УДК 004.056.52:621.391

Каптур В.А
Каптур В.А.
Kaptur V.A.

КОМПЛЕКСНІ СИСТЕМИ ФІЛЬТРАЦІЇ КОНТЕНТУ В МЕРЕЖІ ІНТЕРНЕТ

КОМПЛЕКСНЫЕ СИСТЕМЫ ФИЛЬТРАЦИИ КОНТЕНТА В СЕТИ ИНТЕРНЕТ

INTEGRATED FILTERING CONTENT SYSTEMS IN THE INTERNET

Анотація. Наведено визначення комплексної системи фільтрації контенту в мережі Інтернет та запропоновано класифікацію засобів фільтрації на основі раніше розробленої моделі. Проведено детальний аналіз недоліків використання систем фільтрації, що побудовані лише на одному засобі, та показано яким чином ці недоліки можуть бути подолані за рахунок комбінування з іншими засобами фільтрації. Запропоновано концепцію гібридної системи фільтрації контенту, що базується на одночасному використанні DNS та proxy серверів.

Аннотация. Приведено определение комплексной системы фильтрации контента в сети Интернет и предложено классификацию средств фильтрации на основе ранее предложенной модели. Проведён детальный анализ недостатков использования систем фильтрации, построенных лишь на одном средстве и показано, каким образом эти недостатки могут быть преодолены за счёт комбинирования с другими средствами фильтрации. Предложена концепция гибридной системы фильтрации контента, которая базируется на одновременном использовании DNS и proxy серверов.

Summary. Defined of integrated content filtering system on the Internet and proposed classification of filtering based on a previously proposed model. A detailed analysis of the disadvantages of using filtering systems built on only one mean and illustrates how these shortcomings can be overcome by combining with other means of filtration. A concept of a hybrid system of content filtering, which is based on the simultaneous use DNS and proxy servers.

Зважаючи на бурхливий розвиток мережі Інтернет, який спостерігається в останні роки, все більшої актуальності набуває проблема фільтрації контенту, під час його передавання від інформаційного ресурсу до споживача. Необхідність вирішення таких завдань, як захист дітей в мережі Інтернет, блокування небажаної реклами та шкідливого програмного забезпечення, дотримання корпоративної політики щодо заборони доступу до розважальних ресурсів в робочий час та навіть цензура (для країн із законодавчо обмеженою свободою слова) створили передумови для розвитку цілої індустрії систем фільтрації контенту (СФК). Результатом цього розвитку стала поява великої кількості програмних і програмно-апаратних рішень, призначених для блокування доступу до інформаційних ресурсів мережі Інтернет на різних рівнях (безпосередньо на робочій станції користувача, на стороні провайдера, в точках обміну навантаженням операторів телекомунікацій тощо).

Під час попередніх досліджень [1] було розроблено узагальнену модель фільтрації контенту в мережі Інтернет, яка дозволила провести класифікацію існуючих засобів, видів, методів та підходів до фільтрації на основі аналізу процесу передавання http-навантаження. Навіть поверхневий аналіз різних засобів фільтрації (веб-клієнт, proxy-сервер, DNS-сервер, міжмережний екран) показав наявність переваг і недоліків у кожного з них [1, 2] та дозволив зробити висновок про доцільність комбінування різних варіантів і підходів з метою створення комплексних СФК. Проте створення таких систем сьогодні стримується недостатнім рівнем дослідження спільної роботи різних засобів фільтрації, як з теоретичної (оцінка надійності, швидкодії тощо), так і з практичної (оцінка можливості спільної роботи конкретних реалізацій) точок зору.

Метою статті є визначення найбільш стабільної з точки зору взаємного усунення недоліків конфігурації комплексної системи фільтрації контенту та визначення оптимальних з точки зору збільшення швидкодії параметрів її роботи.

Під комплексною системою фільтрації контенту (КСФК) будемо розуміти систему, що складається з двох або більше засобів фільтрації, кожен з яких може використовувати різні види, методи та підходи до фільтрації контенту. В межах КСФК засоби можуть працювати у послідовному (підсилюючи один одного) або в паралельному (доповнюючи один одного) режимі. В послідовному режимі адреса або контент, що пройшли процедуру фільтрації на одному із засобів, потрапляють на вхід іншого засобу з метою повторного оброблення (наприклад, за допомогою інших методів). У

свою чергу, паралельний режим припускає фільтрацію адреси через використання одного засобу фільтрації, а фільтрацію контенту із використанням іншого.

З метою спрощення розуміння подальшого матеріалу запровадимо умовні позначення для систем фільтрації контенту із урахуванням класифікаційної моделі запропонованої в [1]. Загальна структура умовного позначення наведена на рис. 1.

Рисунок 1 – Умовні позначення систем фільтрації контенту

Відповідні елементи позначення (рис. 1) можуть приймати такі значення (відповідно до порядку слідування):

1. Засоби фільтрації: міжмережний екран — «F» (firewall), DNS-сервер — «D», proxy-сервер — «P» (proxy), веб-клієнт (або інше прикладне програмне забезпечення) — «B» (browser).

2. Архітектура: централізована — «C» (centralized) [3] та децентралізована — «D» (decentralized).

3. Види фільтрації: за контентом — «C» (content), за адресою — «A» (address).

4. Підходи до фільтрації фільтрації: “чорні” списки — «B» (black), “білі” списки «W» (white).

5. Методи формування списків: вручну — «M» (manual), експертами — «E» (experts), автоматично (у випадку якщо відбувається фільтрація за контентом і аналіз виконується самою системою) — «A» (auto).

Таким чином, система, що базується на використанні proxy-серверів, як засобів фільтрації, передбачає децентралізовану архітектуру, фільтрацію виключно за адресою на основі «чорних» списків, при ручному методі їх формування може бути позначена, як F|D|A|B|M.

У разі, коли СФК має дві або більше властивостей одночасно (наприклад, підтримує фільтрацію, як за «білими», так і за «чорними» списками) відповідний розділ позначається обома літерами одночасно (наприклад, «BW»). В свою чергу комплексні СФК (на відміну від звичайних) можуть мати дві або більше літери в секції «Засоби» (наприклад, «DB» у разі одночасного комбінування фільтрації на DNS-сервері та на веб-клієнті) за умов, якщо інші властивості у системі фільтрації однакові або можуть позначатися, як серія записів про звичайні СФК.

Як було зазначено раніше, кожен з розглянутих засобів фільтрації має низку недоліків, які можуть бути повністю або частково вирішені за рахунок комбінування з іншими засобами. В табл. 1 зведено найбільш очевидні недоліки кожного з чотирьох засобів (які визначено шляхом дослідження популярних технічних рішень, що реалізовані на одному засобі фільтрації [4…8]) та проведено оцінку можливості їх усунення шляхом спільного (послідовного) використання з іншими засобами. Знаком «+» позначено випадки, коли комбінування дозволяє подолати недолік, знаком «–» позначено випадки, коли подолати недолік завдяки лише комбінуванню неможливо, а одночасно знаками «+» та «–» позначаються випадки часткового подолання недоліку або випадки, коли отримана за рахунок комбінування КСФК буде не ефективною з точки зору значного споживання ресурсів та/або легкості її обходу.

Одним із найсуттєвіших недоліків фільтрації на міжмережному екрані (табл. 1) є стрімке зростання навантаження на обладнання при збільшенні кількості записів у списках фільтрації. Це пояснюється тим, що аналіз при цьому здійснюється для кожного пакета, а не для запитів на створення сесії. Очевидно, що використовуючи міжмережний екран лише для спрямування навантаження на DNS або proxy-сервер зазначений недолік може бути подоланий, однак при цьому роль самого міжмережного екрана, як засобу фільтрації контенту, значно зменшується. У свою чергу, використовуючи як основний засіб фільтрації веб-клієнт, але при цьому обмежуючи використання інших засобів безпосередньо на міжмережному екрані, можна перекласти завдання фільтрації безпосередньо на вбудовані до веб-клієнта механізми та досить надійно обмежити використання інших видів навантаження.

Таблиця 1 – Комбінування засобів фільтрації для усунення недоліків

Засоби фільт-рації	Основні недоліки фільтрації за допомогою засобу	Можливість усунення недоліку за рахунок комбінування з іншим засобом фільтрації			
		F	D	P	B
F	Зростання навантаження на обладнання при збільшенні кількості записів у списках фільтрації	X	+	+	+
	Обмеженість діалогу з користувачем		+	+	+
	Висока ймовірність помилкового обмеження та неефективність фільтрації за контентом		–	+	+
D	Можливість обходу системи обмеження доступу шляхом конфігурування іншого DNS-сервера або внесенням запису до hosts-файла	+ –	X	+	–
	Неможливість фільтрації за контентом	–		+	+
	Неможливість фільтрації запитів, що не містять доменних імен	+		+	+
P	Неможливість або складність обмеження навантаження, що передається за допомогою відмінних від HTTP протоколів	+ –	+ –	X	–
	Великий час оброблення запитів при збільшенні кількості записів у списках фільтрації	–	+		–
B	Складність адміністрування та підтримки при зростанні кількості вузлів	–	+	+	X
	Можливість обмеження лише для певного веб-клієнта	+	+ –	+	
	Складність збирання та оброблення статистичної інформації	–	+	+	

Іншим важливим недоліком міжмережного екрана, як засобу фільтрації, є обмеженість діалогу з користувачем (табл. 1) при блокуванні того або іншого ресурсу. Це пояснюється тим, що прикладний процес (у випадку, коли міжмережний екран працює на іншому вузлі) не може відрізнити блокування пакета на міжмережному екрані від його втрати під час передавання крізь мережу. Вирішити цю проблему можна лише комбінуючи міжмережний екран із іншими засобами фільтрації з метою делегування їм функцій фільтрації (як і в попередньому випадку).

Останній з наведених в табл. 1 недоліків міжмережного екрана – висока ймовірність помилкового обмеження та неефективність фільтрації за контентом – пояснюється тим, що під час фільтрації за контентом міжмережний екран має не тільки обробляти кожен пакет окремо, а й не має можливості проводити аналіз всього інформаційного блока. Так, наприклад, завдяки фрагментації блока даних на пакети перша частина ключової фрази, яка підпадає під шаблони фільтрації може знаходитись в одному пакеті, а друга частина в іншому. В такому випадку обидва пакети подолають обмеження. Вирішенням проблеми, як і в попередніх випадках, може стати доповнення міжмережного екрану фільтрацією на proxy-сервері або безпосередньо на веб-клієнті. Винятком є лише використання DNS-сервера, який не може бути використаний для фільтрації контенту взагалі.

Таким чином, комбінації «FP» та «FB» дозволяють вирішити основні недоліки міжмережного екрана як засобу фільтрації, проте при цьому фактично відбудеться заміщення одного засобу іншим. При цьому комбінація «FD» дозволяє вирішити лише частину недоліків та не може розглядатися як ефективний засіб за необхідності забезпечення фільтрації за контентом.

Як зазначено в табл. 1 одним з суттєвіших недоліків фільтрації із використанням DNS-сервера є можливість обходу системи шляхом конфігурування іншого DNS-сервера або внесенням запису до hosts-файла. Повністю вирішити цю проблему можливо лише за рахунок направлення всіх запитів до proxy-сервера. В свою чергу використання міжмережного екрану або веб-клієнта дозволяє вирішити проблему частково (наприклад, блокування звернення до інших DNS-серверів) або взагалі не позбавляє СФК зазначеного недоліку.

Другим, з наведених (табл. 1), недоліком систем фільтрації на DNS-серверах є неможливість фільтрації за контентом. Враховуючи наведені вище пояснення щодо оброблення контенту міжмережним екраном вирішити цю проблему шляхом комбінування саме із цим засобом

неможливо. Проте надсилання запитів через proxy-сервер або з власного веб-клієнта дозволяє в повній мірі забезпечити фільтрацію контенту.

Ще одним очевидним недоліком систем фільтрації на DNS-серверах є неможливість фільтрації запитів, що не містять доменних імен. Це пояснюється тим, що такі запити просто не потрапляють до DNS-сервера. Для подолання недоліку може бути використаний будь-який з трьох інших засобів за рахунок фільтрації запитів, що містять лише IP-адресу.

Таким чином, в повному обсязі вирішити позначені проблеми фільтрації контенту на DNS-серверах можна лише за рахунок комбінування із proxy-сервером, який використовується для фільтрації контенту та фільтрації запитів, що містять IP-адреси.

Одним із основних недоліків засобів фільтрації на основі proxy-серверів є неможливість або складність обмеження навантаження, що передається за допомогою відмінних від HTTP протоколів. Очевидно, що частково цю проблему можна вирішити за допомогою міжмережних екранів (наприклад, за рахунок блокування інших видів навантаження). Однак, враховуючи значний обсяг споживання обчислювальних ресурсів, більш прийнятним буде організація додаткової фільтрації із використанням DNS-серверів. Такий підхід дозволить блокувати доступ до будь-яких вузлів (незалежно від протоколу, що використовується для інформаційного обміну), що адресуються за допомогою доменних імен.

Іншим недоліком proxy-серверів (як засобів фільтрації) (табл. 1) є відносно великий час оброблення запитів при збільшенні кількості записів у списках фільтрації. Це пояснюється тим, що порівняно, наприклад, із DNS-сервером на proxy-сервери покладено цілу низку різноманітних функцій (кешування, передавання контенту тощо), які вимагають додаткових обчислювальних ресурсів. Вирішенням цієї проблеми може бути розділення функцій фільтрації між proxy та DNS серверами.

Як показано в табл. 1 найбільш вагомим недоліком фільтрації на веб-клієнті є складність адміністрування та підтримки при зростанні кількості вузлів. Очевидно, що підтримувати та стежити за цілісністю великої кількості копій програмного забезпечення веб-клієнта значно складніше ніж за системами фільтрації, що розміщені на одному сервері. Як і в попередніх випадках вирішенням проблеми може бути делегування функцій фільтрації іншому засобу (proxy або DNS сервера).

Два інших недоліки фільтрації контенту за допомогою веб-клієнта, а саме: можливість обмеження доступу лише для певного веб-клієнта та складність збирання та оброблення статистичної інформації мають таку саме природу, що й у попередньому випадку. При цьому, якщо перший з цих недоліків може бути подоланий за допомогою будь-якого з альтернативних засобів фільтрації, то другий недолік, як правило, долають лише за допомогою DNS або proxy-серверів.

Таким чином, аналіз інформації з табл. 1, показав, що найбільш стійким до зазначених вище недоліків є комбінування засобів фільтрації до КСФК є пара DNS та proxy-сервер. Розглянемо цю конфігурацію більш детально. На рис. 2 зображено модель роботи гібридної (одночасно децентралізованої та централізованої) системи фільтрації контенту при одночасному використанні DNS та proxy-серверів, а також при використанні виключно proxy-сервера як засобу фільтрації.

Як видно з рис. 2 користувацькі запити на отримання доступу до інформаційних ресурсів мережі Інтернет з відповідних мереж надходять до proxy-серверів, які можуть бути встановлені, як на стороні організації, так і на стороні провайдера або в будь-якій іншій точці мережі Інтернет. У свою чергу proxy-сервери здійснюють перевірку наявності запитаної адреси серед заборонених. При цьому, якщо в першому випадку (верхня частина рис. 2) proxy-сервер здійснює перевірку по всьому списку відразу, то для випадків КСФК (нижня частина рис. 2) перевірка на proxy-сервері здійснюється тільки для частини записів (наприклад, тільки для записів, що являють собою IP-адресу).

У разі, якщо запитана адреса знайдена в переліку заборонених, користувачу повертається зміст інформаційного повідомлення про блокування (замість запитаного контенту). В іншому випадку одноступінчатий засіб фільтрації (верхня частина рис. 2) надсилає запит безпосередньо до інформаційного ресурсу (за необхідності здійснивши перетворення доменного імені в IP-адресу за допомогою звичайного DNS-сервіса). У свою чергу, КСФК (нижня частина рис. 2) передбачає пересилання запитів, що містять доменні імена, для фільтрації на централізованому DNS-сервері. Для випадку використання КСФК централізований DNS-сервер також здійснює перевірки наявності запитаної адреси серед заборонених використовуючи при цьому лише ту частину записів, що містять доменне ім'я. Як і у попередньому випадку, якщо запитана адреса знайдена в переліку заборонених, proxy-сервера повертається адреса службового веб-севера, який видає інформаційне повідомлення про блокування ресурсу на будь-який запит. В іншому випадку DNS-сервер повертає коректну IP-

адресу запитаного ресурсу, а після отримання відповіді від DNS-сервера, proxy-сервер, що входить до складу КСФК, надсилає запит безпосередньо до інформаційного ресурсу.

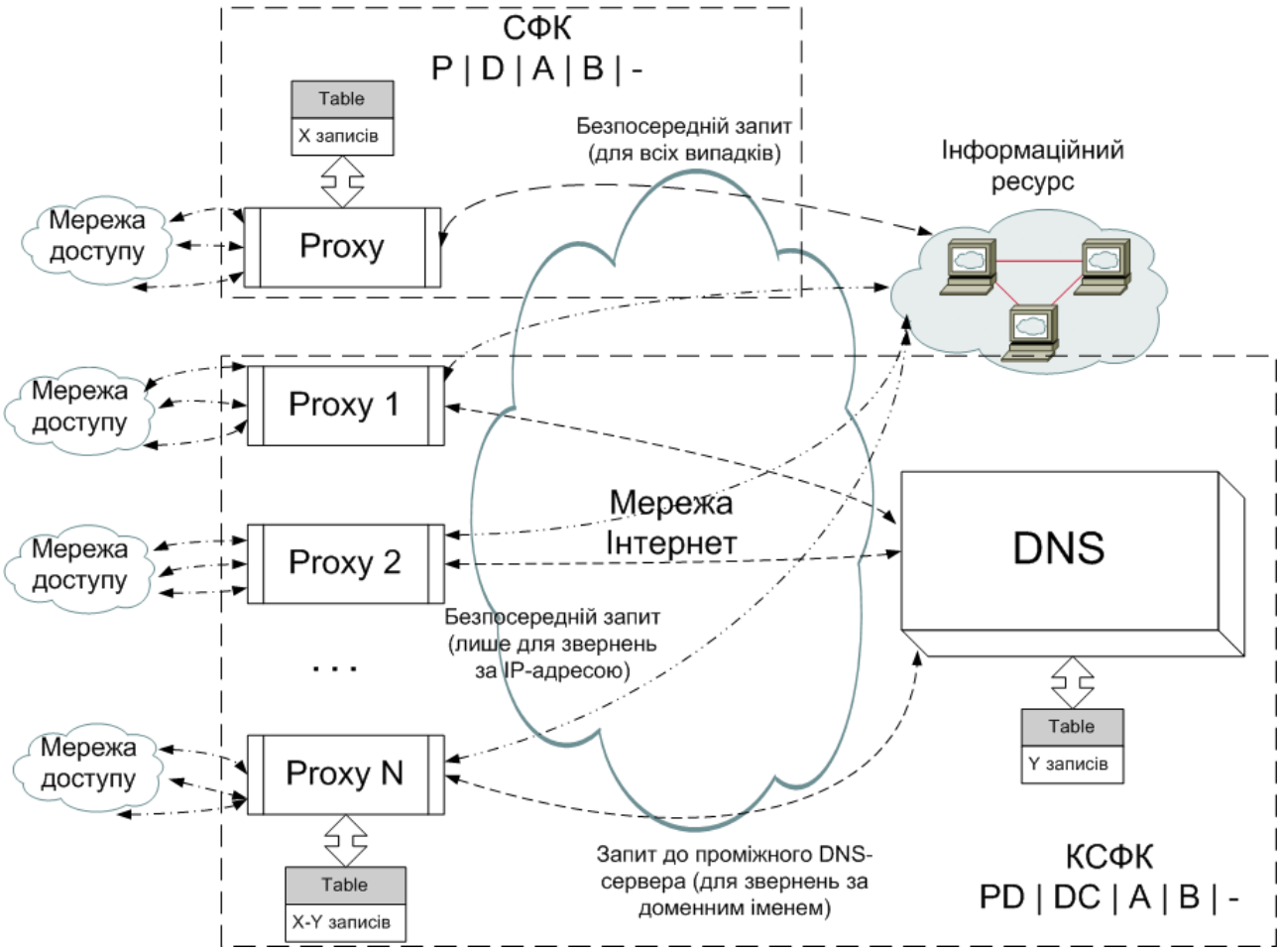

Рисунок 2 – Модель роботи гібридної системи фільтрації контенту
при одночасному використанні DNS та proxy-серверів

Очевидно, що використання КСФК при цьому не тільки дозволяє забезпечити фільтрацію альтернативних до HTTP видів навантаження (за рахунок додаткової фільтрації на DNS), але й розподілити навантаження між серверами, що розміщені безпосередньо в мережі організації, та центральним DNS-сервером. Окремою перевагою зображеного на рис. 2 підходу є збереження в таємниці тієї частини списку блокування, що міститься на DNS-сервері.

Слід зазначити, що представлена модель потребує дослідження з точки зору визначення достатніх значень параметрів обчислювальної потужності серверів для забезпечення роботи КСФК із заданим рівнем якості обслуговування. При цьому основним параметром, за яким може оцінюватись рівень якості обслуговування є середній час опрацювання комплексною системою фільтрації контенту одного запиту. Таким чином, подальші дослідження мають бути спрямовані на розробку математичної моделі та дослідного зразка зазначеної вище КСФК та її тестування як в реальних умовах, так і під штучно збільшеним навантаженням.

За результатами проведеного дослідження можна зробити такі висновки:

1. Створення систем фільтрації, що використовують одночасно декілька різних засобів стримується недостатнім рівнем дослідження їх спільної роботи як з теоретичної, так і з практичної точок зору.

2. Запропонована у статті концепція комплексних систем фільтрації контенту може бути покладена в основу створення ефективних механізмів блокування доступу до небажаних інформаційних ресурсів.

3. Запропонована у статті система умовних позначень дозволяє в зручній формі охарактеризувати систему фільтрації контенту будь-якої складності або навіть комбінацію декількох систем.

4. Проведений аналіз основних недоліків фільтрації за допомогою кожного із засобів та визначення способів усунення цих недоліків за рахунок комбінування з іншим засобом фільтрації дозволив визначити, що найбільш стійкою є пара DNS та proxy-сервер.

Література

1. *Каптур В.А.* Узагальнена класифікаційна модель фільтрації контенту в мережі Інтернет / В.А. Каптур // Збірник наукових праць Військового інституту телекомунікацій та інформатизації НТУУ "КПІ". – 2011. – № 1. – С. 65 – 70.
2. *Воробієнко П.П.* Єдина система обмеження доступу до нецільових ресурсів мережі Інтернет в освітніх закладах України / П.П. Воробієнко, В.А. Каптур, В.А. Коляденко, В.О. Самодід // Комп'ютер у школі та сім'ї . – 2009. – № 8. – С. 30-34.
3. *Kaptur V.A.* Centralized system of http-traffic filtration / *V.A. Kaptur* // Наукові праці ОНАЗ ім. О.С. Попова. – 2010. – № 2. – С.16 – 19.
4. *SkyDNS.* [Електронний ресурс]. – Режим доступу: https://www.skydns.ru/.
5. *Internet Censor.* [Електронний ресурс]. – Режим доступу: http://www.icensor.ru/.
6. *DansGuardian.* [Електронний ресурс]. – Режим доступу: http://dansguardian.org/.
7. *KidsControl.* [Електронний ресурс]. – Режим доступу: http://www.kidscontrol.ru/.
8. *Rejik.* [Електронний ресурс]. – Режим доступу: http://rejik.ru/.