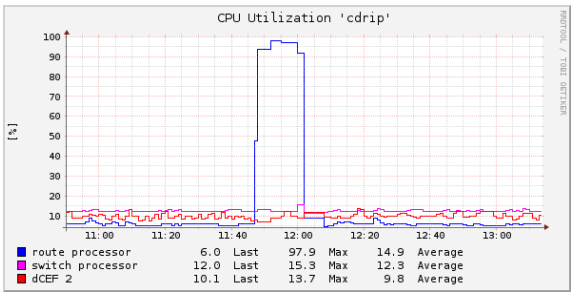
ulm university universität ulm

Bachelorarbeit
mad-traffictool

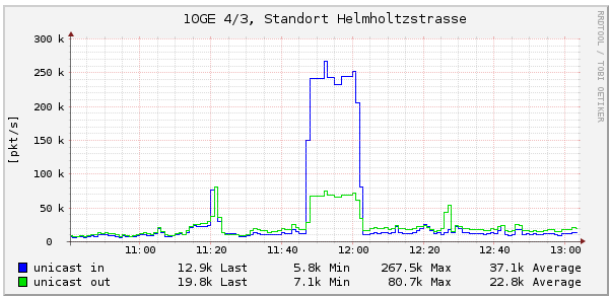
Benjamin Droemer, Medieninformatikstudent
Betreut durch: Karl Gaissmaier, kiz Abt. Infrastruktur

Effiziente Verkehrs- und Anomalieanalyse in großen Netzen

Netzinfrastrukturen werden schnell komplex und heterogen und damit steigt schnell der Aufwand für Überwachung und Wartung. Ein weiteres Problem sind die sehr großen Datenmengen in unterschiedlichen Formaten, die von den jeweiligen Geräten zur Verfügung gestellt werden. Als Beispiel kann das Netz der Universität Ulm mit mehr als 40.000 Switchports an über 700 Switchen verschiedener Herstellern dienen. Darüber hinaus sind weitere Infrastrukturelemente wie Verteiler, Stromversorgung und interne und externe Anbindungen zu berücksichtigen. Zur Lösung dieser Aufgabe werden derzeit am kiz für das Netz der Universität Ulm eine Reihe von skriptbasierten Werkzeugen eingesetzt.



Der Router ist überlastet



Scheinbare Quelle der Last

Das obige Beispiel zeigt ein Beispiel bei dem ein Router der Uni Ulm durch Überlast ausgefallen ist. Auf den Bildern ist die Auslastung des Routers und die eines Netzwerkverteilers,zu sehen nach welchem die Spur zum Verursacher endet. Mit dem in dieser Arbeit zu entwickelnden Werkzeug soll eine solche "root cause Analyse"möglichst automatisiert ermöglicht werden.

Aufgabenstellung

Aufgabe dieser Bachelorarbeit ist der Entwurf und die Realisierung eines Werkzeugs zur Sammlung und Aufbereitung des relevanten Datenaufkommen eines Switchports als Grundlage für die Erkennung von Problemen und Anomalien im Verhalten einzelner Komponenten.

Vorgeschlagene Arbeitsschritte

- Übersicht bestehender Werkzeuge und Protokolle zur Sammlung und Analyse von Datenverkehr in komplexen Netzen

- Analyse der Eignung von Auswertungssystemen wie Chukwa (https://chukwa.apache.org), Zipkin (https://github.com/twitter/zipkin) und der darunter liegenden Datensammelwerkzeuge für den Anwendungsbereich Netzmonitoring
- Entwurf eines Lösungskonzeptes
- Umsetzung und Validierung insbesondere mit den Zielen
  - Integrierbarkeit mit den bestehenden Lösungen zum Netzmanagement am kiz
  - Skalierbarkeit und Robustheit der vorgeschlagenen Lösung

Herausforderungen und Kontext

Die reine Anzahl an Switchports stellt bereits eine große Herausforderung dar. Performante Algorithmen müssen entwickelt oder komplexe Datenanalysewerkzeuge genutzt werden, um alle benötigten Prozesse in angemessener Zeit abarbeiten zu können. Alle an der Uni Ulm eingesetzten Switche erlauben es, über das Netzwerkmanagementprotokoll SNMP abgefragt zu werden weisen jedoch teilweise herstellerspezifischen Erweiterungen auf die geeignet abstrahiert werden müsse. Diese Schnittstelle ermöglicht es, sämtliche für dieses Vorhaben benötigten Daten (Menge der in beide Richtungen gesendeten Unicast-, Broadcast-, Multicastpakete sowie der Fehlermeldungen) abzufragen es muss aber ein geeignetes Speicherkonzept gefunden werden.
Das Werkzeug soll sich in die bereits vorhandenen Softwaretools des Netzwerkmanagements integrieren. Dabei soll es vorhandene Daten über die Netzwerkinfrastruktur (Wie z.B. über die Switche) aus den beim kiz bereits vorhandenen Repositories beziehen. Dadurch ist sichergestellt, dass zum Netzwerk neu hinzugefügte Switche mit einbezogen, entfernte Switche jedoch nicht mehr abgefragt werden.
Die aktuelle Datentransferrate jedes Switchports soll wenn möglich im Minutentakt abgefragt und so gespeichert werden dass die Daten auch für offline Analysen zur Verfügung stehen (z.B. in einer geeigneten Datenbank). Zur Minimierung der vorzuhaltenden Datenmenge sollen alle Daten, nach Ablauf einer jeden Zeitperiode, als Durchschnittswert der vergangenen Stunden, Tage, Wochen, Monate und Jahre abgespeichert werden. Dadurch können einzelne Werte miteinander vergleichen werden, zum Beispiel die aktuelle Datenmenge gegenüber der von vor einem Jahr. Dazu ist ein geeigneter Kompromiss zwischen Speicherbedarf und Detailtiefe zu finden. Ein Beispiel für diese Problematik ist die Speicherung Roh-Daten über einen geeignet langen Zeitraum zu speichern, um Anomalien, wie Abweichungen zum Normalfall, zum Beispiel bezogen auf den Wochentag, zu erkennen. Dabei soll das System offen realisiert sein damit es Netzadministratoren einfach möglich ist, Templates für neue Abfragen zu erstellen.

Weiterführende Themen

In einem ersten Ansatz wird das Verhalten an einem Switchport nur im zeitlichen Vergleich mit sich selbst bewertet. Weiterführende Konzepte können Ports in Gruppen Clustern (z.B. CIP Pool System, Server, ...) und Abweichungen und Auffälligkeiten innerhalb einer Referenzgruppe zu identifizieren. Neben der Überwachung und Erkennung von Fehlern sind auch die automatisierte Anwendung von Maßnahmen denkbar wie das automatische isolieren oder abschalten von ports im Fehlerfall.