
Хакофон OpenBSD: Исправление цикла ожидания

Джереми Эндрюс (Jeremy Andrews), Пятница, 27 Май 2005, 02:56

Боб Бек (Bob Beck) - разработчик [OpenBSD](#) из Эдмонта в Канаде. Он - один из приблизительно 60 разработчиков OpenBSD, в настоящее время занятых в необъявленном отеле где-то в центре города Калгари на [хакофоне OpenBSD 2005](#). Боб участвовал в [настройке инфраструктуры](#), и отвечал за [ежегодное барбекю](#) дома у создателя OpenBSD, Тео де Раадта ([Theo de Raadt](#)). После двух дней работы, в течении которых он помогал воплощать хакофон в реальность, он наконец сел, чтобы позаниматься `sramd` и почитать электронную почту. Одно из электронных писем в его почтовом ящике привлекло его внимание, что вылилось в работу на целый день, о которой он говорит: "Некоторые дни длятся очень, очень, очень долго..."

В следующей статье, Боб от первого лица расскажет про исследование, которое началось просто как проблема производительности RAID, но, в конечном счете, привело к проблеме с циклом ожидания. После исправления это дало внушительный прирост производительности. Боб отмечает: "Цикл ожидания - это место, где ядро вращается, пока нет никакой работы в пространстве пользователя. Там же происходит перехват и обслуживание многих прерываний и передача задач в очереди на устройства, после чего они делают `tsleep` в ожидании прерывания от карты, сигнализирующего, что работа сделана." Боб объясняет, что до сегодняшнего исправления, прерывания обрабатывались нормально только тогда, когда работали пользовательски приложения, но не когда система простаивала, и ядро просто ждало ввода/вывода устройства. Читайте полный отчет Боба об этом дне, потраченном на изучение проблемы, создание исправления и анализ показателей производительности.

Боб Бек пишет:

html]

День на OpenBSD хакофоне... Некоторые дни длятся очень, очень, очень долго...

Итак, это понедельник на хакофоне OpenBSD. Я устал. Я провел два дня, собирая здесь целый мир, и делая барбекю для всех с Тэо. Теперь настало *мое время*. Я могу сесть и заняться тем, чем хочу. Я открываю несколько окон emacs окон и начинаю переделывать внутренности `sramd` и читать почту. Хм, интересное сообщение от Марко.

Марко Перебума (Marco Pereboom) и меня в течение нескольких месяцев преследовали проблемы работы RAID. Ужасная скорость чтения, и это для нескольких машин. Некоторые быстры, а некоторые действительно тормозят. Марко, Арт Грабовский и Тоби Веингартнер (Marco, Art Grabowski, Toby Weingartner) обнаружили условие в `locore.s` в цикле ожидания, при котором мы можем терять прерывания! Хм. Надо взглянуть. Ага, гонка! ОК, здесь нам поможет Микки Шалэйефф (Mickey Shalayeff), а Тоби и Дэйл Ран (Dale Rahn) помогут нам погрузиться в ассемблерные исходники `locore.s`. Так, я начинаю закапываться в это и чувствую, что я нахожусь уже не в

<http://www.osrc.info/plugins/content/content.php?content.98>
пространстве пользователя. А и черт с ни, это серьезно, мне
действительно хочется это исправить и это интересно...

Ладно, мы устраняем гонку, и скорость возрастает с 14 МБ/с до 50 МБ/с. Уже лучше... но есть еще какие-то проблемы... А это что? Хм... похоже, что наш цикл ожидания влияет код APM! Если мы не будем дергать его слишком часто, то производительность пойдет вверх... Марко и Тоби начинают экспериментировать с запросом APM только каждые N раз, думая о том, чтобы вставить заглушку. Я ненавижу заглушки. Я говорю: "Вызывайте этот код, только если armd активен!". Круто! Новый скачек производительности до 100 МБ/с. Но постойте! Вступает Тео. Нет, мы не можем полагаться на наличие пользовательских программ только потому, что некоторые глючные ноутбуки настаивают на наличии APM для управления температурой. Мы не можем на это рассчитывать... (Ворчание, вздохи...) Назад к чертежной доске, и Тоби мудро решает прочитать спецификацию APM... Когда все прояснилось, мы были практически ослеплены, нарастает шум из угла и все начинают участвовать в работе. Когда мы вызываем функцию простоя APM из цикла ожидания, APM BIOS может выполнить инструкцию 'hlt'. При этом, когда мы получаем прерывание, то мы просыпаемся, но наш цикл ожидания исполняет *еще одну* инструкцию 'hlt' после вызова APM, вместо того, чтобы работать дальше!. (Я буду долго вспоминать взгляд Тэо, поскольку я очень долго объяснял ему это...)

Теперь, когда мы знаем что работает неправильно, давайте исправим это. Найдите правильную семантику. В комнате поднимается шум. Тео де Раадт, Арт, Толо, Марко, Бек, Вейнгарт, Никлас, Дэйл, Нейт и другие начинают бросать различные идеи. "сделайте "... "нет, слишком "... "сделайте мягкую "... "нет мы не можем делать здесь 'hlt', в отличие от APM"... "сделайте "... "нет уж, мы ненавидим "... "сделайте все по таймеру "... "Эй, "... (Арт напечатал что-то ужасное на моем экране)... "Вот, делайте "... Да! Точно. Проверяем счетчики профилировщика, и вызываем функцию idle BIOS только когда счетчики CPU_IDLE растут. (в отличие от проверки на каждый вход в цикл ожидания, когда мы, возможно, находимся в драйвере, сделавшем tsleep). Да, это правильный выбор для нас ...

Теперь ami выжимает 105 МБ/с.

Мой ноутбук улучшил свои показатели с 12 МБ/с до 25 МБ/с ...

Ноутбук Нейта - с 12 МБ/с до 40 МБ/с ...

"Нифига себе, раньше быстро работали только машины с неправильными APM!"

"это также влияет на криптографию в пространстве пользователя..."

"это затрагивает драйвера Ethernet"

"Эй, да от этого увеличивается срок службы аккумулятора".

"Это затрагивает ВСЕ на i386"

Мы идем за пивом, ждем появления отчетов тестирования, и без сомнений, это было здорово. Возвращаемся и публикуем... Ух ты, да мы провели целый

Да, а я все еще не начал свою работу в пользовательско пространстве... но, все-таки, на хакофоне случаются странные вещи. Такие, когда в комнате собирается достаточное количество народу, чтобы выловить проблему и действительно ее понять. Иногда бывает забавно забраться в те направления работ, о которых ты и не думал, что когда-либо будешь ими заниматься.

/html]