



Эффективность работы сетевых сетевых приложений приложений

Юрий Кравцов

по материалам компании Fluke Networks

Введение

Успешное управление уровнем сервиса в сетях (Service Level Management, SLM) требует применения упреждающих мер, в то время как большинство отделов ИТ сегодня работает в режиме реагирования. Предотвратить проблему проще, чем устранять ее последствия. Данная статья покажет, как ваш отдел ИТ сможет перейти от устранения проблем в пожарном порядке к планированию работы сети с помощью соответствующих инструментов.

Управление уровнем сервиса (SLM) получило широкое распространение как метод рационального использования ИТ-ресурсов для нужд конкретного бизнеса. Основная идея SLM - постоянный контроль качества поставляемого информационного сервиса. Его задача - удовлетворять требования клиентов к сети и постоянно повышать эффективность работы. Применяя такой подход, заказчик получает средство оценки ИТ по доходности вложений, а не по общей стоимости владения.

В статье обсуждаются решения, нужные для реализации успешной SLM-стратегии. Рассматривается, как правильно задать соглашения об уровне сервиса (Service Level Agreements, SLA), используя целевые параметры (Service Level Objectives, SLO) для достижения целей конкретного бизнеса. Описаны часто встречающиеся ловушки и показано, как их избежать, рассмотрены способы выбора параметров SLO и возможности профилактики с помощью уже существующего инструментария.

Правильный выбор целевых параметров SLO очень важен для успеха, он зависит от того, на чем должны быть сфокусированы ваши измерения - на работе клиента, сервера или сети. Выбор статистических параметров (средних значений или процентных соотношений) влияет на стратегию управления; необходимо правильно задать пороговые значения, которые определяли бы выполнение или невыполнение требований к системе. В статье оценены различные варианты мониторинга, их сильные и слабые стороны и потенциальное влияние на бизнес.

Традиционное управление уровнем сервиса основано исключительно на мониторинге доступности ресурсов. Сервис (сеть, сервер или приложение) должен быть "живым", "не падать" до 99.999% времени. Конечно, пользователю такая оценка сервиса кажется действенной и показательной. Однако такой подход не поможет достичь ключевых целей SLM, удовлетворить требования клиентских приложений и обеспечить постоянное улучшение характеристик, и причина тому проста: "живой" сервис может иметь такие низкие рабочие характеристики, что его будет невозможно использовать. Кроме того, такой подход не только не содействует росту эффективности, но и приводит к обратному

результату - ведь он фокусирует внимание менеджмента на редко происходящих событиях, а не на типичном поведении системы. Для максимизации эффективности необходимо, чтобы управление не только основывалось на факте доступности или недоступности ресурсов, но и жестко привязывалось к конкретным рабочим характеристикам. Следует учитывать не только текущий статус оборудования и инфраструктуры, но и реальную практику работы конечного пользователя. Главное - применяемый способ должен экономить вам время, а не приводить к его расходу. К счастью, инструментарий SLM уже достиг такого уровня развития, который может удовлетворить всем этим требованиям. Современные методы SLM должны активно поддерживать переход от политики реагирования к превентивному управлению за счет применения четырех ключевых средств: многоуровневых отчетов, раннего обнаружения проблем, быстрого их разрешения и подбора других вариантов. Кроме того, все эти функции должны быть просты в освоении, управлении и применении.

Для эффективного управления SLM необходимо точно определить целевые параметры SLO, которые будут выступать в качестве критериев. Существует три ключевых переменных, которые необходимо оценить: время отклика для конечного пользователя, время реакции сервера и задержка сигнала в сети. Однако способ измерения может быть разным (пассивным или активным), и это определяет, достигнете ли вы желаемого результата. Параметры SLO могут быть основаны на средних значениях времени, на процентном соотношении средних значений или на процентном соотношении транзакций. Многие существующие инструментальные средства позволяют отслеживать параметры SLO по усредненным по времени значениям, но ведь средние значения не обязательно отражают реальное состояние большинства пользователей! В чем-то это похоже на "среднюю температуру по больнице". Отслеживание SLO на базе процентного соотношения транзакций, напротив, прекрасный метод с технической точки зрения, он точно соответствует реальной ситуации у пользователей. Однако сейчас практически не существует решений, которые могли бы использовать этот подход в масштабах всего предприятия.

Настройка пороговых значений в параметрах SLO должна основываться на реальных потребностях пользователей. Первое пороговое значение - нижнее - отражает границу, ниже которой пользователь начинает испытывать проблемы с сетью. Второе - верхнее - показывает, в какой момент недостаточная эффективность системы начинает приводить к существенным издержкам, которых надо избегать.

Превентивное управление с помощью уровней сервиса SLM

Для успешного внедрения SLM почти всегда приходится прилагать усилия, чтобы изменить мышление ИТ-специалистов заказчика. Подавляющее большинство отделов ИТ решает проблемы по мере возникновения, причем почти всегда в пожарном порядке. Сотрудники отдела постоянно заняты, но суммарная эффективность работы низка. Применяя же подходы SLM, отделы ИТ могут предвидеть проблемы заранее, быстро находить их источники и устранять сбои до их возникновения, переходя от реагирующей схемы работы к упреждающей. Подобное изменение поведенческой модели, конечно, требует обучения персонала, но зато специалисты получают нужный стартовый импульс и затем оценивают рабочие характеристики сети с новой точки зрения. Инструментарий SLM - это не просто методы мониторинга и анализа. Он гарантирует, что необходимые ресурсы будут соответствовать потребностям пользователей и целям бизнеса. Первое требование для внедрения этой системы - обеспечить свободное время для стратегических действий. Некоторые средства настолько громоздки в установке, управлении и использовании, что у ИТ-специалистов не остается времени ни на что другое. Чтобы быть эффективным, средство SLM должно быть легко в управлении и предоставлять функциональные возможности, которыми действительно можно пользоваться. Простота использования инструментария SLM зависит от того, насколько быстро можно установить систему, удобно ли ее использовать, ежедневно управлять ею. Все это определяется архитектурой средств SLM и особенностями той среды, в которой они применяются. Средство или система, которыми невозможно управлять в масштабах огромного предприятия, могут оказаться пригодными для организации меньшего размера. Инструментарий, слишком дорогой для маленького предприятия, возможно, будет приемлем для большой корпоративной сети. Если же инструмент требует постоянной координации между несколькими ИТ-группами (например, одни отвечают за настольные приложения, другие - за приложения глобальной сети), то это может вызвать напряженность между отделами, и в результате большинство попыток увеличить эффективность окажутся бесполезными... Средства SLM-управления должны активно поддерживать, а точнее, провоцировать переход от реагирующего управления к превентивному. Для этого используются четыре основных типа функциональных возможностей: многоуровневые отчеты, раннее обнаружение проблем, быстрое их разрешение и поиск альтернативы. Все они будут рассмотрены дальше.



Значимые переменные

Один из первых этапов реализации SLA – выбор ключевых переменных. На каких переменных будет основана система соглашений об уровнях сервиса? В реальной жизни конфликт между желаниями пользователя и возможностями отдела ИТ возникает часто. Пожелания конечного пользователя обычно трудно численно определить, их смысл неоднозначен, хотя они так или иначе связаны со временем отклика системы. ИТ-специалисты же всегда хотят иметь четкие параметры, которые легко контролировать. Если ИТ-отдел не контролирует группу серверов, то он не хочет и нести ответственность за проблемы с ними. Компромиссным решением может стать проведение более широких измерений при более четком дележе ответственности: нужно будет отслеживать все важные переменные и детально расписать штрафные санкции за их несоответствие желаемому значению.

Производительность работы для конечного пользователя

Время отклика на запрос конечного пользователя нужно отслеживать независимо от того, применяются ли соглашения SLA в системе или нет. Это облегчает обмен информацией между пользователем и специалистами ИТ, ведь самый простой способ выразить в цифрах ощущения пользователя

от работы сети – измерить время транзакций и его составляющих. Нужно ли отслеживать в пассивном режиме реально существующих в системе пользователей или следует создать агентов специально? Первый вариант важен для достижения целей SLM, поскольку именно ради реальных пользователей все и делается. Зато второй дает на выходе детально определенную базовую характеристику системы, что полезно для устранения неполадок во всей сети. Лучше использовать комбинацию пассивного мониторинга реальных пользователей и нескольких искусственно созданных агентов, тогда вы получите преимущества каждого из методов.

Производительность работы сервера

Время отклика сервера также нужно отслеживать, независимо от применения соглашений SLA. Очень важно уметь быстро определить, серверы ли виноваты в том, что ухудшается время отклика на запрос конечного пользователя.

Есть несколько проблем с оценкой времени отклика сервера. Если для повторного выполнения одних и тех же транзакций используются искусственно созданные агенты, то результаты сразу же становятся недействи-

тельными, поскольку не связаны с реальным пользователем. Если сервер выполняет кэширование информации, то данные о задержках нельзя считать достоверными. Если транзакции выполняются в случайном порядке, то теряется главное преимущество искусственно созданных агентов – их детерминизм. Эти проблемы

можно устранить за счет пассивного мониторинга характеристик сервера по всем транзакциям и всем

пользователям, заодно это со-

здаст прочные предпосылки к обеспечению производительности в будущем.

Характеристики работы сети

Задержка в сети состоит из пяти компонентов: сериализации (преобразования в последовательную форму), организации очередей, задержки распространения сигнала в среде передачи, задержки обработки данных и задержки протокола. Сериализация или задержка передачи сигнала – это время, необходимое для отправки всех битов пакета по среде передачи. Оно зависит от размера пакета и от скорости доступа.

Задержка организации очередей – это время, в течение которого пакет находится в буфере и ожидает своей очереди на передачу.

Задержка распространения сигнала или задержка по расстоянию – это время, в течение которого пакет физически перемещается по среде передачи. Оно зависит только от расстояния и типа среды передачи.

Задержка на обработку – это время, которое требуется маршрутизатору или коммутатору на то, чтобы подготовить пакет к доставке. Оно обычно пренебрежимо мало в сравнении с другими задержками.

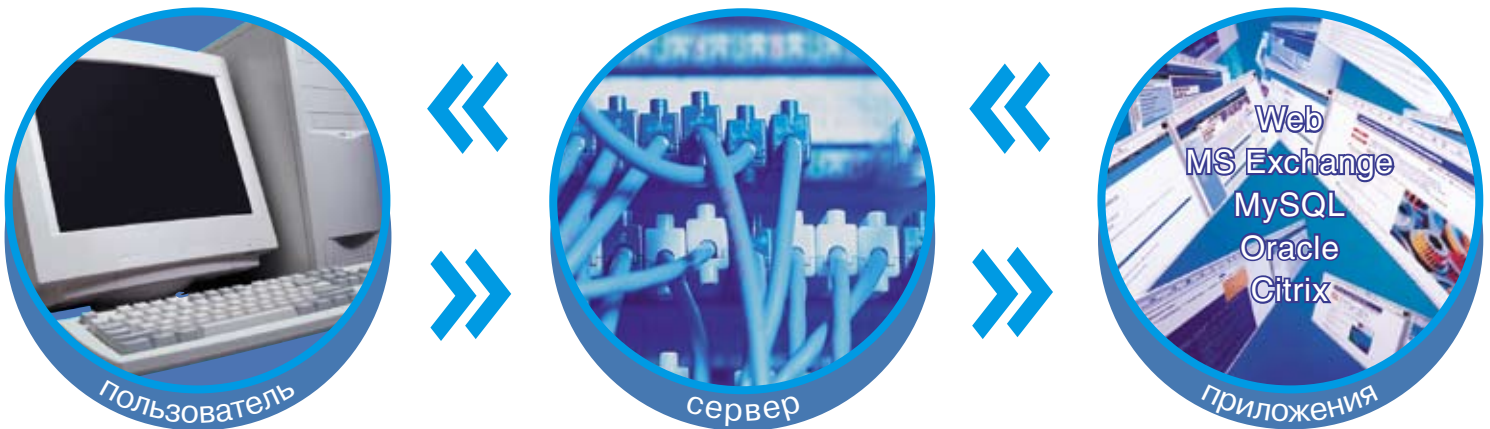
Задержка по протоколу – это время ожидания пакета из-за базовых протоколов; например, в разделяемой среде пакет должен ждать, пока нужный узел не запросит доступ. Значение такой задержки может быть самым разным в зависимости от протокола. Измерение задержки сигнала за счет отслеживания общих прикладных пакетов дает наиболее точные данные о производительности сети, поскольку измеренные значения отражают именно то, что в действительности испытывают пользователи.

Доступность сервисов

Доступность сервиса нужно отслеживать в явном виде – это часть стратегии управления SLM. При традиционных подходах к управлению отслеживается доступность сети и сервера. Можно дополнить это активными агентами или программными зондами, которые будут периодически пробовать выполнить выбранные транзакции. Если такие средства настроены на запуск каждые 15 минут, то в среднем уже через 7.5 минут существенный сбой будет обнаружен. Однако промежуточные кратковременные сбои могут остаться незамеченными, и их не удастся отследить как параметр SLO. Более частый опрос, конечно, поможет обнаружить и такие кратковременные неисправности, но только за счет дополнительной нагрузки на систему.

представления о том, что испытывают пользователи на практике при работе в сети. Например, пусть девять пользователей наблюдают время отклика в 0.5 секунды, а десятый – в 90.0 секунд. Среднее время отклика составит 9.5 секунд, и это отличается на целый порядок от того, что в действительности испытывает каждый из пользователей. По этой причине управление на основе средних значений может оказаться слишком сложной задачей. Если десятый пользователь получит время отклика в 180.0 секунд, а не 90.0, при том, что у остальных пользователей показатель останется по-прежнему 0.5, то среднее значение почти удвоится – и это несмотря на то, что только один пользователь из всех столкнулся с ухудшением характеристик. Некоторые производители опираются на усеченное среднее значение, чтобы уменьшить такую чувствительность к отдельным значениям (выбегам); для этого они отбрасывают любое измерение, превышающее предварительно установленное пороговое значение. В предыдущем примере пороговое значение в 2 секунды приведет к тому, что усеченное среднее составит 0.5, без учета значения 90-секундного выбега. Однако в таком подходе есть опасность: за счет отсека он может замаскировать существующие проблемы с производительностью. Если время отклика для семи пользователей уве-

тывает пользователь. Если 95% транзакций имеют время отклика менее 3 секунд, то значения остальных 5% незначительны. Соглашения SLA на базе усеченных средних значений игнорируют все значения времени отклика, которые превышают предварительно установленный порог; если же все значения времени срабатывания превышают порог, то измерения как такового и вовсе нет. Соглашения SLA на основе процентного отношения транзакций игнорируют только заранее установленное процентное отношение (в нашем примере это 5%) значений времени отклика. Соглашения SLA на основе процентного отношения транзакций предпочтительнее, чем SLA на основе временных средних значений, однако тогда и выбор производителя системы SLM будет более ограниченным. С технической точки зрения отслеживать и давать отчет о процентном соотношении по сравнению со средними значениями – более сложная задача, так что этот вариант поддерживает меньше производителей. Некоторые производители выбирают гибридный метод составления отчетов, выдавая процентное соотношение средних значений, а не процентное отношение транзакций. Соглашения SLA на базе такого гибридного метода потребуют, например, чтобы 95% из 5-минутных средних значений в течение месяца составляли менее 5 секунд.



Переменная статистика

Следующее важное решение при реализации системы SLM касается статистики. Должны ли соглашения SLA базироваться на средних значениях по времени или на процентном отношении транзакций? SLA на основе временных средних потребует, например, чтобы 95% транзакций имели время отклика менее 3 секунд. Преимущество выбора SLA на базе временных средних значений состоит в том, что почти каждый производитель систем SLM поддерживает их определение, обеспечивая достаточно свободный выбор инструментария. Но временные средние не дают точного

личивается с 0.5 до 2.5 секунды, то усеченное среднее по-прежнему останется равным 0.5 – и это несмотря на то, что 80% пользователей столкнулись с ухудшением характеристик! Из-за неоднородной природы большинства сред правильно выбрать пороговое значение для отсечки почти невозможно. В реальной жизни бывали случаи, когда из-за такого усечения участки с наихудшей производительностью выдавали отчеты, которые представляли их чуть ли не лучшими во всей сети по эффективности! Соглашения SLA, основанные на процентном отношении транзакций, не подвержены таким эффектам и напрямую связаны с тем, что испы-

Подведем итог: соглашения SLA могут основываться на средних временных значениях, на процентном отношении временных средних или на процентном отношении транзакций. Соглашения SLA на базе временных средних значений дадут результаты со смещением; такие результаты могут не отражать реально существующее у пользователя положение дел. Соглашения SLA на основе процентного отношения транзакций технически более совершенны, однако пока не получили такого широкого распространения, как соглашения на базе временных средних значений.

Важные параметры

Другое важное решение касается точного описания целей. Сколько целевых значений мы получим для каждой переменной? Какие временные интервалы должны быть заданы для каждой переменной? Какие пороговые значения и процентные соотношения следует принять? Эти важные параметры должны основываться на потребностях пользователя, чтобы по ним можно было определить реальное положение дел.

Существует как минимум два интересующих нас пороговых значения: порог значимости и критический порог. Все значения меньше порога значимости для пользователя незаметны; это не значит, что они малы – просто величина задержек попадает в те пределы ожиданий пользователя, которые не вызывают у него раздражения. Задержки, превышающие критический порог, приводят к тому, что пользователь фактически лишается сервиса и не может работать должным образом. Задержки, по значению попадающие между двумя описанными пороговыми значениями, обычно воспри-

нимаются пользователями как инерционность, “подтормаживание” приложения и сети. Два этих естественных пороговых значения в численном виде обычно заранее неизвестны, однако их можно оценить экспериментально, с помощью сотрудничающих или невольно участвующих в оценке пользователей. Например, типичные значения порогов для загрузки веб-страниц часто составляют 3 и 8 секунд. Однако пороговые значения практически всегда сильно зависят от метода сетевого доступа и от типа самого приложения. Так, пользователи, имеющие доступ к развлекательному portalу через спутник, более спокойно относятся к задержкам, чем те, кто обращается с запросом в службу технической поддержки по выделенной линии. Для каждого приложения и группы доступа следует задать отдельное соглашение SLA.

Следует также учитывать, что пользователи чувствительны не только к абсолютным значениям задержек, но и к их колебаниям. Увеличение процентного соотношения эффективно контролирует такие колебания.

Например, пусть сначала соглашение SLA утверждает, что 95% времени работы отклик при транзакции должен быть менее 3 секунд, а 98% времени – менее 8 секунд. Цель состоит в том, чтобы за определенный период времени увеличить эти отношения до, скажем, 96% и 99% соответственно. Уменьшение 3-секундного порога мало повлияет на бизнес, поскольку это уже и так вполне приемлемое значение.

Итак, пороговые значения (обычно их два) в соглашениях SLA должны учитывать требования пользователей. Эти требования меняются в зависимости от приложений и от метода доступа к сети. Задержки, значения которых ниже нижнего порога, не влияют на пользователя, поскольку он их не замечает; задержки больше верхнего порога обходятся предприятиям очень дорого. Процентные отношения, если соглашение SLA их поддерживает, должны быть скорректированы с течением времени, чтобы работа системы постоянно улучшалась и совершенствовалась ■

S

W

E

N

Google основывает антиспамовскую коалицию

11

Компания Google объединилась с компанией Sun Microsystems в целях создания новой антиспамовской коалиции, пишет известный обозреватель www.eweek.com. Коалиция будет действовать на некоммерческих основах. Группа создает специальный веб-сайт StopBadware.org, призванный помочь пользователям персональных компьютеров бороться с такими спамовскими программами, как: программы показа рекламы на компьютерах; программы-шпионы, занимающиеся сбором персональной информации о компьютере и пользователе; программы, позволяющие хакеру “закрепиться” на взломанном компьютере и предотвратить свое обнаружение; а также препятствовать другим угрозам. “StopBadware.org – это обозреватель, нацеленный на помощь в борьбе с “badware”. Мы будем работать в сфере поиска и сбора надежной, объективной информации о загружаемых приложениях, чтобы помочь потребителю сделать правильный выбор того, что он будет устанавливать на своей машине. Мы готовы быть ведущим информационным центром”, – заявляют создатели проекта.

Часы, показывающие примерное время

12

В последнее время в мире продается все меньше часов – большинству из нас они просто не нужны, ведь мы можем узнать, который час, взглянув на экран мобильного телефона. Именно поэтому успех может быть обеспечен только тем часам, которые не похожи на другие. Компания Talus представила оригинальную концепцию часов, названных About Time. Это электронные часы с LCD-дисплеем, в центре которого вместо привычных часов и минут отображена лишь одна цифра, указывающая на ближайший час. Рядом помещаются текстовые надписи, соответствующие тому, как мы обычно определяем время, например, «около шести», «примерно четверть седьмого» или «почти полночь». Создатели необычных часов говорят о том, что таким образом им удалось «очеловечить» электронное устройство. Предполагается, что часы смогут «говорить» на нескольких языках, а также показывать особые сообщения на праздники, например, на Новый год. А если вдруг владельцу необычного устройства захочется узнать точное время, он сможет это сделать, нажав на специальную кнопку. В этом случае часы отобразят не только текущее время, но также день и год. Дата появления этого устройства в продаже пока неизвестна.

