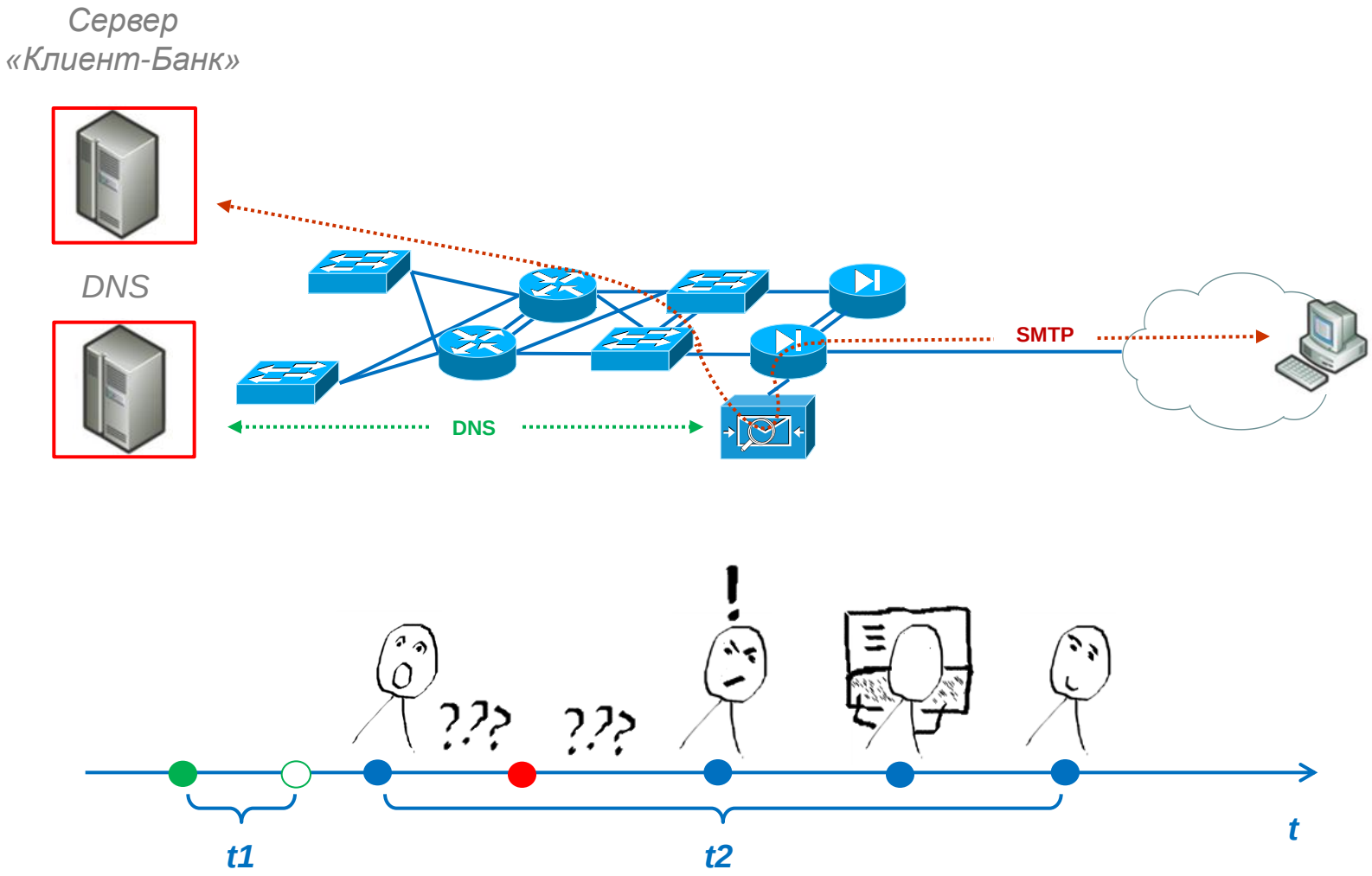
Как узнать о деградации сервисов раньше пользователя?

СП ООО «Солидекс ПИ»

Ровнов Павел

1 ноября 2013

Пример 1. Деградация сервиса «Клиент-Банк»



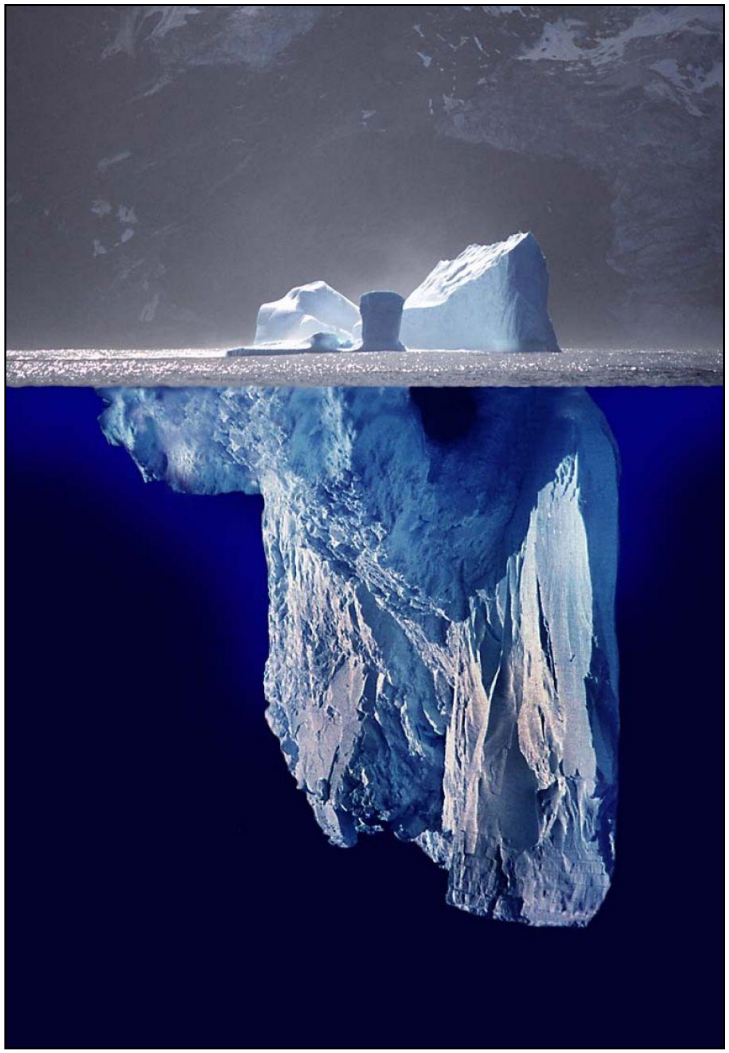
О чем пойдет речь?

Solidex

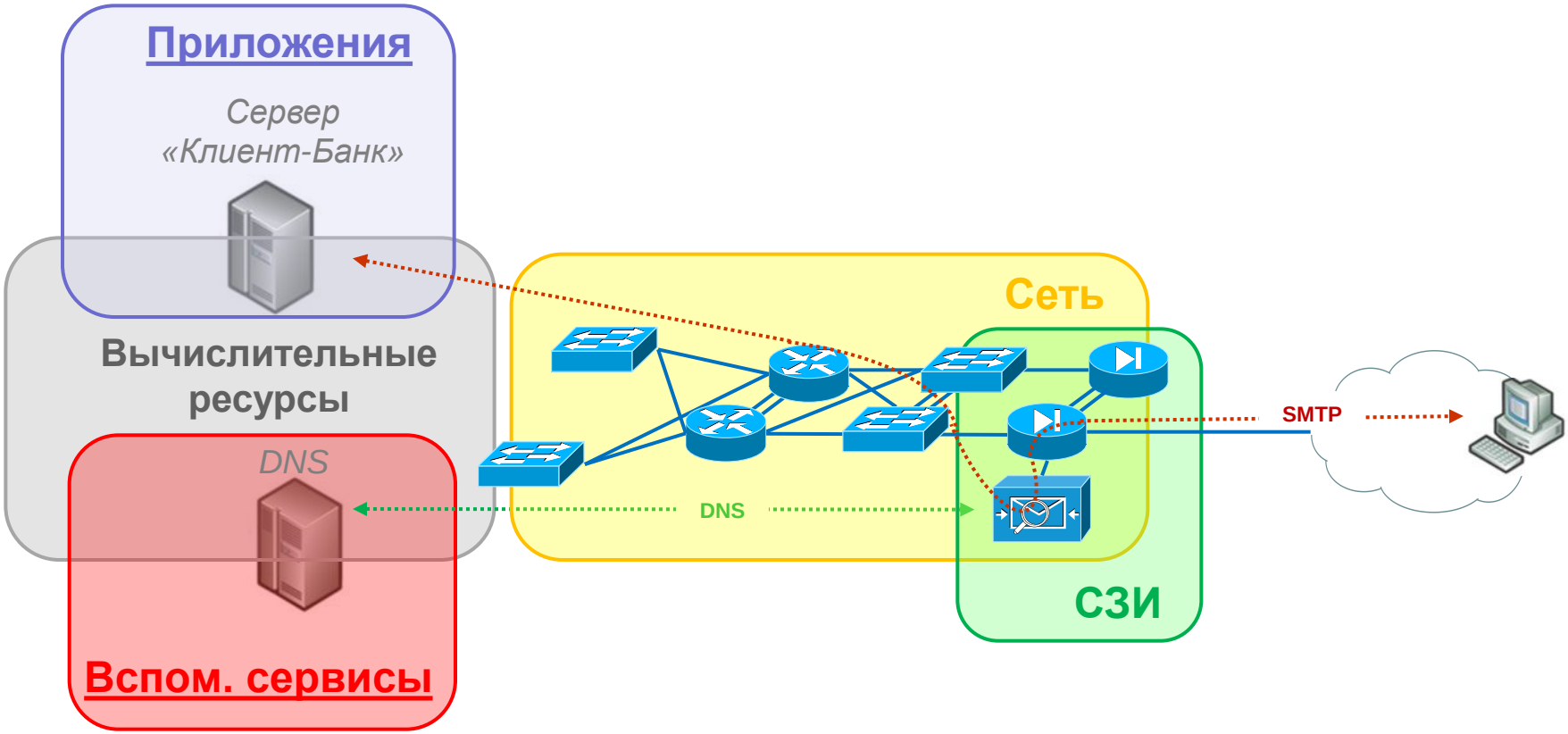
- Качество функционирования сервисов с точки зрения конечного пользователя
- Мониторинг: что, где и когда?
- Обработка событий и своевременное оповещение

Что замечает пользователь?

- **Доступность сервиса**
- ART, или время отклика



Доступность сервиса ~ доступность компонент



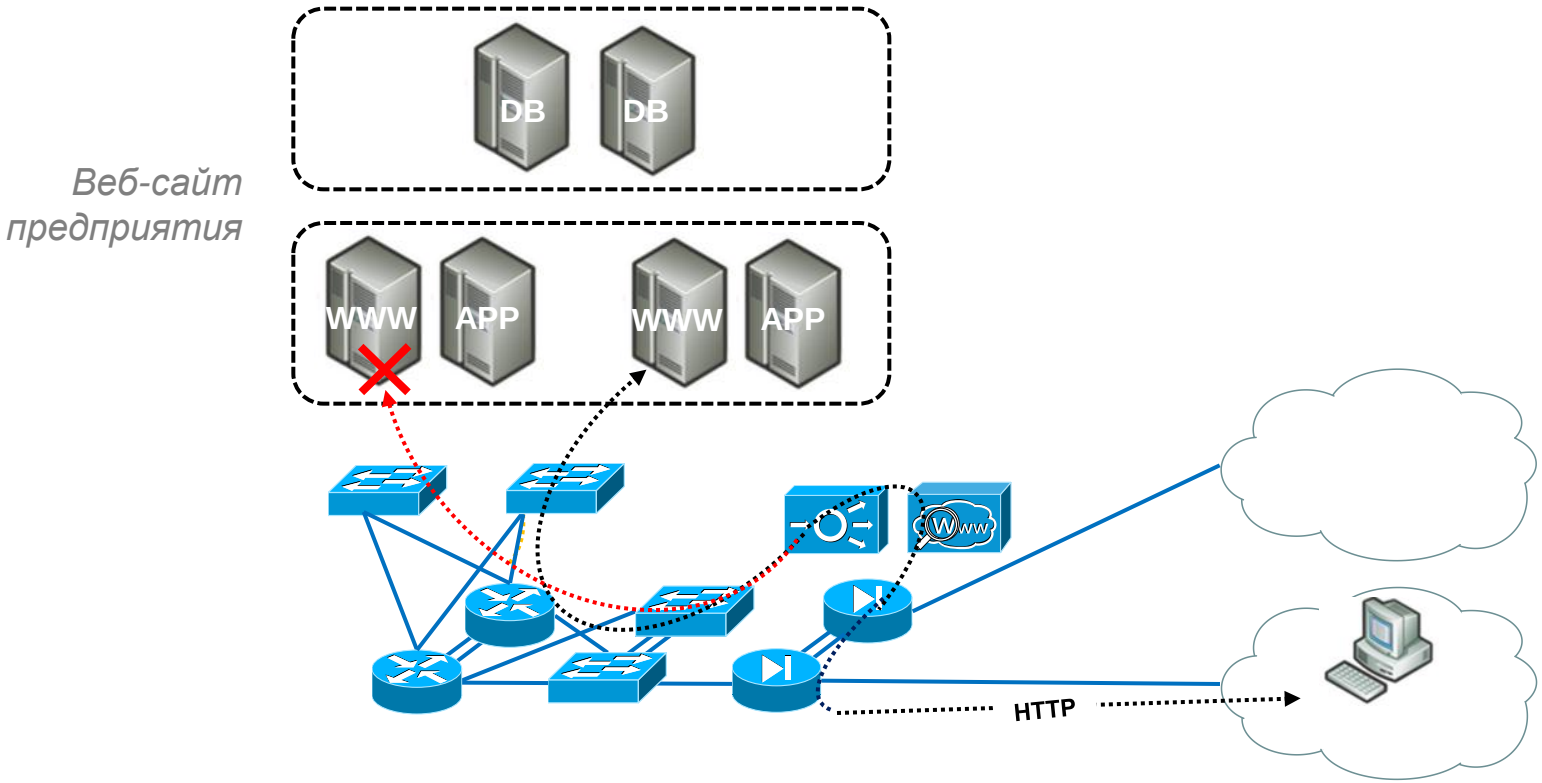
Основные причины инцидентов – программные ошибки и недоступность вспомогательных сервисов

Пример 2. Программный сбой

- 2 веб-сервера + балансировщик в составе системы «Интернет-банк»
- После ввода корректных учетных данных – возврат на логин-страницу
- Сообщений о неверном вводе учетных записей веб-приложением не зарегистрировано
- В чем причина?

Пример 2. Программный сбой (2)

- Не обеспечивается «HTTP session persistence»
- Пользователь аутентифицируется на 1-м сервере, хотя запросы поступают на оба



Пример 3. Недоступность вспомогательного сервиса

- Система «Клиент-Банк»
- Пользователи аутентифицируется при отправке сообщений по SMTP
- Недоступность LDAP-сервера с учетных записей пользователей – недоступность сервиса почты

Мониторинг доступности и информационная безопасность: вместе или врозь?

- По жизни – врозь:
  - Административное разделение
  - Раздельные средства мониторинга приложений и управления ИБ
- С точки зрения сервиса – вместе:
  - Главное – доступность сервиса, природа проблемы не столь важна
  - Мониторинг как инструмент ИБ

Пример 4. DoS-атака на веб-сайт

Solidex

- Регулятор: «какую DoS-атаку сможете выдержать?»
- DoS-атака оказалась успешной…
- Какой тип атаки, кто слабое звено?

Пример 4. DoS-атака на веб-сайт (2)

- Тип DoS-атаки – SYN-flood
- Модуль Anti-SYN-flood межсетевого экрана хранит состояния 4 М сессий
- Период хранения состояния TCP-сессии – 3 сек
- Переполнение памяти модуля происходит для >1.3 М SYN-пакетов/сек

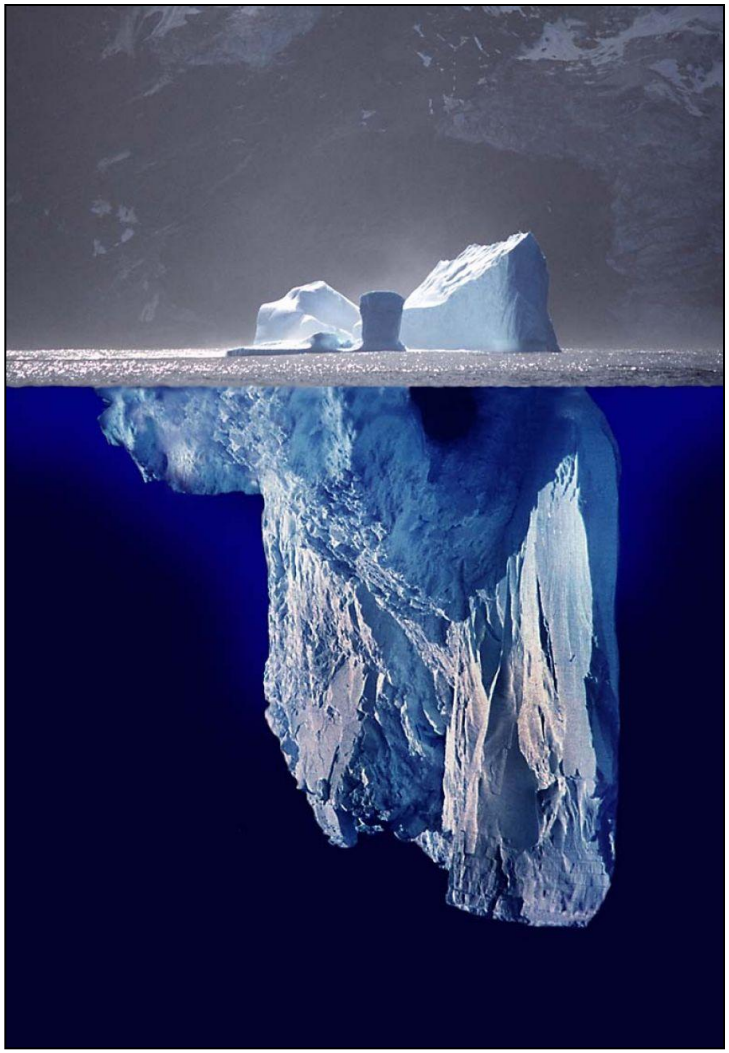
Доступность сервиса: что контролировать?

Solidex

- Состояние (up/down) компонентов сервиса и связей между ними
- Доступность вычислительных ресурсов для прогнозирования и обнаружения ситуаций их исчерпания

Что замечает пользователь?

- Доступность сервиса
- **ART, или время отклика**



Что значит «сервис тормозит»?

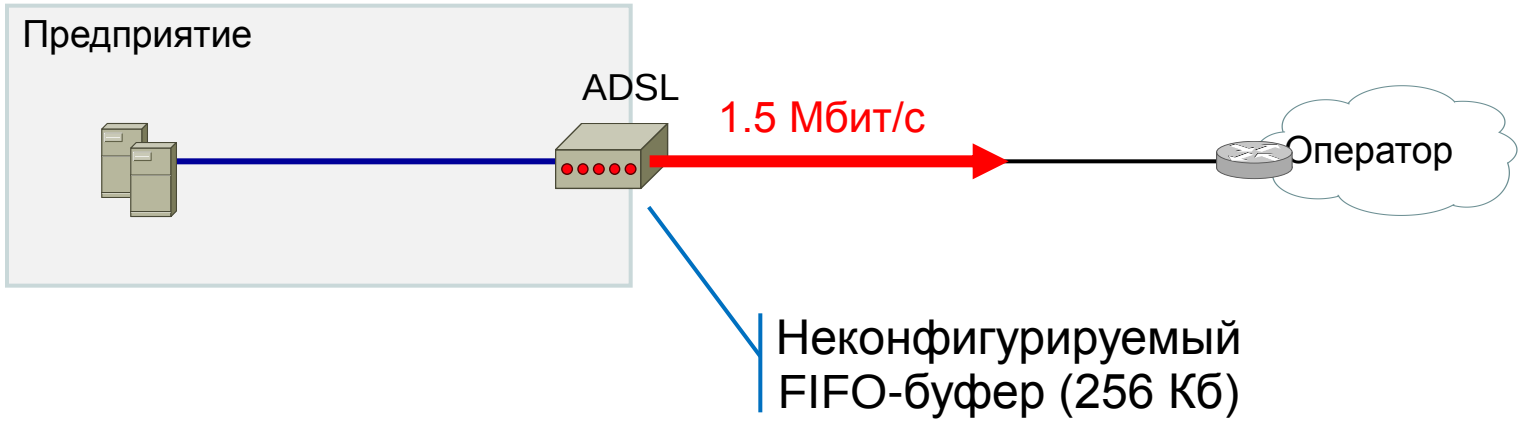
Solidex

- Время отклика приложения выходит за рамки комфортной работы
- Для интерактивных приложений порог ART – 100мс (Top-Down Network Design, CiscoPress, 2011)
- Два сценария: WAN и LAN

Источники повышения ART в WAN

- Конкуренция за WAN-канал и как результат буферизация и задержки
- RTT – Round Trip Time
- Ошибки в WAN-канале

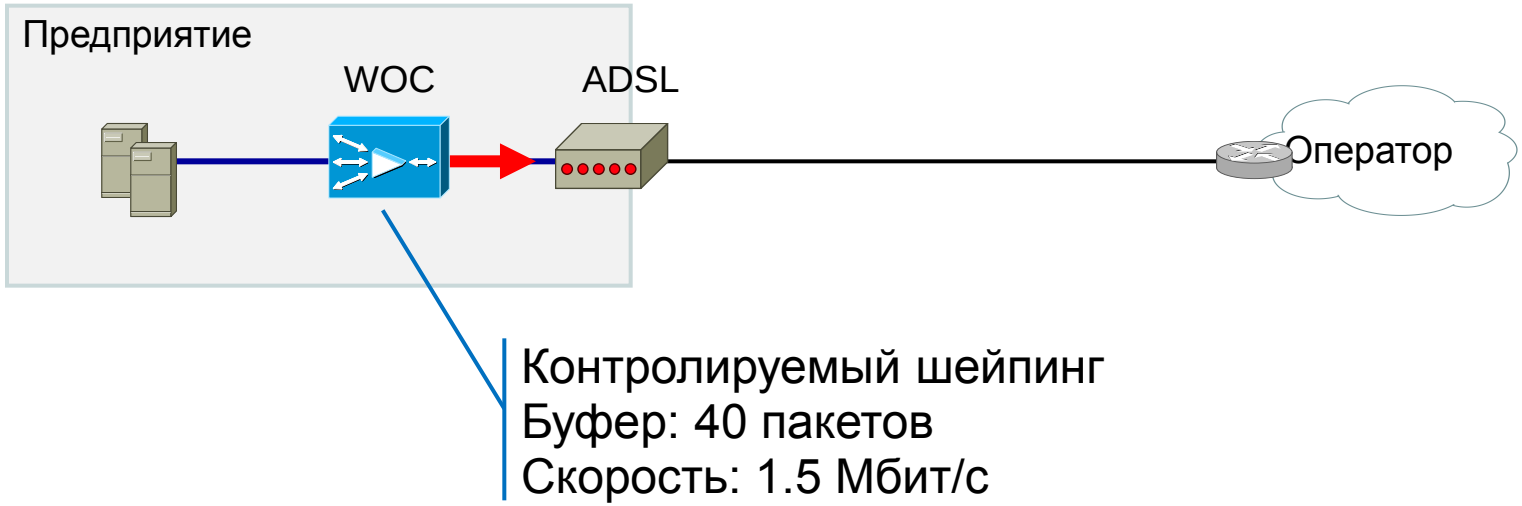
Пример 7. Излишняя буферизация на CPE



Network Access Link Properties

- Network latency measurements (?): Latency: 160ms Loss: 1.0%
- TCP connection setup latency (?): 160ms
- Network background health measurement (?): no transient outages
- Network bandwidth (?): Upload 1.6 Mbit/sec, Download 9.5 Mbit/sec
- Network buffer measurements (?): Uplink 1700 ms Downlink is good

Пример 7. Как ситуация была улучшена?



Network Access Link Properties

- Network latency measurements (?): Latency: 160ms Loss: 0.5%
- TCP connection setup latency (?): 160ms
- Network background health measurement (?): no transient outages
- Network bandwidth (?): Upload 1.5 Mbit/sec, Download 9.6 Mbit/sec
- Network buffer measurements (?): Uplink 57 ms Downlink 120 ms

ART в LAN

- Каналы связи «широкие»
- Запрос от пользователя до сервера «долетает» за миллисекунды
- Как правило, нет точек конкуренции за полосу пропускания (congestion point)

ART всецело зависит от задержек обработки запроса пользователя:

- клиентским приложением
- сервером

Причины задержек на стороне сервера

Solidex

- Недостаточные вычислительные ресурсы сервера
- Задержки при обращении к вспомогательным сервисам

Пример 6. Недостаточные вычислительные ресурсы

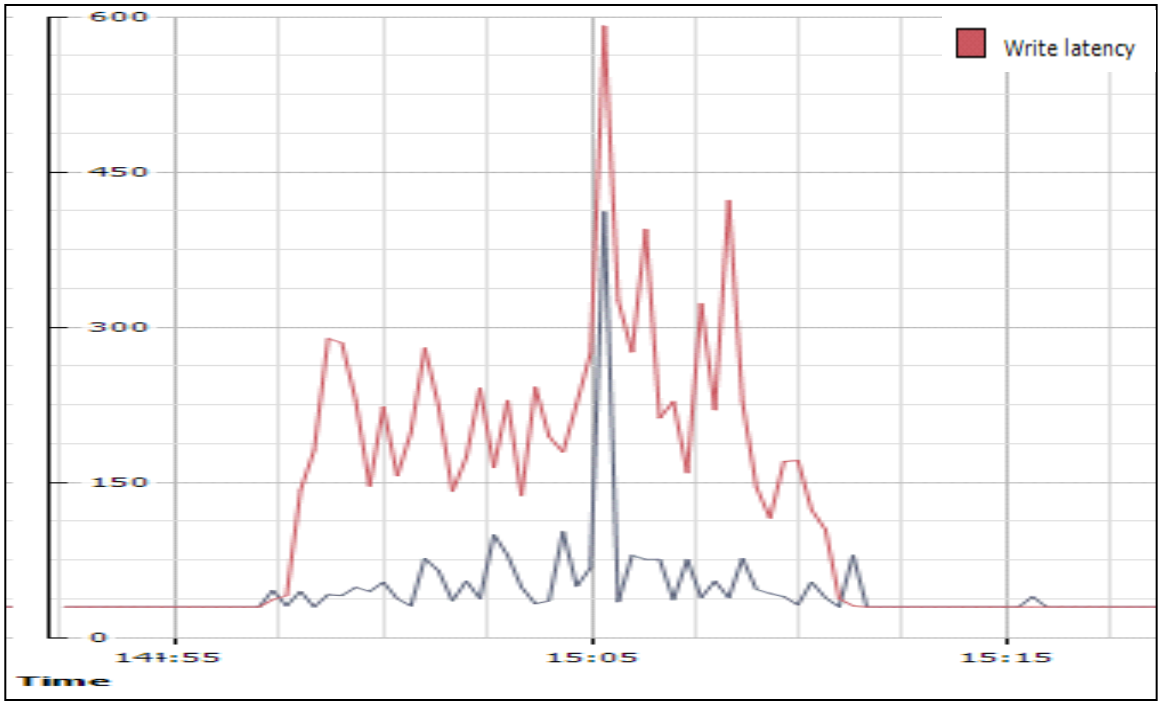
Solidex

- > 10 000 клиентов сервиса почты
- Рассылку писем с вложениями производят параллельно 10 MTA-агентов
- В работе отдельных MTA-агентов возникает задержка до 20 секунд
- В чем дело?

Пример 6. Недостаточные вычислительные ресурсы (2)

Solidex

Disk Latency > 500 ms и как результат заполнение буфера входящих сообщений – ситуация «TCP Window Full»



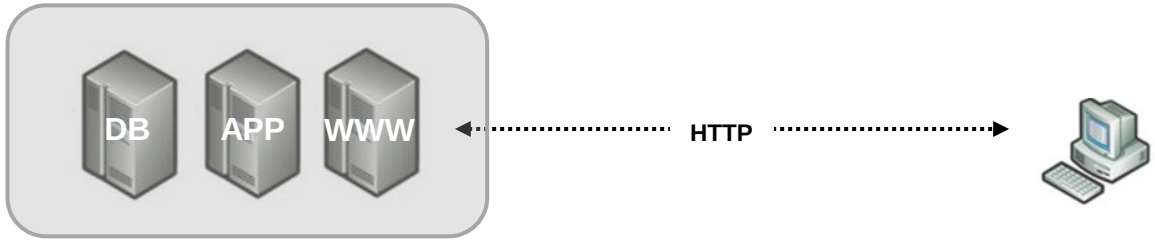
ART: что контролировать?

- В congestion-точках (на пограничных устройствах):
  - сбросы пакетов, очередь на интерфейсе в сторону WAN
- Для сетевых устройств:
  - % CPU, % Memory, % Network
- Для серверов приложений:
  - % CPU, % Memory, % Network, IOps, Disk Latency

Проблема контроля ART: время отклика end-to-end

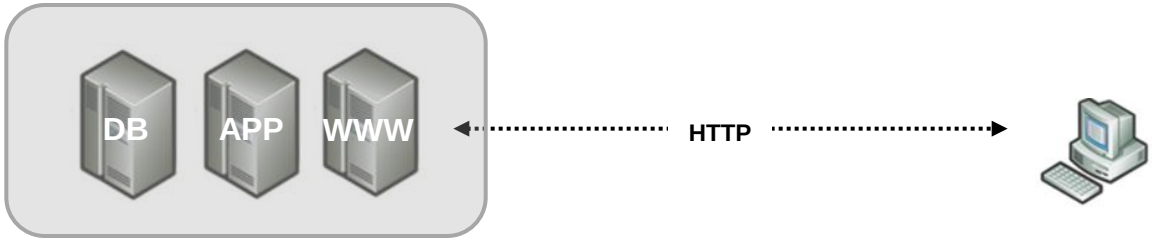
I. «Штиль». Очевидно, ART < 100 ms

CPU~1% MEM~20%
100 IOps NET~2%



II. ART - ? Возможно, есть проблемы

CPU~70% MEM~60%
100 IOps NET~2%



Измерение ART end-to-end

- Пробные/синтетические запросы
- Вычисление и *управление* ART средствами WOC – WAN Optimization Controller

Инструменты сбора информации о качестве сервиса

- Syslog
  - программная доступность
  - контроль связей между компонентами сервиса на программном уровне
- SNMP (polling)
  - параметры производительности сетевых устройств
- VMware API:
  - % CPU, % Memory, % Network, IOPs, Disk latency

Инструменты сбора информации о качестве сервиса (2)

Solidex

- WMI/RPC
  - %CPU, %MEM для отдельных процессов
- Пробные запросы/синтетический мониторинг
  - время отклика вспомогательных сервисов
  - для сервисов вне нашего контроля

Инструменты сбора информации о качестве сервиса (3)

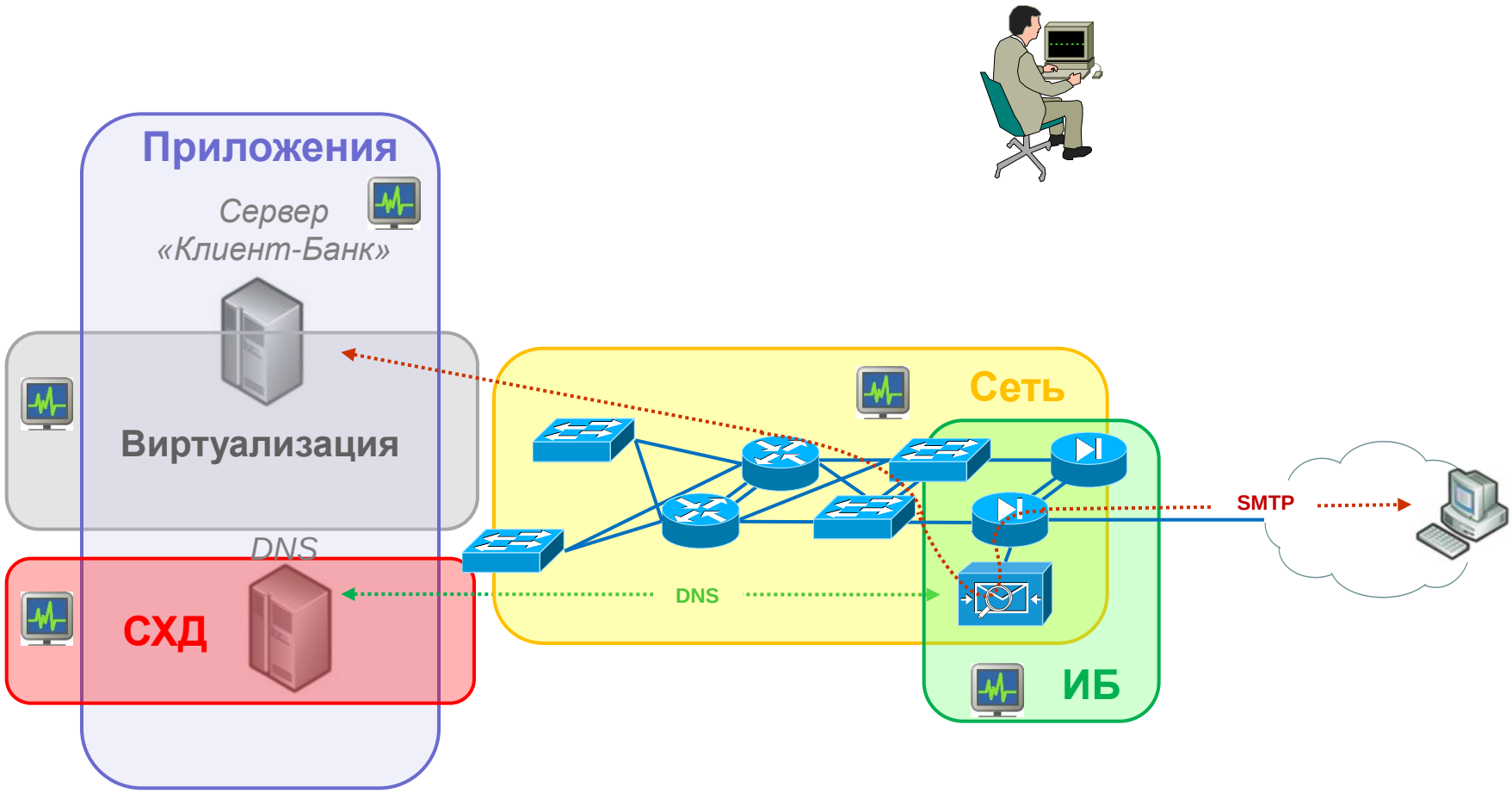
- Специализированные сенсоры:
  - WOC – Wan Optimization
  - App Ctrl / IPFIX / sFlow – информация о приложениях

Обработка данных

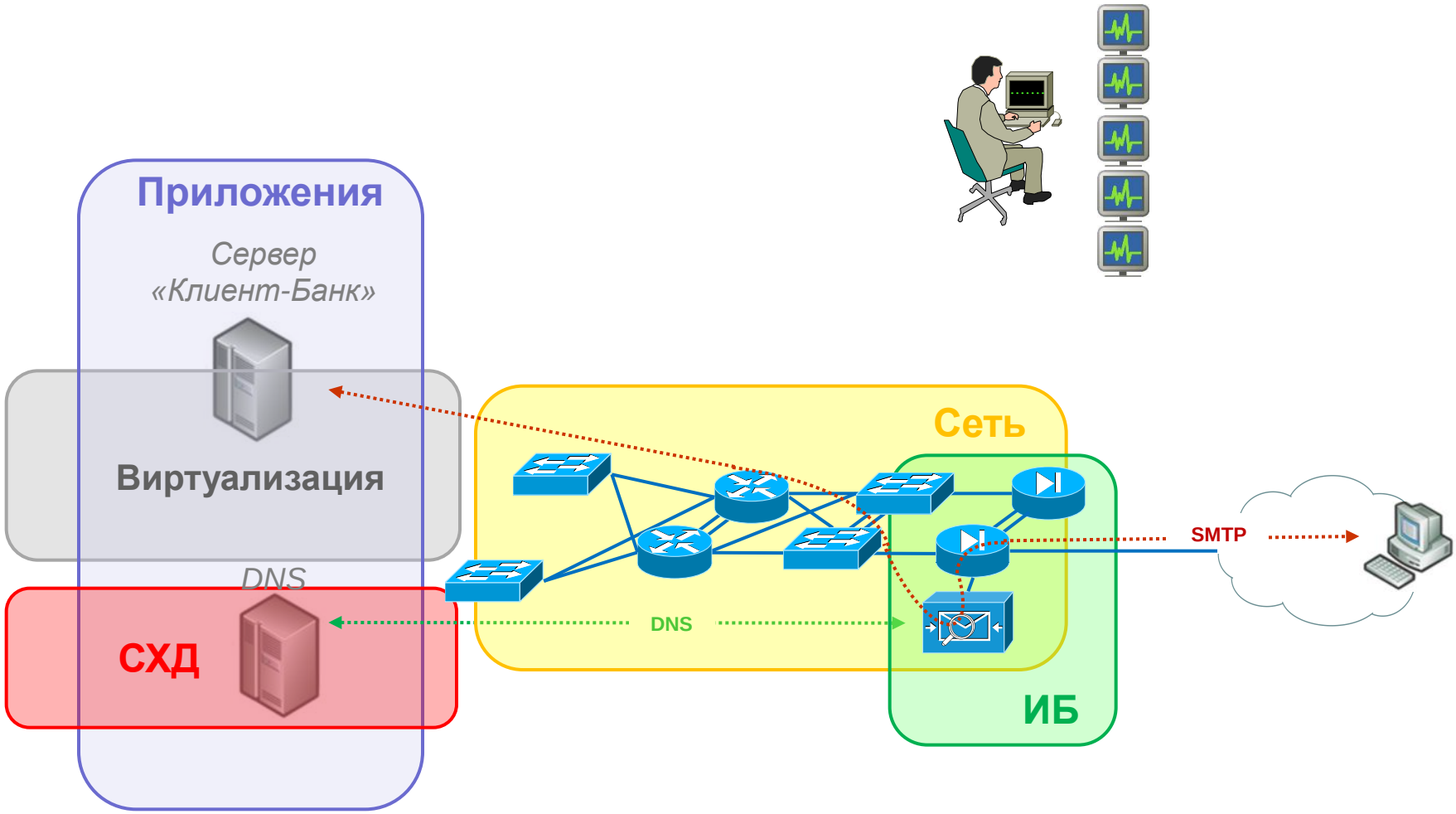
Solidex

- Метрики сервиса определены и собраны
- Где находятся собранные данные?

Административные домены vs. Сервис



Административные домены vs. Сервис (2)



Данные собраны в единой точке, что дальше?

- Фактически – 2 ситуации:
  - Сбор ради сбора (например, в угоду PCI DSS)
  - Аналитика «из коробки» и наборы правил от производителя для борьбы со всеми «болезнями»

Необходимость в собственных правилах

- Без правил (сбор ради сбора):
  - Можно лишь провести расследование постфактум
- Аналитика «из коробки»:
  - Принцип «черного ящика» – как правила работают и что делают?
  - Не знает о критических точках Вашего сервиса

Пример 8. Обнаружение аномальной активности на рабочих станциях

- Задача:
  - Обнаружение следов успешных атак вредоносного программного обеспечения
- Правило:
  - Обнаружить C&C-каналы (например, HTTP Post-запросы к сайтам в списке Botnet-серверов)
  - Оповестить администратора через сообщение консоли инцидентов

И спустя время аномальная активность обнаружена…

Проблема обработки событий: человеческий фактор

- Отсутствие процедур по обработке событий
- Несвоевременное обнаружение деградации сервиса в результате уклонения от процедур

Пример 9. Несвоевременное обнаружение сбоя

Solidex

- Широковещательный шторм в ЦОД
- Сервисы ЦОД не доступны
- В чем причина?

Пример 9. Несвоевременное обнаружение сбоя (2)

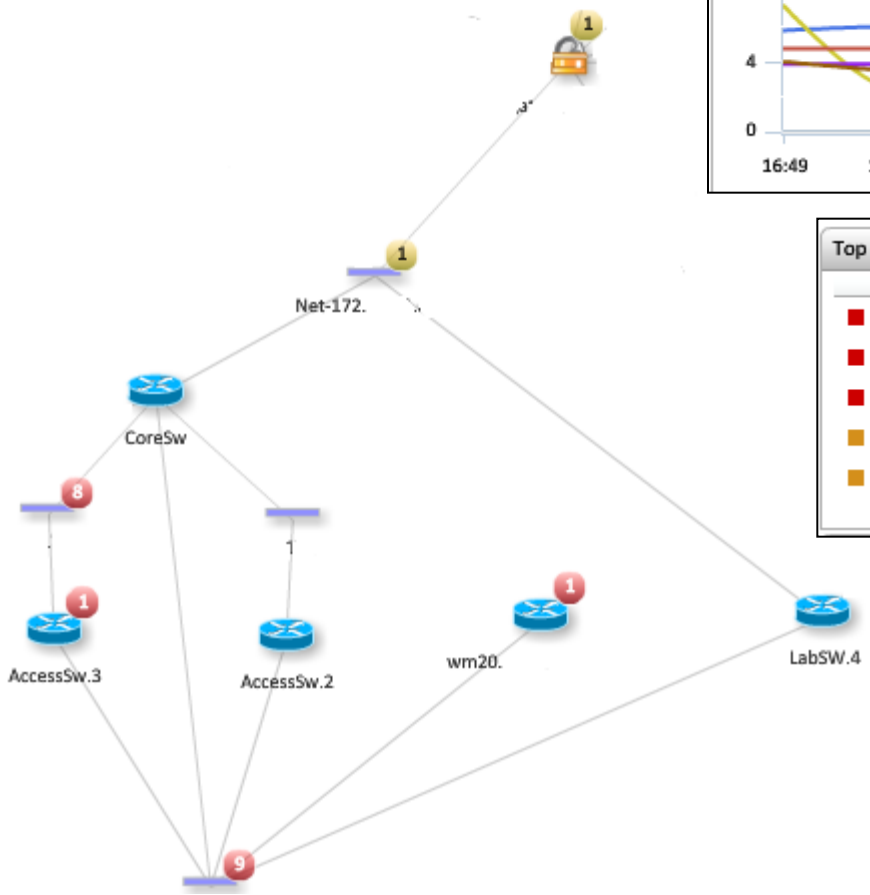
Solidex

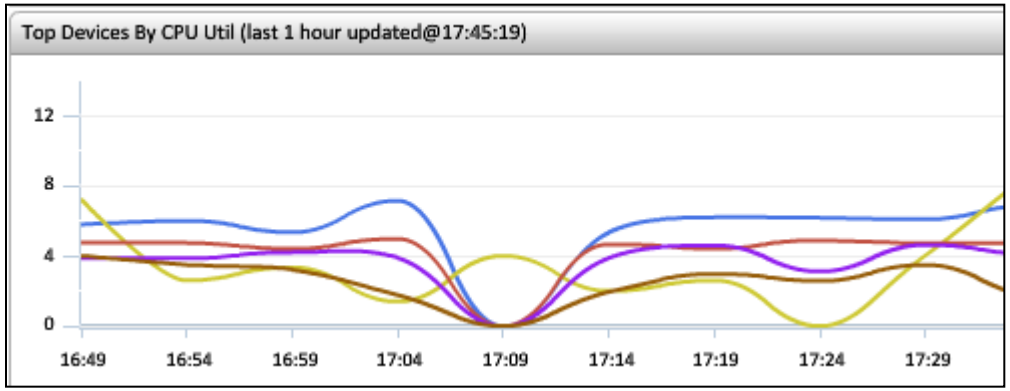
- Сообщения о программной ошибке на коммутатора ядра на протяжении 2-х месяцев сбрасывались на Syslog-сервер
- Программная ошибка вылилась в сбой процесса STP и широковещательный шторм
- Сервисы ЦОД были недоступны клиентам Заказчика ~24 минуты
- Время восстановления отдельных сервисов – 1 день

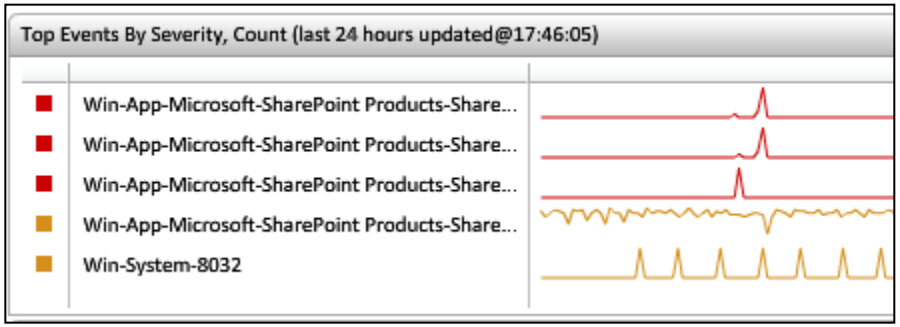
Своевременное оповещение

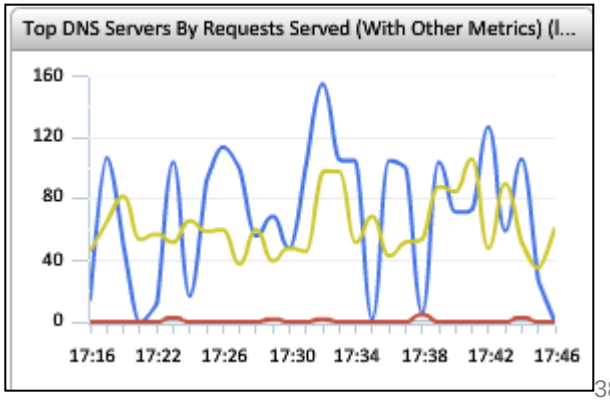
- Для наиболее критичных инцидентов – оповещение по почте или SMS
- Размещение инцидентов на dashboard в соответствии с приоритетом
- Непрерывный тюнинг правил для исключения ложных и избыточных инцидентов

Примеры решений









Пример 10. Решение проблемы «зависания» голосового шлюза

- FXO-линии голосового шлюза не всегда освобождались при завершении разговора
- Спустя время – все 4 линии заняты

Пример 10. Решение проблемы «зависания» голосового шлюза (2)

- Мониторинг
  - состояния FXO-линии – по SNMP с интервалом 1 минута
  - программных ошибок ОС шлюза – по Syslog
- 2 правила для определения момента, когда линии «подвисли» и оповещения администратора:
  - «3 линии заняты более 3-х минут»
  - «длительный разговор – более 20 минут»
- В интервале в 10 минут до «зависания» линий обнаружены сообщения о программной ошибке. Ошибка устранена обновлением версии ОС шлюза

Пример 11. Сервис совместной работы (SharePoint)

- «Медленную» работу сервиса совместной работы (открытие страниц веб-приложения ~10 секунд)
- Недоступность сервиса совместной работы – «500 Internal Server Error»

Пример 11. Сервис совместной работы (SharePoint) (2)

- Мониторинг
  - ART и состояния сервиса (по коду HTTP-ответа) с помощью синтетических HTTP-запросов с интервалом 1 минута
  - потребляемых процессами SharePoint системных ресурсов – по WMI с интервалом 3 минуты
  - программных ошибок SharePoint
- Правила для определения момента деградации качества сервиса с оповещением администратора:
  - «(ART > 1 сек) AND (HTTP.Response.Code = 200)» или «HTTP.Response.Code = 500»

Пример 11. Сервис совместной работы (SharePoint) (3)

- Событию деградации сервиса предшествует утечка памяти через процесс SharePoint OWSTimer.exe
- Утечка памяти является результатом программной ошибки и исправляется патчем

Выводы

- Для своевременного обнаружения деградации сервиса:
  - контроль доступности и ART
  - разработка правил
  - оповещение

solidex