Костенко В. В.
асистент, Криворізький національний університет
Чубаров В. А.
к.т.н., доцент, Криворізький національний університет

ПРОБЛЕМА НЕКОРЕКТНОЇ КОМУТАЦІЇ ETHERNET КАДРІВ В МЕРЕЖЕВИХ КОМУТАТОРАХ РІВНЯ ДОСТУПУ.

Розглянуто причини та наслідки некоректної комутації Ethernet кадрів в мережевих комутаторах рівня доступу проаналізовано механізми моніторингу працездатності мережевого обладнання та їх недоліки.

Проблема некоректної комутації відбувається через hash конфлікт у комутаційний матриці. Під конфліктом hash мається на увазі така ситуація, коли через організацію внутрішньої логіки комутатора деякі MAC-адреси вважаються рівними один одному. У результаті такої "помилки" у таблицю комутації попадає тільки перший MAC, а кадри, призначені конфліктуючим MAC, поширюються по всій мережі. Уперше цей конфлікт, і його наслідки були помічені у провайдерських мережевих структурах рівня доступу ще в 2010-2012 роках. Деяким системним адміністраторам проблема вже була відома і вирішувалася вона різними способами. На офіційному форумі технічної підтримки комутаторів D-Link є коментар, який містить технічні подробиці цієї проблеми [1]. А в даному проекті буде описана природа цього явища, і запропонований варіант зменшення впливу цієї проблеми на якість обслуговування рівня доступу.

Причина "конфлікту MAC". У пам'яті комутатора MAC-адреси не зберігаються у своєму натуральному вигляді. Зберігаються лише деякі значення hash-функції, обчислені на основі MAC і VLAN[3]. Виглядає це приблизно так: val1=hashfunc(mac1+vlan1). Інший запис для іншої пари mac2 і vlan2 буде, відповідно, val2=hashfunc(mac2+vlan2).

На рисунку 3.6 наведена схема для більш детального розуміння причини "конфлікту MAC" у пам'яті комутатора.

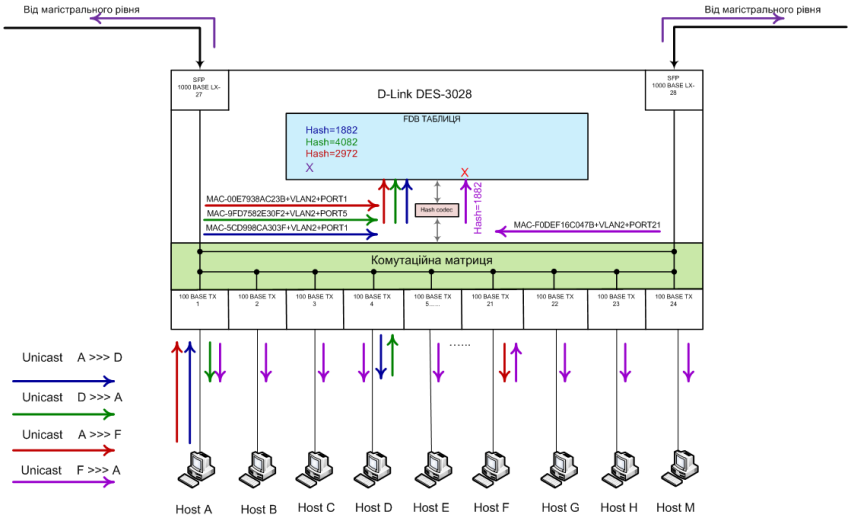

Рис. 1. Схема комутатора для більш детального розуміння причини "конфлікту MAC" у пам'яті FDB таблиці

При цьому для зберігання подібних значень пам'ять виділяється в умовах найсуворішої економії. Із цього випливає, що унікальність кожного запису реалізована з деякими допущеннями. Тому ймовірність того, що val2 буде рівно val1 суттєво відрізняється від нуля. Якщо таке відбувається, то mac2 уважається рівним mac1. Із усіх таких MAC-адрес комутатором буде вивчений тільки перший, тому що відмінностей між ними комутатор не бачить.

З модельного ряду D-Link моделі в порядку убування ймовірності конфлікту можна розташувати так:

-DES-3028 (чипсет BCM 5347), DES-1228/ME/A1 (чипсет BCM 5347, апаратний аналог 3028)

-DES-3200-28/A1/B1 (чипсет BCM 53262), DES-1228/ME/B1 (чипсет BCM 53262), DES-1210-28/ME/B2 (чипсет BCM 53262)

-DES-3528, DES-3526

-DES-3200-28/C1

Можливо, варто було б помістити серію 35xx і 3200-28/C1 в один ряд, але точних описаних даних у літературі знайти не вдалося.

Офіційні коментарі від інженерів D-Link говорять, що чипсети всіх сучасних моделей піддаються даним проблемам. Оскільки основних виробників усього два: Broadcom і Marvell, то в багатьох вендорів спостерігається дана проблема тією чи іншою мірою на різних серіях комутаторів. Чипсети минулих моделей, мали 2-х рівневий хеш, а не однорівневий як нові, тому на них практично відсутня дана проблема, на що вказують тести вендора D-Link. Виробники чипсетів збільшили швидкість роботи FDB(Forward DataBase) таблиці, але як побічний ефект одержали проблему з хешами. Також багато чого залежить від того, як реалізована пам'ять під FDB таблицю й скільки біт відпущене під один хеш запис.

Наявність проблеми хешування MAC зовсім не говорить про те, що дану модель застосовувати неможливо. Кожна модель комутатора має свою маркетингову нішу й припускає її правильне використання. Наприклад моделі DES-3028 і DES- 3200 більшою мірою розраховані на використання з операторською моделлю QinQ. Модель DES-3528 розрахована на корпоративний сегмент, де ціна встаткування має менше значення ніж необхідний функціонал. Модель Des-1210-xx розрахована на використання з операторською моделлю без QinQ.

Як видно із проведених тестів [3] у моделі DES-3028 найнижча стійкість хэш функції до довгих послідовностей (найбільший відсоток колізій утворених хешей), тому при використанні даної конкретної моделі було рекомендовано уникати їх багатокаскадних послідовних з'єднань. В інших моделях дана проблема відсутня.

При експлуатації даних комутаторів, підтверджується, що на моделі DES-3028 проблема виражена гостро, на DES-3200-28/A1/B1 - проблема проявляється менше, а на інших моделях (з перерахованих) може бути помічена тільки при явно помилковому проектуванні мережі.

Діагностика й виявлення проблеми.

а) Пристрій у мережі доступний, MAC-адреса коректна (див. біт для групового розсилання), але на порту не виявляється.

б) За допомогою спеціального функціонала enable flood_fdb. Комутатор почне стежити за MAC-адресами й вести в пам'яті копію (тобто цей функціонал - винятково моніторинг) конфліктів. Переглянути таблицю конфліктів можна командою show flood_fdb, приклад виконання якої наведений у таблиці 1.

Табл. 1. Приклад show flood_fdb - конфлікт FDB

Value	VLAN ID	MAC Address	Time Stamp
3865	24	00-22-B0-04-6A-17*	9978796
3865	1511	00-1A-79-11-85-E4	9978796
2438	115	00-D0-5C-78-2D-70	2716954

Однакове value указує на конфлікт. Зірочка говорить про те, що MAC-адреса присутня у таблиці комутації. Таким чином, "проблемним" у даному прикладі є MAC 00-1A-79-11-85-E4.

В DES-3200-28/C1 такого механізму немає, тому що вважається, що дана модель не піддана проблемі "конфлікт hash".

Самою очевидною проблемою, звичайно ж, буде відсутність усіх "пересічних" MAC-адрес у таблиці комутації. При одержані кадра, адресованого такому MAC-у, комутатор не зможе визначити порт призначення DLF (Destination Lookup Failure) і кадр буде відправлений в усі порти, які є учасниками даного VLAN, крім того у порт, звідки даний кадр був отриманий. Тим самим комутатор (світч) перетворюється в концентратор (хаб), тобто не справляється зі своїм основним завданням - комутацією трафіка. Чим більше пересічних MAC-адрес у мережі, тим більше зайвого трафіка по ній переміщається. Якщо трафік поширюється на абонентів, які оплачують різні тарифні плани, то у випадку "конфлікту hash" трафік абонента з більшою смугою, потрапить у порт абонента з меншою смугою. Частина кадрів при такій ситуації може втрачатися й кінцевий споживач не одержує небхідну якість надаваного сервісу.

Це й було виявлено засобами моніторингу навантаження на абонентські порти рівня доступу. Скрін навантаження абонентського комутатора, де зіставлені графіки завантаження портів у момент проблеми, надано на рисунку 2.

З рисунка 2 видно, що в моменти часу з 4 до 6 ранку, з 12:00 по 12:15 а так само з 14:00 по 14:15 трафік у всіх графіках ідентичний по своїй структурі. Отже, природа виникнення цього трафіка є саме конфлікт hash MAC-адрес на комутаторі рівня доступу.

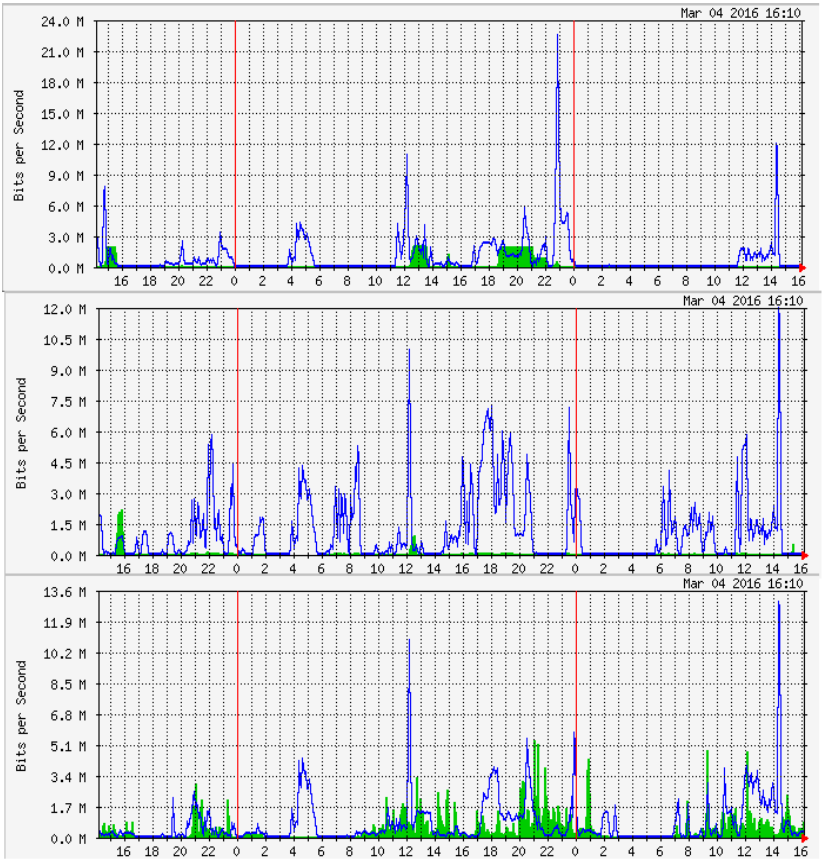

Рис. 2. Графіки завантаження портів у момент проблеми конфлікту hash

Одночасно з моніторингом бази даних завантаження портів, було ухвалене рішення включення команди show flood_fdb у базу даних подій.

Таким чином, щоразу коли комутатор сам у себе своїми засобами виявляє конфлікт hash, у базу даних попадає й фіксується ця подія. На рисунку 3. зображений фрагмент інформації з фіксації конфлікту hash засобами самого комутатора і поява запису про цей конфлікт в базі даних подій комутатора.

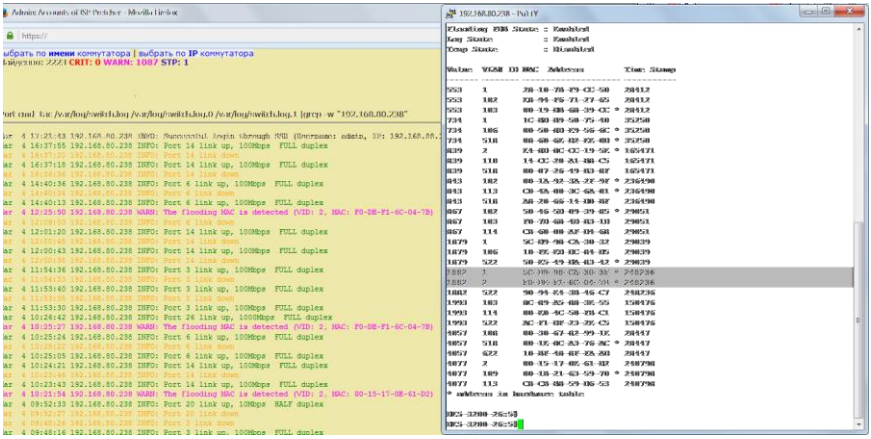

Рис. 3. Фрагмент інформації з фіксації конфлікту hash засобами самого комутатора і поява його в базі даних подій.

При детальному аналізі й зіставлення отриманих фактів, удалося з'ясувати, що час на графіках завантаження портів у момент проблеми збігається із часом у таблиці подій. У такий спосіб вдається знаходити мережні пристрої, яких з'являються конфлікти hash.

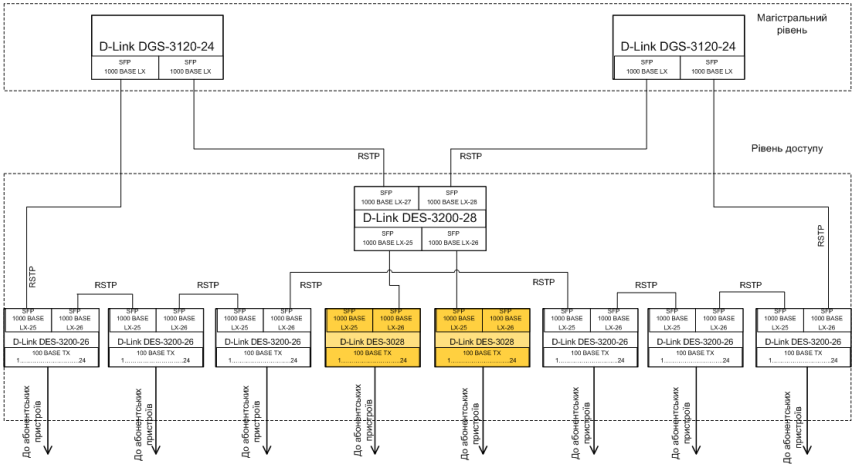

Рис. 4. Схема зміненої фізичної топології мережі рівня доступу з метою мінімізації конфлікту hash.

Як варіант розв'язку, пропонується зменшити вплив конфлікту на якість надаваного сервісу. Пропонується фізичне виключення проблемних комутаторів з кільцевої топології й включення їх по окремому фізично незалежному оптичному лінку. А у випадку неможливості такого розв'язку або відсутності вільних волокон, пропонується сегментація мережі на Vlan-и меншого розміру, аж до vlan-per-user. Широкомовний домен при цьому поменшається, кількість переданого по мережі трафіка теж. На рисунку 4 наведена схема зміненої фізичної топології мережі рівня доступу з метою мінімізації конфлікту hash на якість надаваного сервісу.

У випадку vlan-per-user проблема буде зведена до мінімуму, але не виключена повністю. І причина тому є присутність ще двох VLAN, які не можуть бути зменшені до мінімальної кількості VLAN-ів. Це виникає у зв'язку з присутністю керуючого (management) і мультикаст (mvr) VLAN-ів. Оскільки конфлікти запросто відбуваються між різними VLAN, то перетинання з абонентським MAC може викликати "флуд трафіку" керуючого VLAN-у або погіршення роботи IPTV.

На рисунку 4. наведений приклад конфлікту hash між різними VLAN.

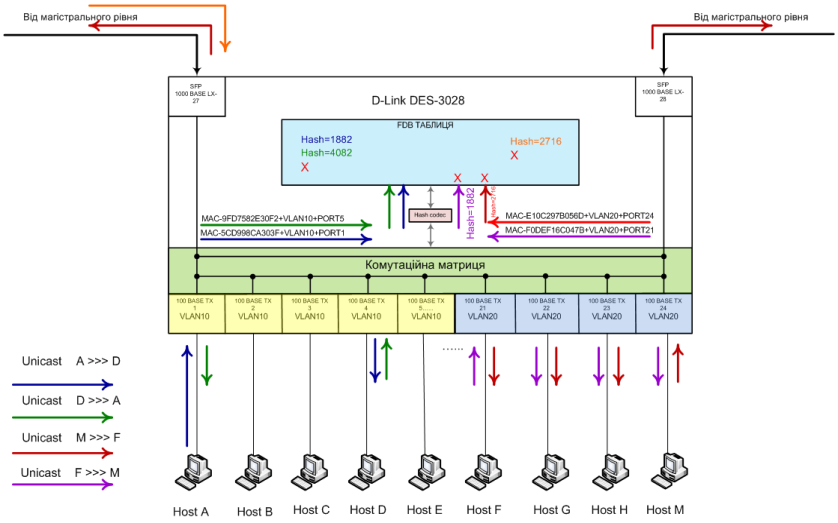

Рис. 5. Приклад конфлікту hash між різними VLAN.

З рисунка 5. варто зробити висновок, що MAC E1-0C-29-7B-05-6D і F0-DE-F1-6C-04-7B із-за конфлікту hash не можуть потрапити до FDB таблиці.

Отже, конкретно для цих MAC-адрес комутатор буде працювати в режимі HUB і отже трафік усередині сегмента конкретно цього VLAN (у наведеному випадку VLAN20) буде поширюватися на всі порти.

Наступна проблема - сам механізм виявлення таких конфліктів (enable flood_fdb), де проводяться обчислення, аналогічні тим, що виконуються в ASIC (Application-Specific Integrated Circuit)[3]. Тобто це не добування проблемної адреси з комірок пам'яті, а обчислення, що проводиться паралельно роботі чипа. Тільки в цьому випадку витрачаються вже ресурси CPU. Звідси випливає, що великий потік трафіка може привести до непотрібного навантаження на CPU комутатору. На практиці, на жаль, так і виходить. У моделі DES-3028 навантаження на CPU може доходити до 100% тим самим роблячи пристрій недоступним. При тому, якщо комутатор виступає в ролі релей-агента, то абоненти перестають одержувати адреси від DHCP. На DES-3200-28/A1/B1 ситуація трохи відрізняється - комутатор з деякою ймовірністю не може відповісти на ARP-запит. Коли час життя ARP (Address Resolution Protocol) на маршрутизаторі минає, комутатор на якийсь час стає недоступний. А оскільки моніторинг за працездатністю всіх комутаторів рівня доступу виконаний за принципом перевірки їх доступності, то результатом може стати періодично повторювана " аварія" на карті моніторингу мережі. Це у свою чергу може ввести в оману службу моніторингу, в обов'язки якої входить контроль над працездатністю комп'ютерної мережі в цілому.

ЛІТЕРАТУРА

1. des-3200 ХЭШ [Електронний ресурс] Режим доступу до статті: http://forum.dlink.ru/viewtopic.php?p=653278#p653278
2. Memory management unit architecture for switch fabric EP 1168727 A2 [Електронний ресурс] Режим доступу до статті: http://nag.ru/upload/images/15587/img_EP1168727A2.pdf
3. Проблема хеш коллизий [Електронний ресурс] Режим доступу до статті: http://nag.ru/articles/reviews/15587/raz-tablitsa-dva-tablitsa.html