruch do bram domyślnych bez sprawdzania, czy ruch ten pochodzi z kierunku zgodnego z podanym adresem źródłowym. Wadą podszywania się, z punktu widzenia atakującego, jest to, że duża organizacja z dużym ruchem sieciowym przechodzącym przez sieć zauważy tymczasową niedostępność serwera proxy spowodowaną przez podszywanie się. Organizacja może przeciwdziałać fałszowaniu, konfigurując routery wewnętrzne tak, aby sprawdzały swoje tablice routingu, aby sprawdzić, czy adres pakietu przychodzącego do interfejsu jest zgodny z sieciami dostępnymi przez ten interfejs. Router odrzuca pakiety, których adresy źródłowe są niezgodne z ich rzeczywistym źródłem. Ta taktyka, zwana filtrowaniem odwrotnej ścieżki, wymaga bardziej wydajnego (i droższego) routera, więc zazwyczaj nie jest dostępna dla użytkownika domowego próbującego skonfigurować ochronę sieciową dla kontroli rodzicielskiej.

Tunelowanie.

Bardziej problematyczną taktyką, ponieważ opiera się na zachowaniu aplikacji, a nie na osłabianiu zabezpieczeń warstwy sieci, jest tunelowanie protokołów lub aplikacji. Aplikacja może hermetyzować informacje dowolnej innej aplikacji jako pakiet danych ogólnych i wysyłać je przez sieć. Klienci wirtualnej sieci prywatnej (VPN) używają tego podejścia do przesyłania ruchu przez zaszyfrowany tunel. Tunelowanie protokołów (czasami nazywane dynamicznym tunelowaniem aplikacji) polega na aplikacjach, które wysyłają dane na powszechnie dozwolonych portach. Na przykład, użytkownik może tunelować sesję sieci Web przez aplikację Secure Shell (SSH), która używa portu TCP 22. SSH często jest dozwolone przez zapory ogniowe, a ponieważ używa zarówno uwierzytelniania, jak i szyfrowania,

może być trudno monitorować różnicę między legalna sesja SSH i ukryty tunel. W tym przykładzie przeglądarka internetowa wysyła żądanie strony HTTP na porcie TCP 80, a inna aplikacja działająca w systemie klienta przechwytuje i przekierowuje żądanie portu 80 do tunelu zaszyfrowanego przez SSH. Drugi koniec tunelu SSH może być serwerem docelowym lub serwerem proxy gdzieś między klientem a serwerem. Aplikacja kliencka (w tym przypadku przeglądarka internetowa) nie jest świadoma przekierowania ruchu i nie wymaga zmiany konfiguracji. Jest to podobne do podejścia stosowanego przez tradycyjne VPN. Ostateczna forma tunelowania protokołów znalazła ograniczone zastosowanie, ponieważ protokół IPv6 stał się szerzej dostępny. Wciąż w powijakach jako protokół obejmujący cały Internet, niemniej jednak jest dostępny w niektórych częściach świata, a brokerzy tuneli oferują połączenia tunelujące ruch IPv6 w sieci szkieletowej IPv4. Może to stać się problematyczne, jeśli strona główna lub witryna sądzi, że filtruje żądania wychodzące, ale szuka tylko zawartości w pakietach IPv4. Tunelowanie aplikacji (nazywane również tunelowaniem aplikacji statycznej) wymaga ponownej konfiguracji aplikacji klienckiej w celu przekierowania żądań przez inny port na komputerze klienckim. Zwykle użytkownik przedefiniowuje adres docelowy aplikacji na adres hosta lokalnego (w zakresie 127.0.0.x, odnosząc się do urządzenia lokalnego), a następnie aplikacja ustanawia połączenie z tego portu lokalnego do miejsca docelowego na dozwolonym port sieciowy. Takie podejście wymaga zmiany konfiguracji aplikacji klienckiej. Niektóre produkty VPN Secure Sockets Layer (SSL) wykorzystują to podejście do tunelowania jednego lub więcej protokołów lub całego ruchu przez zaszyfrowany tunel HTTP. Tunelowanie za pośrednictwem protokołu lub aplikacji zazwyczaj wymaga dostępu w celu skonfigurowania istniejących ustawień aplikacji lub zainstalowania dodatkowego oprogramowania. Dlatego tunelowanie jest niedostępne dla użytkowników w organizacjach, które dają użytkownikom końcowym ograniczoną kontrolę nad swoimi komputerami. Tunelowanie jest większym problemem dla rodziców, których zręczne technicznie dzieci mogą instalować aplikacje tunelujące, aby obejść kontrolę rodzicielską.

Szyfrowanie.

Specjaliści ds. Bezpieczeństwa ogólnie zachęcają do stosowania szyfrowania, ponieważ chroni ono poufne informacje podczas przesyłania przez sieć. Jednak gdy użytkownicy szyfrują dane w celu ukrycia transakcji, które naruszają zasady organizacji, z punktu widzenia organizacji szyfrowanie może stać się zobowiązaniem, a nie zasobem. W sesji encryptedWeb korzystającej z protokołu HTTPS (czyli HTTP przez SSL) zawartość sesji jest szyfrowana, a tym samym niedostępna do monitorowania lub filtrowania na podstawie adresu URL lub zawartości danych. Jednak źródłowe i docelowe adresy IP sesji, które są widoczne dla protokołu TCP podczas konfigurowania sesji, pozostają widoczne w sieci. Jest to konieczne, aby routery mogły pobierać zaszyfrowane pakiety danych z jednego końca transakcji na drugi. W związku z tym, chociaż sesja HTTPS jest odporna na filtrowanie według dopasowywania tekstu adresu URL lub filtrowania opartego na treści, blokowanie serwera docelowego jest nadal skuteczne.

Jak wspomniano w poprzedniej sekcji, technologie VPN stanowią kolejne zastosowanie szyfrowania do ochrony treści transakcji. Ponownie, chociaż VPN szyfruje zawartość transakcji, adresy IP punktów końcowych muszą pozostać widoczne, aby móc przetransportować dane do miejsca przeznaczenia. Inna bardziej skomplikowana wersja szyfrowania, zwana steganografią, w rzeczywistości osadza dane w innych informacjach, aby uniknąć wykrycia. Na przykład użytkownik może osadzić wiadomość tekstową w informacjach używanych do kodowania obrazu.

Anonimowość.

Większość monitorowania sieci i filtrowania treści zależy od tożsamości użytkownika. Filtrowanie treści może się odbywać bez informacji o tożsamości, ale jeśli organizacja lub kraj nie zdecyduje się narzucić drakońskiego, szerokiego filtrowania żądań allWeb (a niektórzy wybiorą to podejście), organizacja

może chcieć wymusić wymagania dotyczące filtrowania tylko dla niektórych użytkowników. Na przykład okręg szkolny może chcieć nałożyć ściślejsze filtrowanie na uczniów szkół podstawowych niż na uczniów szkół średnich i stosować bardziej liberalną politykę wobec nauczycieli i pracowników administracyjnych. Zmorą tego podejścia jest anonimowość. Gdy obawy dotyczące prywatności poszczególnych osób prowadzą do stosowania technologii anonimizujących na skalę, która utrudnia ustalenie tożsamości użytkowników, jedynym sposobem przestrzegania zasad i przepisów wymagających filtrowania treści dla niektórych grup jest filtrowanie pod kątem wszystkich grup na poziomie najbardziej rygorystycznych wymagań. W przypadku okręgu szkolnego, jeśli sieć nie jest w stanie odróżnić ruchu uczniów od ruchu nauczycieli, wówczas okręg musi również nałożyć na nauczycieli wymagania dotyczące filtrowania treści uczniów. W rzeczywistości wiele okręgów szkolnych dokonało tego wyboru ze względu zarówno na ogromną liczbę trudnych do znalezienia anonimowych serwerów proxy, jak i na techniczne trudności związane z oddzieleniem korzystania przez uczniów i nauczycieli ze wspólnej infrastruktury sieci szkolnej. Mimo to zewnętrzne anonimowe serwery proxy utrudniają administratorom sieci wymuszenie zgodności z przepisami dotyczącymi filtrowania. Inne zastosowania technologii zwiększających prywatność (PET) obejmują oparte na sieci schematy anonimowości, takie jak sieci mieszane i routing cebulowy. Projekt, znany jako Tor (pierwotnie skrót od routera cebulowego), zyskuje na popularności w ostatnich latach.

Witryny tłumaczeniowe.

Witryny z tłumaczeniami językowymi, takie jak BabelFish (dawniej http://babelfish.yahoo.com), również oferują możliwość uniknięcia filtrów treści. Użytkownik wprowadza adres URL i klika przycisk, aby zażądać tłumaczenia tekstu witryny. Sesja użytkownika odbywa się między komputerem klienckim a Yahoo !, a żądany adres URL jest przekazywany jako dane naciśnięcia klawisza w sesji HTTP, więc potencjalnie zablokowana witryna jest udostępniana, o ile użytkownik może uzyskać dostęp do Yahoo! Jest to szczególny przypadek serwera proxy, ponieważ żądanie wpisane w BabelFish generuje żądanie do serwera na drugim końcu, ale klient nie otrzymuje dokładnych wyników tej odpowiedzi; zamiast tego wyświetlacz zawiera całą grafikę oryginału, ale tekst jest przetłumaczony na żądany język. Tak więc z punktu widzenia narzędzia monitorującego klient jest zawsze połączony z Yahoo !, a język obcy może utrudniać nawet filtr treści przeglądający tekst w pakietach HTTP. Uznając możliwość nadużyć, Yahoo! opublikował dokument z Warunkami użytkowania witryny BabelFish zakazujący korzystania z usługi w zakresie „przedmiotów objętych embargiem Stanów Zjednoczonych, materiałów szerzących nienawiść (np. pamiątki nazistowskie),… pornografii, prostytucji,… [lub] przedmiotów związanych z hazardem” wśród wielu innych klas działania lub produkty. Witryny z tłumaczeniami są również irytujące dla administratorów systemu, ponieważ tłumaczenie witryny w określonym języku na ten język zwykle skutkuje niezmienioną zawartością. Na przykład tłumaczenie strony internetowej w języku angielskim z francuskiego na angielski po prostu przekazuje treść bez zmian.

Usługi buforowania.

Jednym z głównych zastosowań serwerów proxy było zmniejszenie ruchu w sieci poprzez zapisywanie lokalnych kopii często żądanych stron. To zachowanie może obejść filtrowanie w sieci Web, o ile użytkownik może uzyskać dostęp do serwera pamięci podręcznej. Na przykład Google zapisuje w pamięci podręcznej wiele stron, starając się zapewnić bardzo szybką odpowiedź na zapytania wyszukiwania. Duże pliki graficzne i wideo zajmują najwięcej przepustowości, dlatego wyszukiwarka grafiki Google często je zapisuje w pamięci podręcznej. Często pliki graficzne są dostępne nawet dla witryn, które już nie istnieją. Użytkownicy, którzy chcą ominąć blokady w treściach o charakterze jednoznacznie seksualnym, często korzystają z wyszukiwarki grafiki Google. Google udostępnia funkcję Safe Search z trzema poziomami dobrowolnego filtrowania. Domyślne (środkowe) ustawienie filtruje „tylko wulgarne obrazy”, podczas gdy ścisłe ustawienie filtruje „zarówno wulgarne teksty, jak i

wulgarne obrazy". Google zauważa, że „żaden filtr nie jest w 100% dokładny, ale Safe Search powinno wyeliminować większość nieodpowiednich materiałów". Funkcję wyszukiwania można konfigurować dla każdego użytkownika i nie zapewnia ona ochrony hasłem, więc nie jest to technologia filtrowania sieci Web, a raczej dobrowolny filtr wyszukiwania oparty na płci.

PRZYSZŁOŚĆ.

W miarę pojawiania się coraz większej liczby sposobów komunikowania i rozpowszechniania treści, branża filtrowania treści będzie niewątpliwie ewoluować, aby objąć nowe technologie. Obecnie dostawcy sprzedają produkty filtrujące dla poczty e-mail, czatów internetowych, grup dyskusyjnych, komunikatorów, udostępniania plików w sieci peer-to-peer i FTP, a także filtrujących żądania internetowe. Nowe funkcje pojawią się również w „tradycyjnym" filtrowaniu sieciowym, w tym w filtrowaniu tunelowanego ruchu IPv6. Najnowsze wersje produktów do filtrowania domu, Net Nanny i McAfee Parental Controls, oferują teraz możliwość wymuszenia opcji bezpiecznego wyszukiwania w głównych wyszukiwarkach (takich jak Google Safe Search, opisane w sekcji 31.7.6) i zapewniają „rozpoznawanie obiektów", które rozpoznaje określone wersje obiektów internetowych (np. liczniki odwiedzin), które są powszechnie używane w witrynach pornograficznych. Zwolennicy filtrowania treści i ci, którzy muszą go używać i potrzebują niezawodnego produktu, będą zachęceni rozwojem branży, ale być może rozczarowani, że problem nigdy nie został całkowicie rozwiązany. Postępy w rozpoznawaniu obrazów mogą zapewnić znacznie lepsze filtrowanie, ale mogą również stworzyć nowe sposoby zmiany treści w celu obejścia tych narzędzi. Ci, którzy potępiają te produkty jako oprogramowanie cenzorskie, wskażą, że historycznie większość prób cenzurowania mowy ostatecznie się nie powiodła. Ostatecznie obie strony prawdopodobnie wypracują niełatwy kompromis, tak jak to miało miejsce w przypadku sprzedaży „dorosłych" treści drukowanych i wideo. Tak długo, jak niektórzy ludzie będą upierać się przy swoim prawie do rozpowszechniania informacji, które inni uważają za obraźliwe, ten konflikt prawdopodobnie będzie trwał. Debata na temat równowagi między wolnością a bezpieczeństwem będzie prawdopodobnie w coraz większym stopniu zabarwiona obawami związanymi z nielegalną działalnością, terroryzmem i cyberwojną, co może przełożyć się na postęp w dostępnych narzędziach, a także na próby zdefiniowania i uregulowania odpowiedniego wykorzystania tych narzędzi.

PODSUMOWANIE.

Z różnych powodów, jednych lepiej niż inne, grupy ludzi posiadających władzę nad innymi grupami lub za nie odpowiedzialnych chcą kontrolować rodzaj informacji, do których mają dostęp inne grupy, oraz uzyskać wiedzę o tym, kto handluje informacjami. Produkty do monitorowania i filtrowania treści (lub oprogramowanie cenzorujące) zapewniają tego rodzaju kontrolę przy użyciu technologii komputerowej do badania informacji przepływających w Internecie. Zwolennicy wolności słowa i prywatności twierdzą, że filtrowanie treści uniemożliwia legalny, legalny dostęp do informacji. Nawet jeśli w niektórych przypadkach można uznać, że filtrowanie jest uzasadnione, obecne technologie są podatne na błędy, zarówno uniemożliwiające blokowanie niektórych kontrowersyjnych treści, jak i blokowanie niektórych witryn, które nie zawierają takich treści. Większość technik filtrowania obejmuje badanie metadanych, dopasowywanie ciągu tekstu lub liczb w celu określenia źródła lub przeznaczenia żądania lub wiadomości, lub niektórych cech kontekstu komunikacyjnego. Te metadane mogą służyć do blokowania dostępu: serwery mogą być blokowane według adresu lub nazwy serwera. Niedawne dodanie domeny .xxx otworzyło możliwość filtrowania całych domen najwyższego poziomu. Inne metody skoncentrują się na blokowaniu treści, badaniu tekstu lub pozakontekstowych części strony internetowej, w tym grafiki. Badania nad rozpoznawaniem obrazu i dopasowywaniem odcieni skóry postępują, a agencje rządowe wykorzystywały niektóre narzędzia do rozpoznawania obrazu w postępowaniach sądowych, ale rozpoznawanie obrazu nie weszło jeszcze w dużej mierze na rynek

komercyjny. Z każdą ochronną lub nadopiekuńczą strategią wiąże się grupa ludzi oddanych jej klęsce. Filtrowanie treści ma wiele luk w zabezpieczeniach, wśród których główną jest wykorzystanie anonimowości za pośrednictwem technologii zwiększających prywatność, takich jak anonimowe serwery proxy i routing cebulowy. Inne sposoby na pokonanie filtrowania sieci Web obejmują wykorzystanie tunelowania protokołów i aplikacji, szyfrowania, witryn do tłumaczenia sieci Web i usług buforowania. Technologie filtrowania ulegały poprawie przez lata, podobnie jak pomysłowość tych, którzy im poświęcili się udaremnianie. W miarę mnożenia się kanałów informacyjnych i pojawiania się nowych środków komunikacji, ten rodzaj konfliktu między technologiami ochronnymi a obchodzeniem ochrony prywatności prawdopodobnie będzie się utrzymywał.