Olaf Zäncker
VoIP Consulting & Services

White Paper

Version 1.1, deutsch

von

Olaf C. Zäncker

29. Juli 2018

Illuminating the Background

of

VoIP Readiness

VoIP ist eine sehr sensible Verkehrsart. Sie stellt -verglichen mit 'normalen' Datenanwendungen- ganz neue, strengere Anforderungen an IP-Infrastrukturen. Daher sollte vor Einführung von VoIP-Systemen immer eine realitätsnahe 24 / 7 VoIP-Readiness Messung stehen, die bei Nicht-Eignung zuverlässige Rückschlüsse auf die Störungsursachen (root cause diagnosis) zulässt. MOS, PESQ, sowie Min., Max., Mittelwerte & Co. werden diesem Anspruch natürlich nicht mehr gerecht.

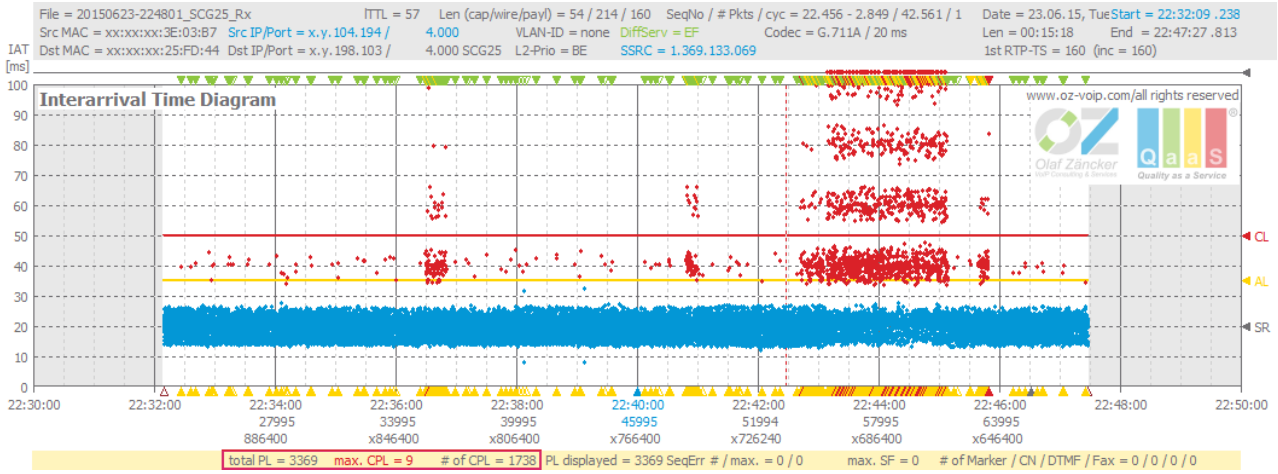
Alle Fakten und Hintergründe (für eilige Leser + ausführlich) **in diesem White Paper**.

Copyright Olaf C. Zäncker 2009 - 2020

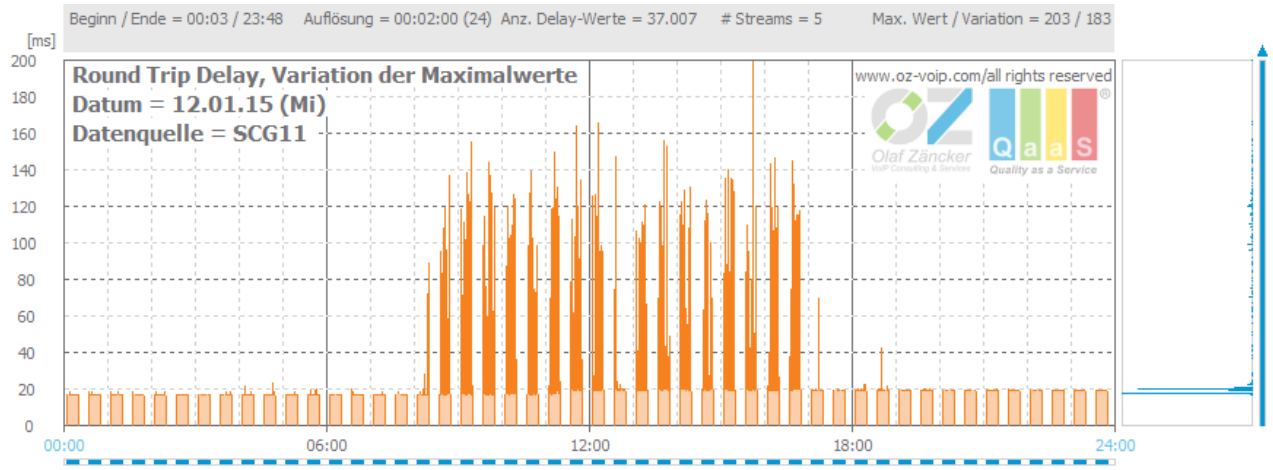
1 Für den eiligen Leser - alle Fakten auf nur einer Seite !

VoIP-Readiness Messungen haben das Ziel, die grundsätzliche Eignung einer IP-Infrastruktur für Echtzeitdienste (VoIP) nachzuweisen. Bereits der Planung (eine dem künftigen VoIP-System angemessene Anzahl und Verteilung von Callgeneratoren) und der Durchführung (repräsentativer Mess-Zeitraum, Erfassung geeigneter Parameter) ist größte Sorgfalt gefragt.

Die meisten Fehler werden jedoch bei der Auswertung gemacht. Nur wenn alle für VoIP relevanten Qualitätsparameter separat und entsprechend ihres individuellen Informationsgehaltes im zeitlichen Verlauf ausgewertet werden (Analyse), ist der tatsächliche Charakter von Störungen erkennbar – eine unabdingbare Voraussetzung, um zuverlässige Rückschlüsse auf deren tatsächliche Ursache(n) ziehen zu können (Diagnose).



Nachvollziehbare, aussagekräftige Messergebnisse müssen zudem sowohl Details (z.B. Jitter, Packet Loss – siehe obige Grafik), als auch Übersichten (z.B. Tagesprofil des Round Trip Delays; siehe unten) beinhalten. Min.-, Max.-, Mittel-, Prozentwerte, Standardabweichung sind somit tabu.



Betrachtet man Details einer state-of-the-art VoIP-Readiness Messung wird sofort klar, warum die sog. MOS-Werte (Mean Opinion Score) in der modernen IT-Landschaft keinen Platz mehr haben !

Um zusätzlich zur IP-Infrastruktur auch das nach der aktiven VoIP-Readiness Messung installierte VoIP-System zu prüfen, empfiehlt sich als Ergänzung nach dessen Inbetriebnahme eine passive VoIP-Abnahmemessung auf Basis der tatsächlichen VoIP-Calls der Benutzer.

>>> *Die genaue Herleitung findet sich auf den nachfolgenden Seiten des White Papers* >>>

2 Bevor es losgehen kann

2.1 Motivation für VoIP-Readiness Messungen

Die Motivation für die Durchführung einer VoIP-Readiness Messung ist das zweifelsfreie messtechnische Feststellen der grundsätzlichen Eignung einer IP-Infrastruktur (LAN, WLAN, WAN) für den dauerhaften störungsfreien Betrieb einer isochronen Echtzeitapplikation (VoIP, aber auch Fax over IP, Video over IP). Dabei wird für einen repräsentativen Zeitraum (24 / 7) künstlicher VoIP-Verkehr in dem zu untersuchenden IP-Netzwerk erzeugt, gemessen und ausgewertet um insbesondere die Eignung (Performance) der aktiven Netzwerk-Komponenten (Switches, Router, Firewalls, etc.) für den permanenten Betrieb von VoIP vor dem Hintergrund der Koexistenz mit / der Auslastung durch Datenverkehr anderer, i.d.R. schon bestehender Applikationen beurteilen zu können.

2.1.1 Sonderfall VoWLAN

In der Praxis wird i.d.R. fälschlicherweise davon ausgegangen, dass eine für das WLAN vorgenommene funktechnische 'Ausleuchtung', in deren Ergebnis eine sog. 'Heatmap' entsteht, das Funktionieren aller Anwendungen -einschließlich VoIP- sicherstellt. Das kann jedoch -streng genommen- nicht einmal für 'normale' Datenanwendungen, wo Störungen oft nur schwer / überhaupt nicht bemerkt werden, geschweige denn für Echtzeitkommunikation wie z.B. VoIP gesagt werden, da sich die Ausleuchtung technisch bedingt zwangsläufig nur auf grundlegende physikalische Größen wie Feldstärke, Signallaufzeiten, etc. beziehen kann.

Wenn also vorgesehen ist, VoIP auch im WLAN zu betreiben, empfiehlt es sich aus diesem Grund durchaus, zusätzlich zur Ausleuchtung mit Hilfe von WLAN-fähigen VoIP-Callgeneratoren eine WLAN-Qualitätsmessung auf VoIP-Anwendungsebene durchzuführen.

2.2 Worum es *nicht* geht

Wie im vorhergehenden Abschnitt erläutert, fokussiert eine VoIP-Readiness Messung auf die aktiven Komponenten eines IP-Netzwerkes im Kontext der Auslastung durch andere Anwendungen. Die Verkabelung selbst ist (sofern heutzutage überhaupt noch erforderlich) im Zuge einer solchen Messung nicht mit abgedeckt – sie wird üblicherweise unter Verwendung von Kabeltestern z.B. in Bezug auf ihren Frequenzgang überprüft. Somit ist es auch überflüssig, Callgeneratoren an einem Arbeitsplatz-Anschluss (sog. Wall-Outlet) anzuschließen – zumal es außerdem so durch unberechtigte, nicht autorisierte Personen zu einem (un)beabsichtigten Unterbrechen der Messung kommen kann. Statt dessen sollten die Callgeneratoren direkt mit den betreffenden Switch-Ports verbunden werden.

2.3 Planung und Vorbereitung

Auch bei VoIP-Readiness Messungen ist die sorgfältige Planung und Vorbereitung wesentliche Voraussetzung sowohl für den reibungslosen Ablauf der Messung selbst, sowie für zuverlässige und aussagekräftige Ergebnisse.

So sollte sich die zu untersuchende VoIP-Infrastruktur zum Zeitpunkt der VoIP-Readiness Messung idealerweise bereits in dem Konfigurationszustand befinden, welcher für den späteren VoIP-Betrieb vorgesehen ist. Das betrifft -sofern geplant- VLANs (einschließlich der damit verbundenen IEEE 802.1Q Priorisierung auf Layer 2) und / oder die DiffServ-Priorisierung (Layer 3; siehe auch nachfolgender Unterabschnitt).

Weiterhin sollte insbesondere bei Standort-übergreifenden Messungen Sorge getragen werden, dass die entsprechenden Ports (z.B. H.323 = 1720, SIP = 5060, RTP / RTCP = Portbereich siehe Herstellerangaben) in den Firewalls freigeschaltet sind.

2.3.1 Vorgaben der VoIP-Hersteller

Idealerweise steht zum Zeitpunkt der VoIP-Readiness Messung bereits fest, welches VoIP-System eingesetzt werden soll, da die Hersteller i.d.R. konkrete Vorgaben für die Parametrisierung und die Toleranzgrenzen der Übertragungsgüte von IP-Infrastrukturen (VoIP-Qualität) machen. Das beinhaltet u.a. folgende Informationen (Aufzählung erhebt keinen Anspruch auf Vollständigkeit):

- Parametrisierung (einige Punkte sind nicht bei allen VoIP-Systemen relevant) :
  - maximale Auslastung des IP-Netzwerkes in % und Anteil von Broadcasts
  - Verwendung von VLANs einschließlich Priorisierung gemäß IEEE 802.1Q (5 = Video oder 6 = Voice); ggf. Application Sharing mit einer darunter liegenden Prio (< 5)
  - Verwendung von DiffServ, d.h. Layer 3-Priorisierung für RTP (Expedited Forwarding = '*EF*') und SIP (Assured Forwarding xx = '*AFxx*')
  - für RTP / RTCP verwendete UDP Port-Bereiche (zur Freischaltung in Firewalls), sowie SIP-Ports (UDP oder TCP), ggf. zusätzlich zu / abweichend von Port 5060
  - Mindest-Bandbreite der geplanten Codecs (insbesondere für WAN-Verbindungen)
  - PoE-Fähigkeit der Switches
  - Konfigurations- bzw. Leistungsangaben für PCs, sofern Soft-Clients vorgesehen sind
- Toleranzgrenzen (Schwellwerte) der Übertragungsgüte (VoIP-Qualität) – idealerweise mit separaten Werten für LAN und WAN:
  - Delay
  - Jitter
  - Packet Loss.

MOS-Werte (Mean Opinion Score), egal auf welchem Wege diese auch ermittelt werden, haben hier natürlich nichts mehr zu suchen, da sie -technisch bedingt- im Falle von Störungen in der IP-Infrastruktur keinerlei konkrete Rückschlüsse auf deren tatsächliche Ursachen ermöglichen.

Die Einhaltung der Hersteller-seitigen Vorgaben vereinfacht besonders im Falle von Störungen der VoIP-Qualität erfahrungsgemäß für den Endkunden die Diskussion(en) mit dem Hersteller bzw. Service-Provider und ist somit dringend anzuraten.

Wurde bezüglich des VoIP-Systems noch keine Auswahl getroffen, können / müssen statt dessen natürlich allgemeingültige Schwellwerte herangezogen werden, die sich an marktüblichen Werten orientieren.

2.4 Mess-Szenario

2.4.1 Anzahl gleichzeitiger VoIP-Testcalls

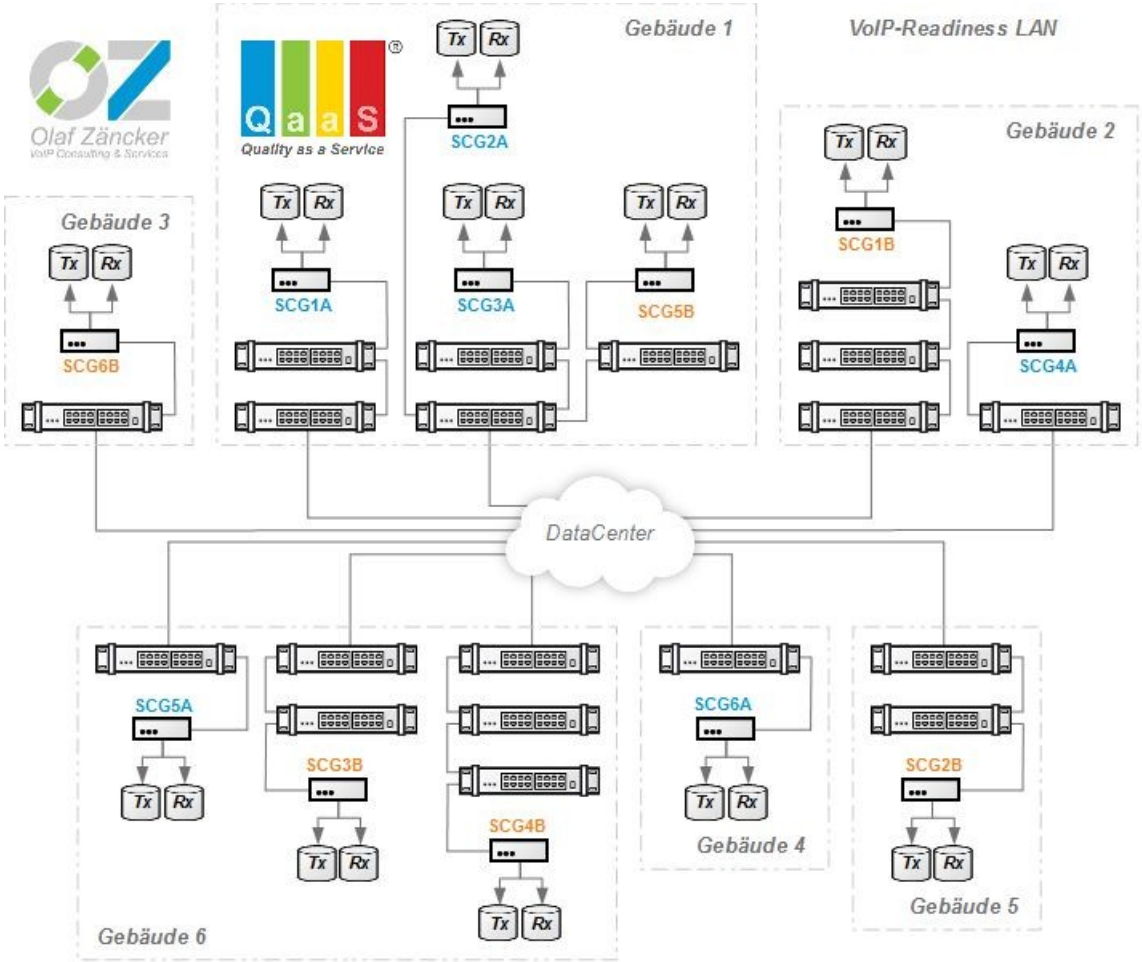
Sofern keine speziellen Anwendungen, wie z.B. VoIP-basierte CallCenter vorgesehen sind, hat sich in normalen Büro- und Produktionsumgebungen bei der Festlegung der Anzahl gleichzeitiger VoIP-Testcalls (aggregierte VoIP-Last im Core-Bereich) ein Wert zwischen 7 .. 10 % (bezogen auf die Anzahl der für den entsprechenden Standort geplanten IP-Phones bzw. VoIP-Clients) als

zuverlässig und ausreichend erweisen.

Im Falle von VoIP-CallCentern kann es hingegen notwendig sein, bis zu 100 % VoIP-Verkehr (bezogen auf die Anzahl der CallCenter-Agenten) zu simulieren.

2.4.2 Anzahl und Verteilung der VoIP-Readiness Messpunkte

Da die IP-Phones / VoIP-Clients normalerweise verteilt im IP-Netzwerk angeordnet sind, lässt sich daraus unmittelbar eine weitere wichtige Anforderung an realitätsnahe VoIP-Readiness Messungen ableiten: die maximal mögliche Verteilung der Messpunkte. Da es kaum sinnvoll ist, die gesamte aggregierte VoIP-Last nur zwischen zwei Punkten zu messen, erübrigt sich zumindest in gängigen Enterprise-Umgebungen somit auch die Notwendigkeit, große (und damit i.d.R. auch teure) Lastgeneratoren einzusetzen. Statt dessen ist der verteilte Einsatz einer Vielzahl kleinerer VoIP-Callgeneratoren im Hinblick auf die Simulation des geplanten VoIP-Systems zielführender. Die genaue Anzahl der VoIP-Callgeneratoren sollte mit der geplanten Größe des VoIP-Systems korrespondieren. Dabei kann es durchaus erforderlich sein, 10, 20 oder noch mehr solcher Geräte einzusetzen. Deren Verteilung in der IP-Infrastruktur sollte sich nach deren physikalischer und logischer Struktur richten und z.B. worst case Konstellationen (schmalbandige Anbindung, kaskadierte Komponenten, etc.) mit berücksichtigen.

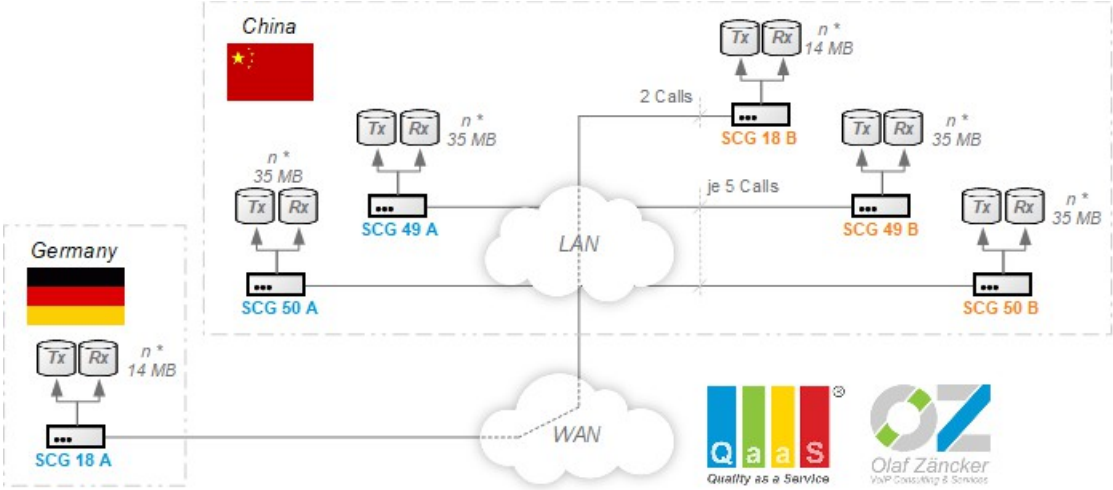


Das obige, aus der Praxis stammende, anonymisierte Szenario zeigt exemplarisch die Messung eines sich über 6 Gebäude erstreckenden LANs mit kaskadierten Switches.

In diesem Beispiel wurden insgesamt 6 Paar Callgeneratoren (SCG), immer aus einem A-

Teilnehmer (blau) und einem B-Teilnehmer (orange) bestehend, eingesetzt. Dabei wurden die Paare grundsätzlich so platziert, dass deren Verkehrsströme notwendigerweise über die komplette Switch-Kaskade und zusätzlich auch Gebäude-übergreifend über den Core-Switch des DataCenters verlaufen.

Bei VoIP-Readiness Messungen im WAN (siehe nachfolgendes, anonymisiertes Praxisbeispiel) ist die für VoIP zugewiesene / reservierte Bandbreite das ausschlaggebende Kriterium.



In diesem Fall wurden für die WAN-Messung vom HQ in Deutschland zur Niederlassung in China mit dem Callgenerator-Paar SCG18A-SCG18B zwei VoIP-Testcalls aufgebaut. Da zusätzlich auch noch das LAN gemessen werden sollte, kamen zwei weitere Callgenerator-Paare (49, 50) zum Einsatz, die je fünf gleichzeitige VoIP-Testcalls erzeugt haben.

2.5 Parameter der Messung

Eigentlich selbstverständlich – doch vor dem Hintergrund der in der täglichen Praxis immer noch anzutreffenden Ping-“Versuche“ sollte nochmals explizit erwähnt werden, dass eine seriöse VoIP-Readiness Messung wirklich nur unter Verwendung von simuliertem RTP-Verkehr (also auf Application Level) unter Einhaltung der nachfolgend aufgeführten Parameter durchgeführt werden kann.

2.5.1 Mess-Dauer

Ein weiteres Kriterium für die Güte und Zuverlässigkeit einer VoIP-Readiness Messung ist die gesamt Mess-Dauer (nicht Call-Dauer; siehe nächster Unterabschnitt). Wie allgemein bekannt, unterliegt die Auslastung von IP-Infrastrukturen nicht nur tageszeitlichen Schwankungen, sondern variiert oft auch von Tag zu Tag. Somit ist ein Zeitraum von 1 Woche für VoIP-Readiness Messungen unbedingt zu empfehlen.

Wenn sich eine VoIP-Readiness Messung über mehrere Zeitzonen erstreckt, ist eine tägliche Mess-Dauer von 24 Stunden wegen dem Versatz der Zeitzonen zwingend vorgeschrieben. Jedoch bringt das auch bei Messungen innerhalb nur einer einzigen Zeitzone den unbestreitbaren Vorteil einer vollständigen Übersicht – 24 / 7 ist also state-of-the-art !

Was bei der Zeitplanung von VoIP-Readiness Messungen gern außer Acht gelassen wird ist die Tatsache, dass im Falle der Nicht-Eignung anschließend eine Störungsbehebung verbunden mit einer Nachmessung der entsprechenden Teile der IP-Infrastruktur erfolgen muss, welche die Wirksamkeit der Störungsbehebung dokumentiert.

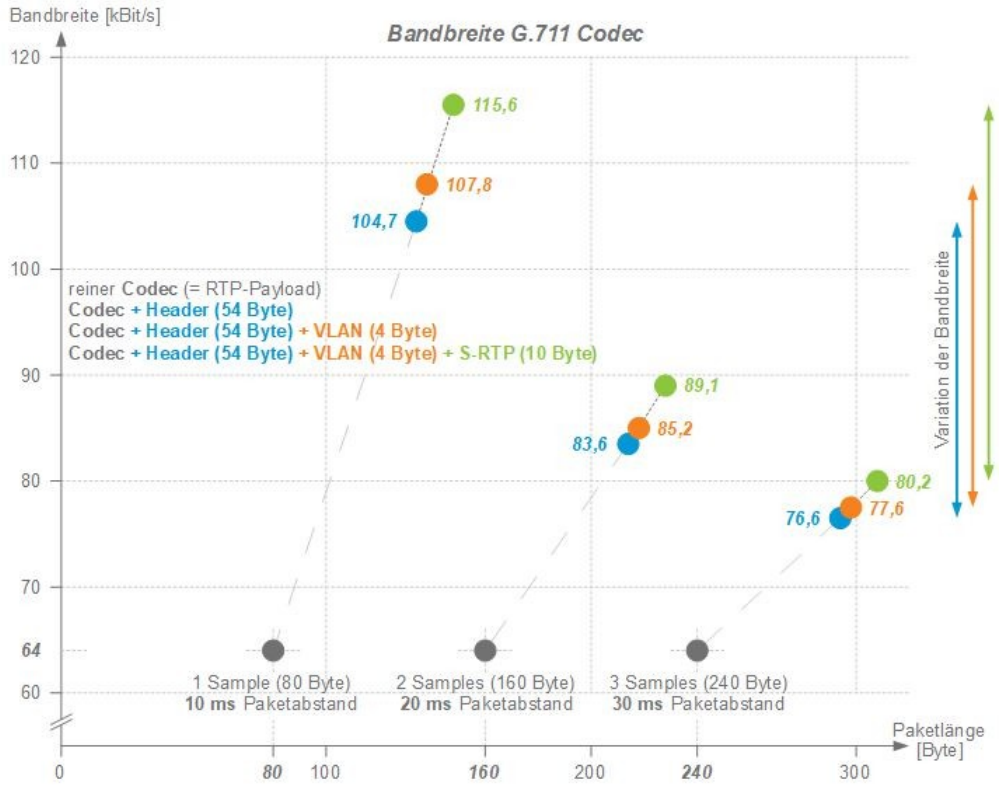
2.5.2 Call-Dauer

Mit Blick auf eine größtmögliche Realitätsnähe einerseits sowie interpretierbare Ergebnisse andererseits haben sich zyklisch wiederholende Call-Sequenzen von 15 Minuten Länge und 15 Minuten Pause -also 2 x stündlich 15 Minuten Calls- bestens bewährt. Verglichen mit dem häufig zu beobachtenden Benutzerverhalten (viele, oft nur kurze Calls im Gegensatz zu einigen wenigen langen, ggf. sogar sehr langen Calls) stellt dieser Wert ein sehr vernünftiges Optimum dar.

Sinnvollerweise ist das Aktivitätsprofil aller Callgeneratoren dabei zeitlich identisch, so dass sich im Core-Switch auch tatsächlich die vorgesehene aggregierte (maximale) VoIP-Last (siehe Unterabschnitt '*Anzahl gleichzeitiger VoIP-Testcalls*') einstellt.

2.5.3 Codec

Codecs sind für die Digitalisierung (genauer: **Co**dierung und **Dec**odierung) der analogen Sprache zuständig. Am verbreitetsten ist der bereits aus dem ISDN bekannte, von der ITU-T spezifizierte Codec G.711. Er komprimiert die Sprache nicht und erzeugt also auch kein zusätzliches Codierungs- (bzw. Dekodierungs-) Delay. Der G.711 ist daher in LANs sehr populär, zumal dort kaum Mangel an Bandbreite herrscht. Auf WAN-Verbindungen steht jedoch naturgemäß weniger Bandbreite zur Verfügung, weswegen dort i.d.R. komprimierende (Narrowband) Codecs, wie z.B. der G.729 Verwendung finden. Im Zuge der Weiterentwicklung von VoIP-Systemen wurden schließlich noch die sog. Wideband-Codecs für erhöhte Audio-Qualität eingeführt. Insbesondere einige der neueren Soft-Clients (VoIP-Applikationen auf PCs) neigen -i.d.R. von den Benutzern unbemerkt- zudem noch dazu, den Codec sogar teils mehrfach auch während eines Gespräches zu wechseln.



Völlig unberücksichtigt bleibt bei solchen Diskussionen immer auch die Tatsache, dass die von einem bestimmten Codec benötigte Bandbreite in nicht unerheblichem Maße davon abhängig ist,

wieviel Sprache (also wieviele sog. Samples) in einem RTP-Paket enthalten ist (sind). Im Fall von G.711 ist so ein Sample, welches 10 ms Sprache enthält, genau 80 Byte lang. Da der gesamte Protocol-Stack (MAC+IP+UDP+RTP Header, ohne VLAN-Tag) eines RTP-Pakets weitere 54 Bytes beansprucht, besitzen G.711-Pakete, wenn sie im Abstand von 10 ms gesendet werden, 134 Byte Länge, bei 20 ms Abstand (zwei Samples) bereits 214 Byte, usw.. Mit dem (Miss-) Verhältnis von Sprache (Payload) zur Gesamtlänge eines RTP-Paketes ändert sich so also auch die (Brutto-) Bandbreite eines Codecs. In der Praxis ist zudem immer wieder festzustellen, dass nicht alle VoIP-Endpunkte sauber auf die selbe SampleRate (Paketabstand) konfiguriert sind, sondern ein Mix von zwei oder noch mehr verschiedenen Paketabständen (und damit Bandbreiten) anzutreffen ist.

Weiteren unmittelbaren Einfluss auf die Bandbreite haben noch VLAN-Tags (zusätzliche vier Byte), S-RTP (Secure RTP gemäß RFC 3711; zusätzliche 10 Bytes, ggf. mehr) und Silence Suppression (nicht quantifizierbare Reduzierung). Die obige Grafik zeigt diesen Zusammenhang exemplarisch für den G.711 Codec.

Weiterhin ist entscheidend, ob über die Netto- (reiner Codec) oder die Brutto- (Codec + Header) Bandbreite gesprochen und ob für die Umrechnung in '*k*' (Kilo) der mathematische (1.000) oder der in der IT übliche Wert von 1.024 verwendet wird. Einige Berechnungen berücksichtigen bei der Brutto-Bandbreite zusätzlich sogar noch Präambel und FCS.

Als wäre das nicht schon genug darf natürlich auch nicht vergessen werden, dass das RTP-Protokoll für die Übertragung der Sprache vorsieht, in der Gegenrichtung RTCP-Reports (Feedback des Empfängers an den Sender; lt. RFC 3550 < 5 %) zu erhalten. Schließlich benötigt noch die Signalisierung (SIP) -mindestens zu Beginn und Ende einer Verbindung- mit den teils sehr langen Paketen einen implementierungsabhängigen Bandbreitenanteil.

Aus diesem Grund und weil die sog. Payload von Datenpaketen (im Falle von RTP also die RTP-Payload = digitalisierte Sprache) in IP-Infrastrukturen ohnehin transparent (also völlig unabhängig vom Codec) übertragen wird, erübrigt sich grundsätzlich die Diskussion hinsichtlich der Codec-Auswahl für VoIP-Readiness Messungen.

Dementsprechend ist G.711 für VoIP-Readiness Messungen in LANs das ultimative Mittel der Wahl. Und im WAN geht es in Fortsetzung des Gedankens lediglich darum, eine entsprechende Bandbreite mit VoIP-Testcalls zu füllen. Sollen im späteren VoIP-Betrieb z.B. komprimierende Codecs Anwendung finden, ist die VoIP-Readiness Messung somit einfach nur mit einer entsprechend geringeren Anzahl von G.711 VoIP-Testcalls durchzuführen.

3 Auswertung

Der grundsätzliche Nachweis der Eignung einer IP-Infrastruktur für die Übertragung von Echtzeitdiensten (VoIP, Fax, Video) ist dabei nur ein Aspekt von VoIP-Readiness Messungen – sofern Fehler / Störungen festgestellt wurden, sollen diese an Hand der Mess-Ergebnisse auch sicher identifiziert und gezielt beseitigt werden können. Und das ist natürlich nur möglich, wenn man möglichst viele detaillierte, vor allem relevante Informationen zur Verfügung hat. Dementsprechend kommt der Aussagekraft der Auswertung eine ganz entscheidende Rolle zu.

3.1 Falsche Mittel zum richtigen Zweck

Selbst wenn die Voraussetzungen für eine VoIP-Readiness Messung in Form der vorgenannten Parameter, wie Anzahl und Verteilung der Messpunkte, Mess- und Call-Dauer, etc. erfüllt sind, kann die Wahl ungeeigneter Auswertemethoden eine Messung wertlos machen.

So behaupten sich die aus der klassischen, analogen Telefonie stammenden Verfahren MOS (Mean Opinion Score), PESQ (Perceptual Evaluation of Speech Quality), etc. einschließlich deren Weiterentwicklungen nach wie vor hartnäckig im Kontext von VoIP-Readiness Messungen. Zwar

befassen sich all diese Standards unbestritten mit der Sprachqualität, jedoch weisen einige der entsprechenden ITU-T Recommendations auf teils erhebliche Ausnahmen und Randbedingungen hin. Auch wird in diesen Dokumenten an keiner Stelle (!) auf die explizite Eignung für VoIP-Readiness Messungen hingewiesen.

Zwei entscheidende Argumente mögen an dieser Stelle genügen, um die Nicht-Anwendbarkeit dieser Verfahren für VoIP-Readiness Messungen zu verdeutlichen:

- bei der digitalen Übertragung macht die Sprache (RTP-Payload) nur einen Teil des Datenpaketes aus, wobei diese natürlich nur als Ganzes (!) entweder verloren gehen (Packet Loss), vertauscht werden (Sequenzfehler) oder verzögert werden (Jitter, Delay) können. Damit ist die oft mathematisch aufwändige Fokussierung auf die RTP-Payload (siehe auch vorhergehender Unterabschnitt '*Codec*') definitiv überflüssig
- die bei diesen Verfahren kalkulierten (!), zu einem Wert zusammengefassten Ergebnisse (z.B. MOS) besitzen kaum noch Informationsgehalt und geben keinerlei konkrete, praktisch relevante Anhaltspunkte für eine eventuell notwendige Störungsbeseitigung. Packet Loss und Jitter beispielsweise können jedoch völlig unterschiedliche Ursachen haben.

Für weitere Einzelheiten zu diesem Thema wird auf das zur selben Reihe '*Illuminating the Background of ...*' gehörende '*White Paper_MOS-PESQ-E Model.pdf*' verwiesen.

Konsequenterweise finden sich in den Vorgaben der VoIP-Hersteller für IP-Infrastrukturen daher auch konkrete Grenzwerte für Packet Loss, Jitter, Delay, etc.. MOS- und ähnliche Werte, selbst wenn sie eine kalkulierte (Schein-) Genauigkeit von zwei Nachkommastellen aufweisen, passen einfach nicht dazu / können dem hohen Qualitätsanspruch nicht mehr gerecht werden.

Der Vollständigkeit halber sei noch darauf hingewiesen, dass auch Packet Loss, Jitter und Delay nicht nur als Minimum-, Maximum-, Mittel- und Prozentwerte, Standardabweichungen, etc. ausgewiesen werden dürfen, da es sich ebenfalls nur um kalkulierte Zusammenfassungen handelt. Wie in den nachfolgenden Abschnitten gezeigt wird, sind für die zuverlässige Bewertung einer IP-Infrastruktur weitere Details, u.a. über den zeitlichen Verlauf dieser Werte unverzichtbar.

3.2 Nachweis der Gültigkeit einer Messung

Eine VoIP-Readiness Messung kann nur dann gültige und damit zuverlässige Ergebnisse liefern, wenn die gesendeten VoIP-Testcalls eine (nahezu) ideale VoIP-Qualität aufweisen. Neben Packet Loss und Sequence Error muss insbesondere auch der Jitter gleich Null sein. Das bedeutet, dass alle erzeugten VoIP- (RTP-) Pakete den gleichen (zeitlichen) Abstand haben – der Fachbegriff hierfür lautet: '*isochron*' (griech: '*iso*' = gleich, '*chronos*' = Zeit).

Ohne einen adäquaten, nachvollziehbaren Nachweis bezüglich der Gültigkeit ist jede noch so detailreiche VoIP-Readiness Messung daher völlig wertlos. Zu einem solchen Nachweis gehören neben einer Skizze des Messaufbaus u.a. auch Aussagen zur Anzahl der Messgeräte, zur ausgewerteten Datenmenge, zur Messdauer sowie eine nachvollziehbare, belastbare Aussage zur Stabilität der erzeugten VoIP-Testcalls.

Das Prinzip ist in der obigen Grafik verdeutlicht. Nur wenn die VoIP-Testcalls auf der Sendeseite

völlig ungestört sind (T_x, linke Seite, grün), können die auf der Empfangsseite ggf. gemessenen Störungen (R_x, rechte Seite, rot) ursächlich der IP-Infrastruktur zugeschrieben werden – ist eine VoIP-Readiness Messung überhaupt erst gültig ! Idealerweise werden die gesendeten VoIP-Testcalls (T_x) daher den gleichen Auswertemethoden unterworfen, wie die empfangenen (R_x).

Werden für die Call-Erzeugung hochwertige, i.d.R. teure Spezialgeräte (Hardware !) verwendet, ist natürlich alles in bester Ordnung - man kann auf die Stabilität der erzeugten Testcalls vertrauen. Das Problem besteht in diesem Fall also nicht in mangelnder Stabilität, sondern in der wahrscheinlich unzureichenden Verfügbarkeit (ein, max. zwei) solcher Geräte in einem Projekt.

Die tägliche Praxis sieht jedoch anders aus – oft werden einfach zwei Laptops verwendet, um nicht nur einen, sondern sogar gleich eine Vielzahl gleichzeitiger VoIP-Testcalls zu erzeugen. Zwar kann man mit einem PC ohne weiteres in guter Qualität telefonieren, jedoch ist das u.a. darauf zurückzuführen, dass VoIP-Clients empfängerseitig über einen sog. Jitter-Buffer (genauer: De-Jitter-Buffer, auch als Receiver-Buffer bezeichnet) verfügen, welcher in der Lage ist, leichte Störungen auszugleichen. Auch sorgen u.a. Funktionen, z.B. FEC (Forward Error Correction) und PLC (Packet Loss Concealment) dafür, dass ein Benutzer nicht jede Störung wahrnimmt. Das bedeutet im Umkehrschluss allerdings nicht, dass PCs automatisch auch isochrone RTP-Ströme mit der Qualität von Mess- bzw. Referenz-Signalen erzeugen können. Gründe hierfür können u.a. das Betriebssystem selbst, andere Applikationen und / oder schwache Performance sein.

Daher ist es auch mehr als fragwürdig, wenn aus Gründen der Einfachheit Callgenerator-Software auf kundeneigenen PCs installiert wird, da man diese natürlich noch weniger unter Kontrolle haben kann, als eigene Laptops.

3.3 Prämissen

Störungen bei der Übertragung von Sprache über IP-Netzwerke können sehr unterschiedliche Ursachen und Auswirkungen haben. Dem entsprechend müssen auch die unmittelbar messtechnisch erfassbaren Parameter, wie Jitter, Packet Loss / Sequence Error und Delay, die einen sehr individuellen Informationsgehalt besitzen, separat ausgewertet werden.

Ebenso essentiell ist es, diese separate Auswertung für den gesamten 24 / 7 Mess-Zeitraum hinweg im zeitlichen Verlauf vorzunehmen, so dass deutlich wird, welcher Parameter wann gestört war. Auf diesem Weg lässt sich u.a. herausfinden, ob die Störungen nur zu bestimmten Zeiten (z.B. werktags während der Regelarbeitszeit) ggf. sogar zyklisch (immer im gleichen Abstand) auftreten oder ob mehrere Parameter zeitgleich betroffen sind. Wie bereits erwähnt, fällt die alleinige Verwendung von Minimum-, Maximum-, Mittel- oder Prozentwerten, o.ä. damit heraus – diese können lediglich ergänzend zu grafischen Darstellungen herangezogen werden.

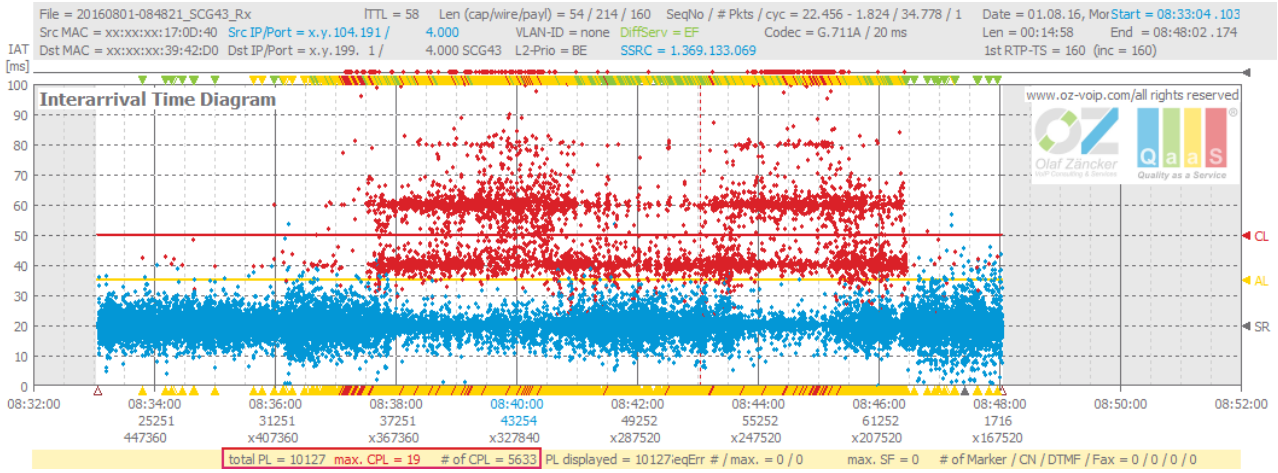
Weiterhin müssen die Auswertemethoden sowohl auf die VoIP-Testcalls, als auch auf den realen VoIP-Verkehr von Benutzern in gleicher Weise anwendbar sein (durchgängiges Auswertekonzept), da sonst der praktische Bezug fehlt.

3.4 Analyse und Diagnose

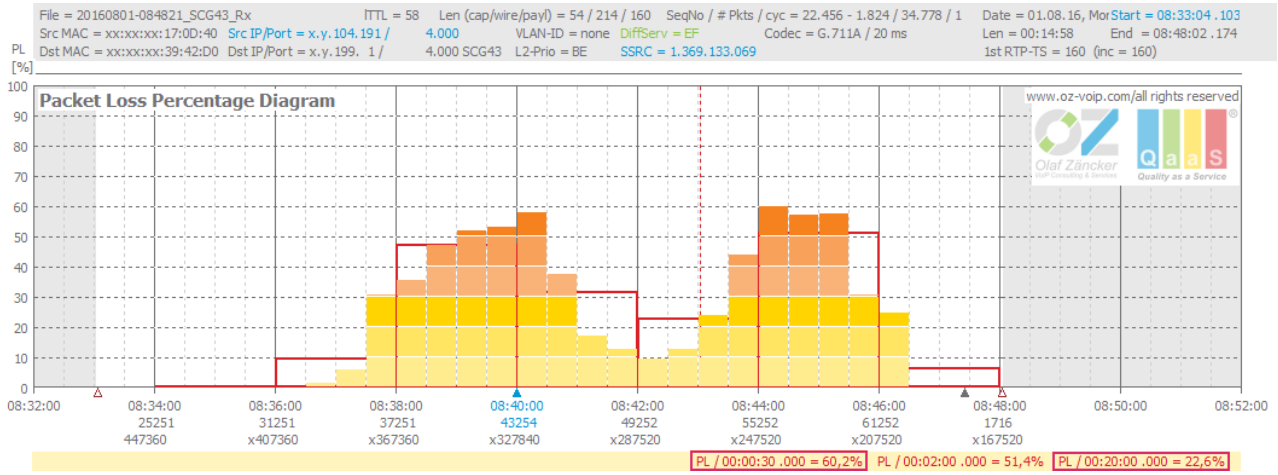
Bei der Analyse geht es um nicht weniger, aber auch nicht um mehr als die detaillierte Auswertung und Aufbereitung / Beschreibung der Messdaten. Diagnose bezeichnet hingegen die auf der Analyse aufbauende Fähigkeit, konkrete Ursachen einer Störung zu erkennen - '*Gnosis*' (griech.) bedeutet nämlich '*(Er-)Kenntnis*'. Beide Begriffe werden in der Praxis gerne miteinander verwechselt, obwohl sie durch ihre Definition klar zu unterscheiden sind.

Geeignete Auswerteverfahren / -tools sind also eine zwingende Voraussetzung. Nachfolgend sind einige praktische Beispiele für die Analyse des zeitlichen Verlaufs von 15-minütigen VoIP-Testcalls gegeben.

Wichtigste Kenngröße der VoIP-Qualität ist der Abstand (Interarrival Time) der RTP-Pakete, der u.a. Informationen über Jitter, Packet Loss (siehe Grafik) sowie weitere Parameter liefert.

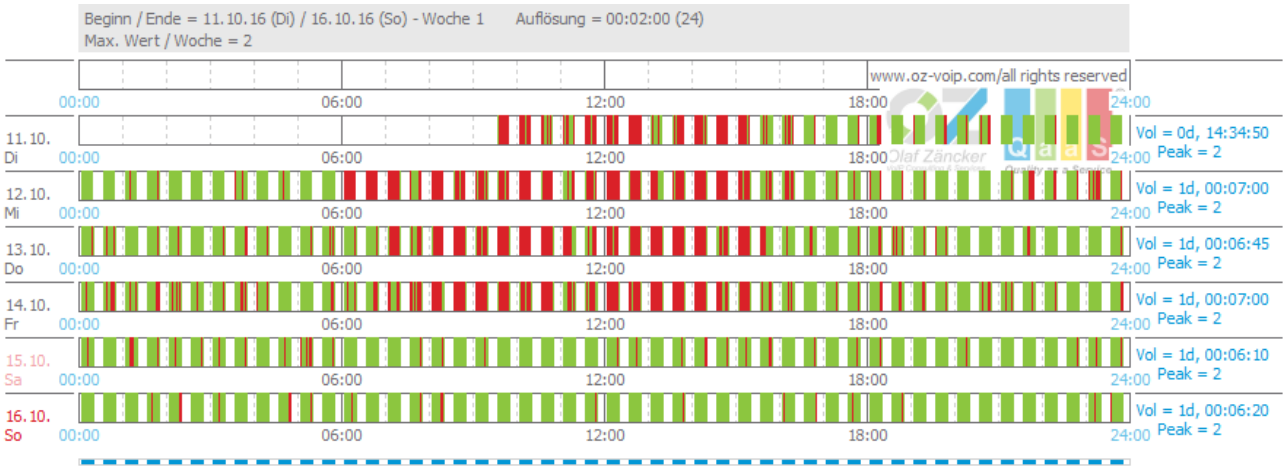


Der zeitliche Verlauf des korrespondierenden prozentualen Paketverlustes ist in der nächsten Grafik dargestellt. Dabei wird nochmals deutlich, dass Maximal- und Mittelwerte immer nur im Kontext eines konkreten zeitlichen Bezuges und als Ergänzung zu anderen Auswertedetails verwendet werden dürfen. So beträgt die mittlere Verlustquote des gesamten 15 ' VoIP-Testcalls bereits 22,6 %, während der innerhalb von 30 Sekunden-Intervallen erreichte Maximalwert der Verlustquote sogar fast bis auf den dreifachen Wert (60,2 %) steigt !



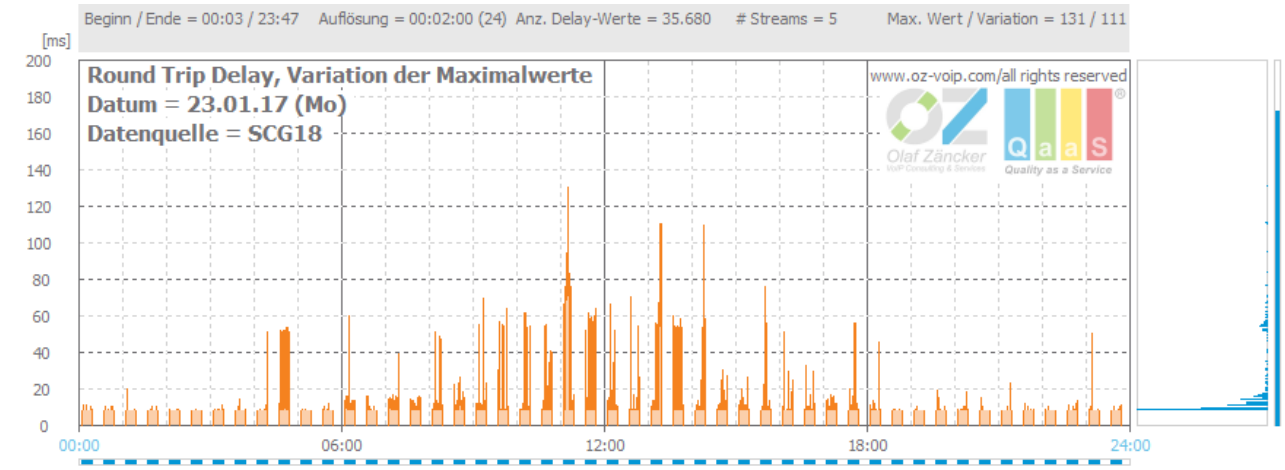
Dabei liefert der nur aus den Grafiken eines VoIP-Testcalls erkennbare Charakter der Paketverluste bereits erste Hinweise auf eine mögliche Ursache. Im nächsten Schritt muss dann die Frage beantwortet werden, ob sich diese Störungen bei dem selben Callgenerator auch zu anderen Zeitpunkten in ähnlicher Form nachweisen lassen. Weiterhin ist zu klären, wie sich die in der selben IP-Infrastruktur eventuell zeitgleich erzeugten VoIP-Testcalls anderer Callgeneratoren darstellen (kontextuelle Analyse).

Neben der unverzichtbaren Fokussierung auf Details ist natürlich auch die makroskopische Gesamtübersicht zwingend erforderlich. Wie an Hand der nachfolgenden, einem anderen Projekt entnommenen, aus insgesamt 48 VoIP-Testcalls pro Tag zu je 15 Minuten Länge erstellten Wochenübersicht eines Parameters leicht nachzuvollziehen, enthüllt sie auf einen Blick, ob ein Zusammenhang mit der Auslastung des betreffenden Teils der IP-Infrastruktur besteht. In diesem Fall (der Parameter ist immer werktags zur Regelarbeitszeit gestört) kann das klar bestätigt werden.



Bei WAN-Messungen über mehrere Zeitzonen lässt sich mittels solcher Wochendarstellungen ggf. auch sagen, an welchem Ende (zu welchen Zeitpunkten) die Störungen verursacht wurden.

Wichtig, in seiner Aussagekraft jedoch häufig überschätzt ist schließlich das Delay. In der u.a. Grafik ist exemplarisch das aus insgesamt 48 VoIP-Testcalls zu je 15 Minuten Länge erstellte Round Trip Delay-Profil eines Tages gezeigt. Die Grundwerte sollten dabei immer ein Abbild der physikalischen Gegebenheiten sein. Im LAN (sehr kurze Übertragungswege) ist von sehr kleinen Werten auszugehen, während im WAN (je nach Länge des Übertragungsweges: innerhalb eines Landes, Kontinents oder sogar interkontinental) u.U. sehr große Werte normal sind. Wie in dem hier gezeigten Beispiel zu erkennen, schwanken die Werte nicht selten auch im Tagesverlauf.



Weitere Anhaltspunkte liefert die gemeinsame Betrachtung von Delay und Jitter. Diese Beispiele illustrieren einen kleinen Ausschnitt der Möglichkeiten professioneller VoIP-Readiness Messungen.

3.5 Andere Medien

Im Laufe der Jahre sind IP-Infrastrukturen zum ultimativen Medium auch für andere gängige (Fax, Video), jedoch auch für vergleichsweise exotische Anwendungen wie z.B. TETRA avanciert.

3.5.1 Fax over IP

Auch zu Zeiten von eMail und Unified Communication hat die Fax-Übertragung immer noch ihre

Daseinsberechtigung. Dabei gibt es -einmal abgesehen von den Möglichkeiten moderner Unified Messaging Systeme- grundsätzlich zwei Möglichkeiten, Faxe in einem IP-Netzwerk zu übertragen. Neben dem robusten T.38 Protokoll kann das Fax-Signal auch mittels unkomprimierter Sprachpakete (G.711 Codec) übertragen (enkapsuliert) werden. Dabei stellt die zeitkritische Fax-Übertragung in Bezug auf Jitter und Packet Loss tatsächlich höhere Anforderungen an die Übertragungsgüte einer IP-Infrastruktur, als Sprache.

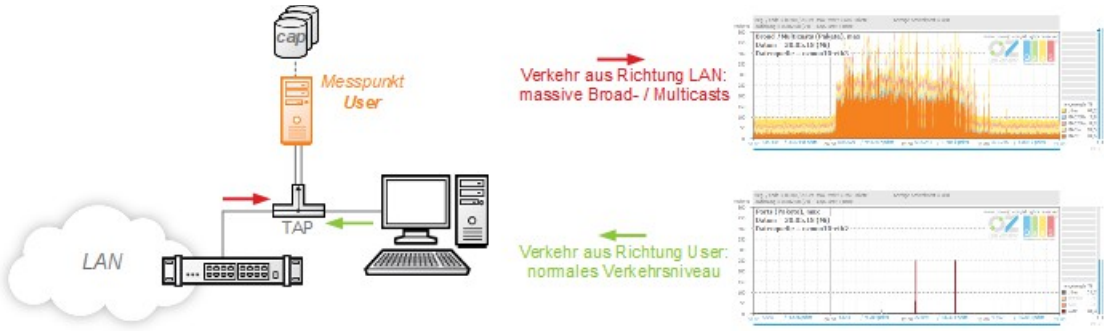
3.5.2 Video over IP

Während Voice und Fax over IP mit konstanten Paketraten und -größen arbeitet, folgt die Übertragung von Video over IP anderen Regeln. Video-Pakete besitzen abhängig vom Bildinhalt eine stark variierende Größe und müssen daher auch nicht in einem exakten (isochronen) Zeitraster transportiert werden. Zusätzliche Robustheit in Bezug auf Störungen, wie z.B. Paketverlust wird auch dadurch erlangt, dass Vollbilder nur in größeren Abständen übertragen werden, dazwischen jedoch nur Bildänderungen (Delta-Frames) erforderlich sind.

Daraus resultiert, dass IP-Infrastrukturen, die für VoIP geeignet (ready) sind, auch bei sporadisch verwendetem Video keine Probleme bereiten sollten. Lediglich wenn Video einem größeren Nutzerkreis zugänglich sein soll, oder Systeme mit sehr hohem Bandbreitenbedarf (z.B. Telepresence) geplant sind, sollten entsprechend modifizierte VoIP-Readiness Messungen mit höherer aggregierter Last in Erwägung gezogen werden.

4 Broad- / Multicasts

Die seitens der Aufgabenstellung bedingte Simulation / Messung / Auswertung von VoIP-Testcalls berücksichtigt zwar die Einflüsse des (non-VoIP) IP-Datenverkehrs auf die VoIP-Testcalls, lässt den IP-Datenverkehr selbst (speziell das eventuelle Vorhandensein eines unzulässig hohen Anteils an Broad- / Multicasts) jedoch notwendigerweise unberücksichtigt. Auch wenn Broad- / Multicasts noch keine messbare Beeinträchtigung der VoIP-Testcalls zur Folge hatten, kann es bei einem zu hohen Anteil von Broad- / Multicasts im späteren VoIP-Betrieb unter realen Bedingungen zu teils erheblichen Störungen kommen.



Aus diesem Grund wird unbedingt empfohlen exemplarisch an einem Arbeitsplatz (Netzwerk-Peripherie) den gesamten bidirektionalen Datenverkehr (natürlich mit Capture-Limit, d.h. ohne Payload) mittels eines passiven Messpunktes (Kombination von TAP und VoIP-Monitor Server) für einen repräsentativen kurzen Zeitraum zu erfassen und auszuwerten (siehe Grafik).

5 Ergebnisdarstellung

Ein praktisch relevanter Ergebnisreport muss mindestens die in diesem WhitePaper erläuterten Aspekte adressieren. Im Einzelnen sind das:

- Management Summary
- Skizze des Messaufbaus mit allen pro Standort eingesetzten Messpunkten sowie Angaben zu IP-Adressen, Verkehrsbeziehungen, Anzahl der VoIP-Testcalls, Priorisierungen, etc.
- allgemeine Angaben zu den VoIP-Testcalls (Codec, SampleRate, Call-Länge)
- Angaben zum zeitlichen Ablauf (Datum von Beginn und Ende der Messung)
- Informationen zur Vorgehensweise bei der Auswertung
- geeigneter Nachweis zur Stabilität der erzeugten VoIP-Testcalls
- Aussagen zu den Messwerten der für VoIP relevanten Qualitätsparameter (Jitter, Packet Loss, Delay)
- Aussagen zu dem exemplarisch an einem Arbeitsplatz erfassten Anteil an Broad- / Multicasts
- bei Störungen der o.g. Qualitätsparameter zusätzlich:
  - Beschreibung der Störung
  - detaillierte Angaben zu Intensität, Häufigkeit, zeitlichem Verlauf, räumlicher Verteilung inkl. entsprechender Grafiken
  - Verifikation der Ergebnisse
  - Trace-Auszüge mit den ermittelten Störungen in Form standardisierter pcap Dateien inkl. Beschreibung (Beweissicherung), um Dritten zu ermöglichen, die Störungen mit einem Tool ihrer Wahl einfach nachzuvollziehen
  - Bewertung der Störungen mit Hinweisen auf die potentielle Ursache (Diagnose)
  - Empfehlungen zur Störungsbeseitigung
- Verzeichnis der Anlagen

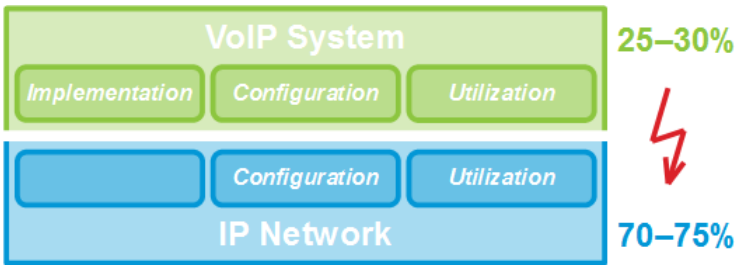
Am Ende eines jeden Ergebnisreports sollte schließlich eine aus den Messergebnissen in nachvollziehbarer Weise abgeleitete belastbare Aussage zur VoIP-Readiness (nicht, bedingt, d.h. nur für bestimmte Teile der IP-Infrastruktur, oder uneingeschränkt) stehen.

6 Maximal mögliche Sicherheit

Eine VoIP-Readiness Messung ist auf jeden Fall ein sehr sinnvoller, empfehlenswerter Schritt um potentielle grundlegende Engpässe in der IP-Infrastruktur (IP-Netzwerk) rechtzeitig aufzudecken.

Eine VoIP-Umgebung besteht jedoch notwendigerweise immer aus den beiden Komponenten IP-Netzwerk sowie VoIP-System. Folglich können natürlich auch beide Komponenten ihre individuellen negativen Beiträge zur VoIP-Qualität leisten. Gemäß langjähriger Erfahrung ist in Bezug auf die, von VoIP-Systemen hervorgerufenen Ursachen im Mittel von einer 25-30 % Quote auszugehen. Diese Ursachen lassen sich im Wesentlichen auf die drei Bereiche Implementierung, Konfiguration und Auslastung zurückführen.

Die nachfolgende Matrix veranschaulicht diesen Sachverhalt grafisch. Bei der Beurteilung der Ergebnisse eines VoIP-Readiness Messung muss daher immer die Tatsache berücksichtigt werden, dass -technisch bedingt und unabhängig vom Messverfahren selbst- maximal die 70-75 % der potentiellen Fehlerquellen im IP-Netzwerk untersucht werden können – das VoIP-System mit seinem 25-30 % -igen Fehlerpotential ist ja noch nicht vorhanden und kann demzufolge auch nicht beurteilt werden.

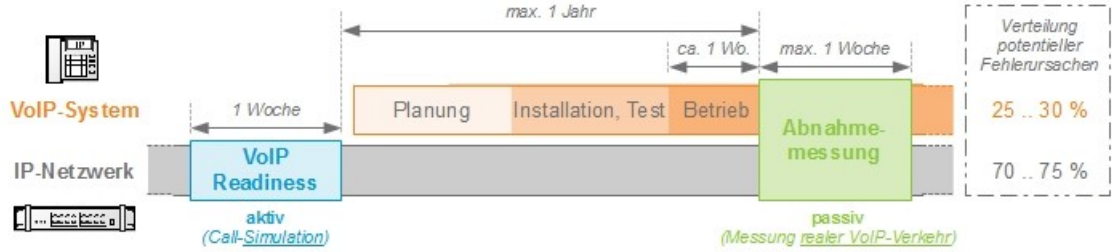


Hinzu kommt, dass sich auch bei sorgfältigster Durchführung unter Berücksichtigung aller, in den vorhergehenden Abschnitten diskutierten Aspekte eine Simulation des VoIP-Verkehrs in Bezug auf die zeitliche und räumliche Verteilung (Menge, Profil, Dauer, Gleichzeitigkeit, etc.) naturgemäß nur bis zu einem bestimmten Grad an das tatsächliche, immer hochgradig individuelle / sporadische Verhalten der VoIP-Nutzer annähern kann. Etwaige Garantien für einen störungsfreien späteren VoIP-Verkehr sind somit unrealistisch / unseriös.

6.1 VoIP-Abnahmemessung

Aus diesen Gründen ist gerade bei größeren VoIP-Systemen unbedingt angeraten, ca. 1 .. 2 Wochen nach Inbetriebnahme des VoIP-Systems, wegen der Aktualität jedoch nicht später als 1 Jahr nach erfolgter VoIP-Readiness Messung mit geeigneten Mitteln eine ergänzende VoIP-Abnahmemessung durchführen zu lassen (siehe nachfolgende Grafik), die in der Lage ist, auch potentielle Störungen des VoIP-Systems mit einzubeziehen.

Anders als bei der VoIP-Readiness Messung (aktive Simulation von VoIP-Verkehr mit Hilfe von Callgeneratoren) wird bei der passiven (also nicht-invasiven) VoIP-Abnahmemessung eine repräsentative Portion des realen VoIP-Verkehrs der Benutzer mittels sog. VoIP-Monitore erfasst und ausgewertet. Wie bereits erläutert werden RTP-Pakete bei der Übertragung immer als Ganzes beeinflusst, so dass auch hier im Interesse maximalen Datenschutzes auf die RTP-Payload (d.h. die Sprache) verzichtet werden kann (Capture-Limit bereits bei der Akquisition der Messdaten). Das technische Setup entspricht dabei dem von VoIP-Troubleshooting Messungen – jedoch mit dem Unterschied, dass es bei der VoIP-Abnahmemessung um den Nachweis der Fehlerfreiheit der VoIP-Umgebung (VoIP-System + IP-Infrastruktur) geht.



Die entscheidenden Argumente für VoIP-Abnahmemessungen bestehen daher auch darin, dass:

- die Gesamtheit aller Einflüsse (IP-Infrastruktur und VoIP-System) auf die VoIP-Qualität
- unter realen Last- / Betriebsbedingungen
- einschließlich eventueller Restpunkte (z.B. Konfiguration des VoIP-Systems)

sichtbar wird.

Das gilt neben dem LAN umso mehr auch für WAN-Verbindungen, da sich die LAN- / WAN-Übergänge und die WAN-Strecke selbst üblicherweise nicht im Zugriff befinden und Probleme mit der VoIP-Qualität weitaus häufiger auftreten.

Wie schon bei der VoIP-Readiness Messung bedarf es auch, oder besonders bei der VoIP-

Abnahmemessung geeigneter Auswertemethoden, um etwaige Störungen oder Restpunkte zuverlässig zu erkennen und in nachvollziehbarer Form aufzubereiten (siehe auch '*White Paper_Passive VoIP-Messungen.pdf*').

7 Fazit

Ab einer bestimmten Größenordnung des geplanten VoIP-Systems ist eine professionelle VoIP-Readiness Messung unverzichtbar. Sie zeigt an Hand einer adäquaten Anzahl simulierter VoIP-Testcalls, wie sich der untersuchte Teil der IP-Infrastruktur zum Zeitpunkt der Messung verhält (Gültigkeitsbereich der Messung). Sind die Messdaten mit geeigneten Methoden ausgewertet und deren Ergebnisse nachvollziehbar aufbereitet, können latente Störquellen in der IP-Infrastruktur vor Einführung des VoIP-Systems gefunden und beseitigt werden.

Wird nach Inbetriebnahme des VoIP-Systems ergänzend eine Abnahmemessung durchgeführt, erhöht das die Betriebssicherheit nochmals und liefert messtechnische Kriterien für die Abnahme.

8 Referenzen, Querverweise

Um den Umfang des vorliegenden WhitePapers in einem vernünftigen Rahmen zu halten, wird an einigen Stellen auf andere, in der selben Reihe '*Illuminating the Background of …*' erschienene WhitePapers verwiesen :

- '*White Paper_MOS-PESQ-E Model.pdf*'
- '*White Paper_Passive VoIP-Messungen.pdf*'