

УДК 004.056.52: 621.391

ВИЗНАЧЕННЯ УМОВ ДОЦІЛЬНОСТІ ЗАСТОСУВАННЯ КОМПЛЕКСНИХ СИСТЕМ ФІЛЬТРАЦІЇ КОНТЕНТУ

Каптур В.А., Поднебесний І.А.

*Одеська національна академія зв'язку ім. О.С. Попова,
65029, Україна, м. Одеса, ул. Ковальська, 1.*

vadim.kaptur@onat.edu.ua

ОПРЕДЕЛЕНИЕ УСЛОВИЙ ЦЕЛЕСООБРАЗНОСТИ ПРИМЕНЕНИЯ КОМПЛЕКСНЫХ СИСТЕМ ФИЛЬТРАЦИИ КОНТЕНТА

Каптур В.А., Поднебесный И.А.

*Одесская национальная академия связи им. А.С. Попова,
65029, Украина, г. Одесса, ул. Кузнечная, 1.*

vadim.kaptur@onat.edu.ua

DETERMINATION THE FEASIBILITY OF USING INTEGRATED CONTENT FILTERING SYSTEMS

Kaptur V.A., Podnebesny I.A.

*O.S. Popov Odessa national academy of telecommunications,
1 Kovalska St., Odessa, 65029, Ukraine.*

vadim.kaptur@onat.edu.ua

Анотація. На основі часової діаграми роботи комплексної та звичайної систем фільтрації контенту отримано аналітичні вирази, які пов'язують між собою їх основні характеристики. На основі отриманих виразів виведені умови доцільності застосування комплексних систем фільтрації контенту порівняно зі звичайними системами. Проведені розрахунки, які дозволили продемонструвати конкретні порогові значення кількості записів у списках блокування, а також порогові значення затримки при передаванні запитів від одного засобу фільтрації до іншого, за яких використання комплексної системи, що складається одночасно з проху- та DNS-серверів, є більш доцільним ніж використання одного проху-серверу.

Ключові слова: комплексні системи фільтрації контенту.

Аннотация. На основе временной диаграммы работы комплексной и обычной систем фильтрации контента получены аналитические выражения, которые связывают между собой их основные характеристики. На основе полученных выражений выведены условия целесообразности применения комплексных систем фильтрации контента в сравнении из обычными системами. Проведены расчёты, которые позволили продемонстрировать конкретные пороговые значения количества записей в списках блокирования, а также пороговые значения задержки при передаче запросов от одного средства фильтрации к другому, при которых использование комплексной системы, состоящей одновременно из проху- и DNS-серверов, является более целесообразным чем использование одного проху-сервера.

Ключевые слова: комплексные системы фильтрации контента.

Abstract. Analytical expressions that interrelate the main characteristics of an integrated and simple content filtering systems are obtained using timing diagram. On the basis of obtained expressions provided conditions for the use of the integrated content filtering systems in comparison of simple systems. The calculations, which will demonstrate the specific limitation of entries number in the blocking list, as well as limitation of delays in the transmission of requests from one part of medium integrated content filtering system to other, can help to determine conditions where using of integrated content filtering system is more appropriate than using a simple proxy server.

Key words: integrated content filtering systems.

Аналіз сучасного стану та перспектив розвитку методів фільтрації контенту в телекомунікаційних мережах дозволяє виділити серед багатьох інших окрему групу технологічних питань, які сьогодні потребують пильної уваги з боку дослідників. Так, наприклад, в [1] було проведено аналіз переваг і недоліків різних засобів фільтрації та запроваджено систему їх позначень. При цьому було уведено поняття комплексних систем фільтрації контенту (КСФК), тобто систем, що складаються з двох або більше засобів фільтрації, кожен з яких може використовувати різні види, методи та підходи до фільтрації контенту. Проте слід зазначити, що недостатній рівень дослідження спільної роботи різних систем фільтрації контенту (СФК) не дозволяє однозначно говорити про доцільність використання КСФК в тих або інших випадках.

Метою роботи є визначення умов доцільності застосування комплексних систем фільтрації контенту порівняно із класичними системами.

На рис. 1 наведено часову діаграму роботи звичайної та комплексної систем фільтрації контенту. Звичайна СФК, яка за схемою, запропонованою в [1], позначена як P|D|A|B|-, являє собою проху-сервер, який встановлюється безпосередньо в мережі, для якої має забезпечуватись фільтрація контенту. У свою чергу, КСФК, позначена як PD|DC|A|B|-, використовує не тільки проху-сервер, а й віддалений (централізований) DNS-сервер. Будемо вважати, що обидві з розглянутих систем використовують лише фільтрацію за адресою та працюють за однаковими підходами – на основі використання «чорних» списків.

Як видно з рис. 1, для звичайної СФК (верхня частина), побудованої із використанням лише проху-серверу, час опрацювання запиту визначається єдиною складовою – часом пошуку адреси в базі даних заборонених ресурсів (t_{proxy_all}). Для спрощення будемо вважати, що всі інші складові часу опрацювання запиту або є значно меншими за пошук, або не відносяться до процесу контентної фільтрації, а їх значення не залежать від типу та класу системи, що використовується.

У свою чергу, для КСФК (нижня частина рис. 1) час опрацювання залежить, насамперед, від типу запиту. Так, наприклад, якщо запит надійшов у вигляді IP-адреси, то цей час, як і в попередньому випадку, визначається процедурою пошуку адреси в базі даних заборонених ресурсів безпосередньо на проху-сервері за тією лише відмінністю, що до складу бази в цьому випадку входять лише IP-адреси, тому значення часу пошуку (t_{proxy_ip}) в цьому випадку буде меншим або рівним ніж аналогічне значення для звичайної СФК ($t_{proxy_ip} \leq t_{proxy_all}$). В іншому випадку, якщо запит надійшов у вигляді доменного імені, на час опрацювання запиту будуть впливати дві основні складові: час пересилання запиту або відповіді до/від централізованого DNS-сервера (t_{inet}) та час пошуку адреси в базі даних заборонених ресурсів безпосередньо на DNS-сервері (t_{dns}). Для спрощення будемо вважати, що час пересилання запиту та отримання відповіді через мережу Інтернет є однаковими.

Для порівняння часу опрацювання запиту зазначеними вище системами будемо вважати, що вони працюють з одним і тим самим списком блокування, що складається з $X_{all} = X_{IP} + X_{DN}$ записів, де X_{IP} – кількість записів наданих у формі IP-адреси, X_{DN} – кількість записів наданих у формі доменного імені (або URL). Очевидно, що при цьому час обробки запиту проху-сервером, який містить повний список (IP-адреси та доменні імена одночасно) буде залежати, насамперед, від способу збереження цих списків та алгоритму пошуку. Так, наприклад, якщо проху-сервер буде використовувати єдиний змішаний список, то час обробки буде визначатися лише загальною кількістю записів та середнім часом виконання операції із порівняння отриманого запиту із записом списку блокування на проху-сервері (τ_{proxy}), тобто $t_{proxy_all} = F(X_{all}, \tau_{proxy})$. В іншому випадку, якщо проху-сервер буде

використовувати два окремі списки та попередньо здійснювати аналіз отриманого запиту на предмет належності до групи доменних імен (URL-ів) або до групи IP-адрес, час обробки буде залежати не тільки від кількості записів в окремих списках, а й від типу отриманого запиту, а в загальному випадку від імовірності отримання запиту того або іншого виду.

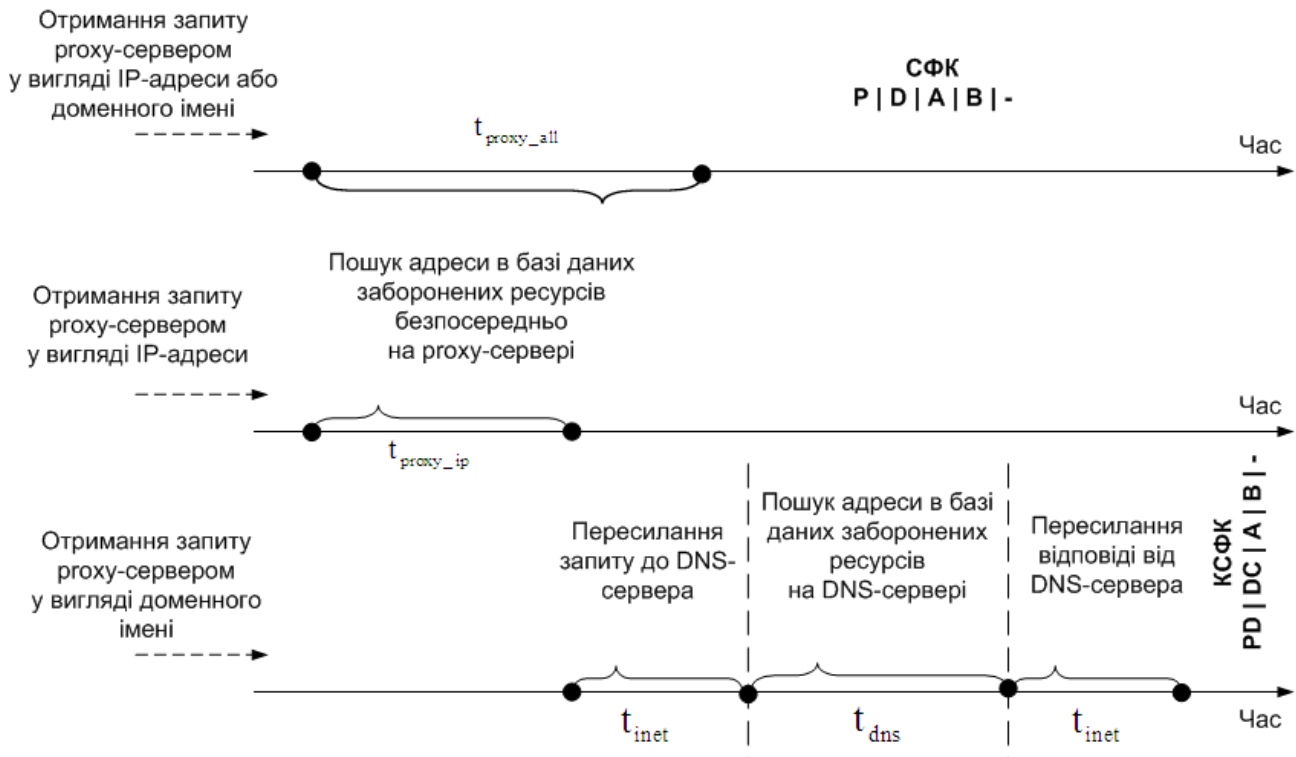


Рисунок 1 – Часова діаграма роботи СФК типу P|D|A|B|- та КСФК PD|DC|A|B|-

З точки зору порівняння СФК та КСФК за часовими характеристиками найбільшу практичну корисність мають відповіді на такі питання, як:

- за якої кількості записів у формі доменних імен (або URL-ів) використання КСФК є більш вигідним ніж звичайної СФК з розглянутого прикладу;
- за якого рівня якості підключення до централізованої частини КСФК (з розглянутого прикладу) є доцільним її застосування порівняно із звичайною СФК;
- яким чином мають змінюватись виробничі характеристики обладнання централізованої частини КСФК при збільшенні кількості записів у відповідних списках блокування та збільшенні кількості організацій, що її використовують.

Очевидно, що основною умовою за якої використання КСФК з розглянутого на рис. 1 прикладу є кращим (з точки зору часових характеристик) за відповідну СФК є
$$\begin{cases} t_{proxy_ip} < t_{proxy_all} \\ t_{dns} + 2t_{inet} < t_{proxy_all} \end{cases}$$
 за умов використання СФК змішаного списку та $t_{dns} + 2t_{inet} < t_{proxy_ns}$ за умов використання СФК окремого списку. При цьому у другому випадку ймовірний вигравш можливий лише при надходженні запиту, що має форму доменного імені.

Зважаючи на те, що умова $t_{proxy_ip} < t_{proxy_all}$ виконується для всіх випадків, у яких списки блокування не містять доменних імен (URL-ів), то найбільш цікавим, з точки зору

дослідження для обох наведених випадків, є умова $t_{dns} + 2t_{inet} < t_{proxy_{ns}} \leq t_{proxy_{all}}$, де $t_{dns} = F(X_{DN}, \tau_{dns})$, τ_{dns} – середній час виконання операції із порівняння отриманого запиту із записом списку блокування на DNS-сервері, а $t_{proxy_{ns}} = F(X_{DN}, \tau_{proxy})$.

Будемо вважати, що час обробки запиту сервером (проху або DNS) рівномірно розподілений на проміжку $\{\tau_{обр} \dots X \tau_{обр}\}$, а події запиту користувачем заблокованого і дозволеного ресурсу є рівномірними. Тоді математичне очікування [2] часу обробки запиту DNS-сервером буде $t_{dns} = \frac{\tau_{dns}(3X_{DN} - 1)}{4}$, а часу обробки запиту, що має форму доменного імені на проху-сервері, $t_{proxy_{ns}} = \frac{\tau_{proxy}(3X_{DN} - 1)}{4}$. У свою чергу, $t_{proxy_{all}} = \frac{\tau_{proxy}(3X_{all} - 1)}{4}$. Використовуючи ці значення отримаємо формалізовані умови доцільності використання КСФК для наведеного вище випадку:

$$\left\{ \begin{array}{l} X_{DN} < \frac{1}{3} - \frac{8t_{inet}}{3(\tau_{dns} - \tau_{proxy})} \\ 1 < X_{DN} \leq X_{all} \end{array} \right\} \quad (1)$$

Отриманий вираз (1) дозволяє дослідити залежність граничної (з точки зору доцільності застосування КСФК) кількості записів у формі доменних імен (або URL-ів) від співвідношення характеристик апаратних платформ DNS- та проху-серверів, а також від якості підключення (часом передавання запиту/відповіді) між проху- та DNS-сервером у відповідній КСФК.

Очевидно, що умова (1) виконується лише у випадку, коли середній час виконання операції із порівняння отриманого запиту із записом списку блокування на проху-сервері є більшим за відповідний середній час операції для DNS-сервера ($\tau_{proxy} > \tau_{dns}$). При цьому чим більшою є відповідна різниця (тобто чим DNS-сервер кращий за характеристиками ніж проху-сервер), тим меншою буде межа доцільності застосування КСФК.

На рис. 2 наведено графік зазначеної вище залежності для чотирьох значень часу відправки запиту (отримання відповіді) між проху- та DNS-серверами у складі єдиної КСФК: 1, 10, 25 та 50 мс. Як видно з графіка, при збільшенні різниці між середнім часом виконання операції із порівняння отриманого запиту із записом списку блокування на проху-сервері та DNS-сервері лише на 1 мкс практично для всіх кривих спостерігається двократне зменшення граничного значення кількості записів, що представлені у формі доменних імен. Проте слід зазначити, що при подальшому зростанні різниці зменшення відбувається менш стрімкими темпами, а також, що для випадку, коли DNS-сервер розміщений в одній локальній мережі із проху-сервером (час відправки запиту (отримання відповіді) між проху- та DNS-серверами у складі єдиної КСФК складає 1 мс) гранична кількість записів, починаючи з якої доцільно застосовувати КСФК в десятки разів менша ніж для випадків, коли канал зв'язку між проху- та DNS-серверами є не надто швидкісним (середній час доставки запиту/відповіді складає 50 мс).

Очевидно, що для визначення доцільності використання КСФК, необхідно також визначити граничні значення продуктивності DNS сервера, за якого буде виконуватися розглянута вище умова. Шляхом перетворення нерівності (1), отримано залежність необхідної продуктивності DNS-сервера залежно від середнього часу виконання операції із порівняння отриманого запиту із записом списку блокування на проху-сервері τ_{proxy} , кількості записів у списку заблокованих ресурсів у централізованому DNS-сервері X_{DN} , а

також від якості підключення (часом передавання запиту/відповіді) між проху- та DNS-сервером у відповідній КСФК: $\tau_{dns} < \tau_{proxy} - \frac{8t_{inet}}{3X_{DN} - 1}$.

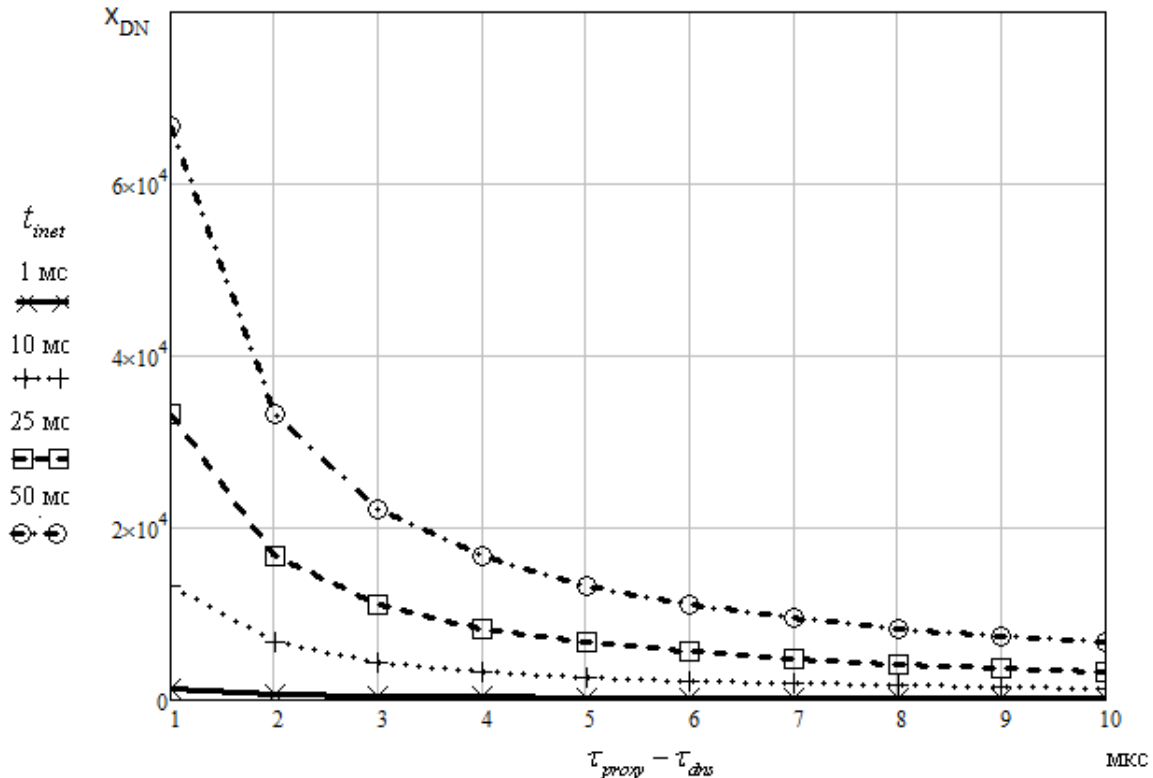


Рисунок 2 – Залежність граничної кількості записів у формі доменних імен від різниці між середнім часом виконання операції із порівняння отриманого запиту із записом списку блокування на проху-сервері та DNS-сервері

Для визначення значення середнього часу порівняння запиту з одним записом у списку заблокованих ресурсів у проху-сервері було проведено низку експериментів на трьох серверах з різною продуктивністю, побудованих на базі процесорів Intel Core i7, Intel Core2Duo та Intel Pentium 4 [3], з об'ємом оперативної пам'яті, що дорівнює 2 Гб та жорстким диском, підключеним по інтерфейсу SATA2. Для найбільш продуктивного сервера, побудованого на базі процесора Intel Core i7, значення τ_{proxy} склало $1,7 \cdot 10^{-8}$ с. На рис. 3 представлені графіки залежності граничного значення продуктивності DNS-сервера від кількості записів у списку ресурсів, які фільтруються, для чотирьох значень часу відправки запиту t_{inet} і трьох отриманих експериментально значень τ_{proxy} .

Як видно з графіка (рис. 3), застосування комплексної системи фільтрації є доцільним лише при досягненні деякої мінімальної кількості записів у таблиці заблокованих ресурсів на стороні DNS сервера, при чому це значення значно зростає навіть за незначного збільшення часу відправлення запиту t_{inet} . Враховуючи, що розглянута КСФК передбачає використання централізованих DNS-серверів, які будуть обслуговувати значну кількість заявок, що надходять від розподілених проху-серверів, припустимо, що значення t_{dns} становить 1 мкс. Виходячи з цього, можна зробити висновок, що при використанні низкопродуктивного проху-сервера з високошвидкісним підключенням до мережі Інтернет ($t_{inet} = 10$ мс), використання КСФК буде доцільним вже при 200000 записів у списку фільтрації DNS-сервера, в той час як при використанні високопродуктивного проху-сервера, граничне значення доцільності використання КСФК становить 800000 записів.

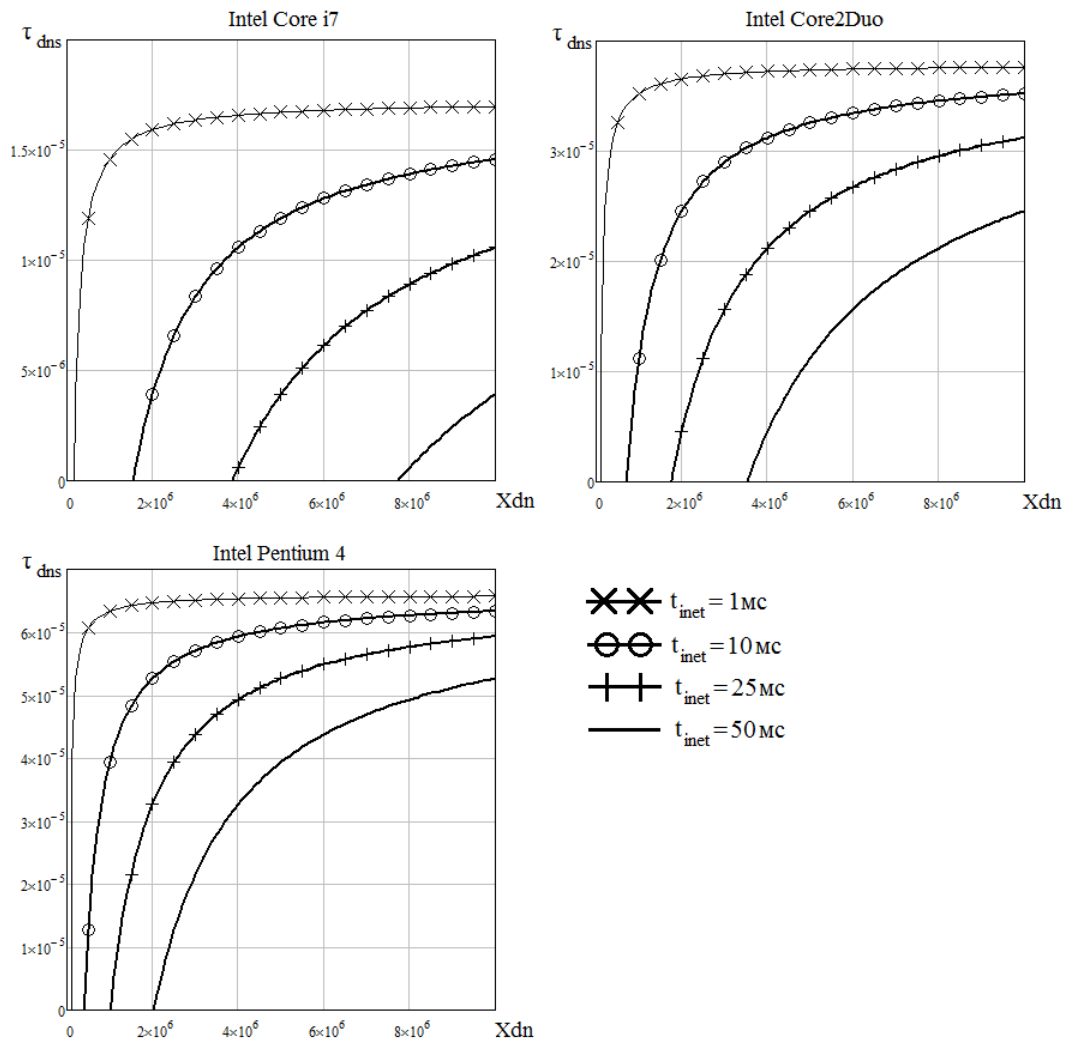


Рисунок 3 – Графіки залежності граничного значення продуктивності DNS- та проху-серверів від кількості записів у списку блокування

Висновки та результати:

1. Використання КСФК з централізованим DNS-сервером буде найбільш ефективно в організаціях, які не можуть установити високопродуктивний проху-сервер, проте мають широкосмугове підключення до мережі Інтернет. Прикладом таких організацій можуть служити загальноосвітні школи, публічні бібліотеки, невеликі державні установи.

2. При досягненні певної кінцевої кількості записів у списку блокування, застосування КСФК буде доцільно для будь-яких організацій при виконанні нерівностей $\tau_{proxy} > \tau_{dns}$ та $1 < X_{DN} \leq X_{all}$.

3. Отримані в результаті проведених досліджень математичні вирази, які дозволяють визначити граничні умови, за яких використання КСФК може надати користувачам кращу якість сервісу, ніж при використанні СФК, не враховують імовірнісних характеристик потоку запитів, які надходять на вхід систем фільтрації. Зважаючи на це подальші дослідження мають бути спрямовані на дослідження статистичного розподілу запитів для установ різних типів з метою формування відповідних профілів, що дозволять використовувати отримані вирази під час оцінки доцільності використання КСФК у конкретній організації.

ЛІТЕРАТУРА:

1. Каптур В.А. Комплексні системи фільтрації контенту в мережі Інтернет / В.А. Каптур // Наукові праці ОНАЗ ім. О.С. Попова. – 2013. – №1. – С. 16–21.
2. Крылов В. В. Теория телетрафика и ее приложения/ В.В. Крылов, С.С. Самохвалова. – СПб.: БХВ-Петербург, 2005. – 288 с.: ил.
3. Intel [Електронний ресурс]. – Режим доступу: <http://www.intel.ua/>.

REFERENCES:

1. Kaptur, Vadim. "Integrated Content Filtration Systems On The Internet" Scientific works of ONAT n.a. Popov 1 (2013): 16-21. Print.
2. Krylov, V., and C. Samohvalova. Teletraffic Theory And Applications. Sankt-Peterburg: BHV-Peterburg, 2005. Print.
3. Intel. Intel Corporation, n.d. Web. 24 June 2014. <<http://www.intel.ua/>>.