

Real-Time-Core- und -Edge-Switches, Teil 1 Core-Switches

# Klassengesellschaft im Netz

Dafür, dass kritische und wichtige Daten bevorzugt behandelt werden, sorgen LAN-Switches mit Datenpriorisierung und etablieren so eine Klassengesellschaft im Netz. Eine Reihe dieser Systeme musste in unseren Real-World Labs beweisen, wie gut sie diese Priorisierungs-Mechanismen beherrschen.

Lastspitzen sorgen in vielen Ethernet-basierten LANs immer wieder dafür, dass es im Netz eng wird, und dann kommt es je nach Aus- und Überlastung zu teils erheblichen Datenverlusten. Aber auch andere übertragungstechnische Parameter, wie Latency oder Jitter, können – wenn sie gewisse Toleranzwerte überschreiten, weil beispielsweise das Netzwerk und seine Komponenten überfordert sind – zu Störungen einzelner Services oder auch der gesamten Kommunikation führen. Im Zeitalter konvergenter Netze, die Echtzeitanwendungen wie Voice- und Video-over-IP aber auch Produktionssteuerungsdaten in das klassische Datennetz integrieren, führen Datenverluste und andere Störungen in der Praxis zu empfindlichen Kommunikationsstörungen und Produktionsausfällen. So haben in realen Projekten schon mangelhaft priorisierende Switches ganze Call-Center außer Betrieb gesetzt.

## Auswirkungen von Datenverlusten

Bei klassischen Dateitransfers arbeitet das Netzwerk mit möglichst großen Datenrahmen. Bei Echtzeit-Anwendungen teilt sich das Feld. Video-Übertragungen nutzen ähnlich den Dateitransfers relativ große Datenrahmen. Voice-over-IP bewegt sich dagegen im Mittelfeld. Messungen mit Ethernet-LAN-Phones der ersten Generation in unseren Real-World Labs haben beispielsweise ergeben, dass diese Voice-over-IP-Lösung die Sprache mit konstant großen Rahmen von 534 Byte überträgt. Aktuelle Lösungen überlassen es dem IT-Verantwortlichen selbst festzulegen, mit welchen Frame-Größen die Systeme arbeiten sollen. Dabei sollte der IT-Verantwortliche berücksichtigen, dass der Paketierungs-Delay mit kleiner werdenden Datenrahmen kleiner wird. Dagegen wächst der Overhead, der zu Lasten der Nutzdatenperformance geht, je kleiner die verwendeten Pakete sind. Generell kann man bei der IP-Sprachübertragung davon ausgehen, dass mittelgroße Frames verwendet werden. Und auch die meisten Web-Anwendungen nutzen mittelgroße Datenrahmen. Kurze Frames von 64 Byte sind dagegen beispielsweise bei den TCP-Bestätigungspaketen oder interaktiven Anwendungen wie Terminalsitzungen zu messen.

Für eine realitätsnahe und aussagefähige Auswertung der Messergebnisse ist es darüber hinaus entscheidend zu wissen, welche Framegrößen in welchen Verteilungen in realen Netzwerken vorkommen. Die Analyse der Verteilung der Framegrößen, die für das NCI-Backbone do-

kumentiert sind, sowie die Ergebnisse der Analyse typischer Business-DSL-Links haben ergeben, dass rund 50 Prozent aller Datenrahmen in realen Netzwerken 64 Byte groß sind. Die übrigen rund 50 Prozent der zu transportierenden



Datenrahmen streuen über alle Rahmengrößen von 128 bis 1518 Byte.

Für die Übertragung von Real-Time-Applikationen ist zunächst das Datenverlustverhalten von entscheidender Bedeutung. Für Voice-over-IP gilt beispielsweise: Ab 5 Prozent Verlust ist je nach Codec mit deutlicher Verschlechterung der Übertragungsqualität zu rechnen, 10 Prozent

führen zu einer massiven Beeinträchtigung, ab 20 Prozent Datenverlust ist beispielsweise die Telefonie definitiv nicht mehr möglich. So verringert sich der R-Wert für die Sprachqualität gemäß E-Modell nach ITU G.107 schon bei 10 Prozent Datenverlust um je nach Codec 25 bis weit über 40 Punkte, also Werte, die massive Probleme im Telefoniebereich sehr wahrscheinlich machen. Auf Grund ihrer Bedeutung für die Übertragungsqualität haben wir daher das Datenrahmenverlustverhalten als primäres K.O.-Kriterium für unseren Test definiert. Die Parameter Latency und Jitter sind dann für die genauere Diagnose und weitere Analyse im Einzelfall wichtig. Sind jedoch die Datenverlustraten von Hause aus schon zu hoch, können gute Werte für Latency und Jitter die Sprachqualität auch nicht mehr retten. Dafür, dass es zu solchen massiven Datenverlusten im Ethernet-LAN erst gar nicht kommt, sollen entsprechend gut funktionierende Priorisierungsmechanismen sorgen. Bei entsprechender Überlast im Netz sind Datenverluste ganz normal, jedoch sollen sie durch die Priorisierungsmechanismen in der Regel auf nicht echtzeitfähige Applikationen verlagert werden. Arbeitet diese Priorisierung nicht ausreichend, kommt es auch im Bereich der hoch priorisierten Daten zu unerwünschten Verlusten.

## Report-Card / interaktiv unter [www.networkcomputing.de](http://www.networkcomputing.de)

### Policy-based Switching

Feature	Gewichtung	Hewlett-Packard HP ProCurve Switch 5308XL	Extreme Networks Alpine 3804	MRV OptiSwitch Master
Bandbreitenmanagement homogene Frames	20%	5	4	5
Bandbreitenmanagement heterogene Frames	20%	5	3	3
Strict-Priority-Switching Layer-2 Burst-Size 1	15%	5	5	5
Strict-Priority-Switching Layer-2 Burst-Size 100	15%	3	3	2
Strict-Priority-Switching Layer-3 Burst-Size 1	15%	3	5	5
Strict-Priority-Switching Layer-3 Burst-Size 100	15%	2	3	2
Gesamtergebnis	100%	3,95	3,8	3,7
		<b>B</b>	<b>B</b>	<b>B-</b>

A>=4,3; B>=3,5; C>=2,5; D>=1,5; E(1,5);  
Die Bewertungen A bis C beinhalten in ihren Bereichen + oder -;

Gesamtergebnisse und gewichtete Ergebnisse basieren auf einer Skala von 0 bis 5.

Zur Report-Card-Wertung: Um die Flut der Messergebnisse in eine Report-Card verdichten zu können, mussten wir einen Algorithmus definieren, der eine in Bezug auf eine generelle, anwendungsunabhängige Echtzeitfähigkeit aussagefähige, relevante und trennscharfe Ergebnisbewertung ermöglicht. Bewertet haben wir die Differenz zwischen den zu erreichenden Sollwerten und den gemessenen Testergebnissen. Hierzu haben wir in den jeweiligen Disziplinen den jeweiligen Mittelwert der Abweichungen errechnet und nach folgendem Schlüssel bewertet: Abweichung vom Sollwert <= 5% = 5, >5% = 4, >10% = 3, >20% = 2 und >30% = 1.



## Qualitätssicherungsmechanismen

Um solchen Problemen vorzubeugen rüsten die Ethernet-Hersteller ihre Switches mit einer zusätzlichen Funktionalität aus, die es ermöglichen soll, bestimmten Applikationen die Vorfahrt im Netzwerk einzuräumen. Diese Priorisierungs-Mechanismen werden allgemein als Class-of-Service oder – verfälschend in Anlehnung an ATM – als Quality-of-Service bezeichnet. Standardisiert ist eine achttstufige Priorisierung auf Layer-2 nach IEEE 802.1p/Q und auf Layer-3 nach RFC 1349/2474/2475. Die jeweils zugeordnete Priorisierung lesen die Systeme aus den Headern der Datenpakete aus.

Wie Switches die Datenpakete dann gemäß ihrer Priorität behandeln, hängt von den jeweils implementierten Queuing-Mechanismen ab. So besteht die Möglichkeit, bestimmten Daten absolute Vorfahrt einzuräumen oder auch für niedrigere Prioritäten Mindestdurchsatzraten zu garantieren. Eine gute Queuing-Strategie sollte folgende Voraussetzungen erfüllen:

- ▶ Sie muss die faire Verteilung der Bandbreite auf die verschiedenen Serviceklassen unterstützen. Dabei sollte auch die Bandbreite für besondere Dienste berücksichtigt werden, so dass es zu bestimmten Gewichtungen bei der Fairness kommen kann.
- ▶ Sie bietet Schutz zwischen den Serviceklassen am Ausgangsport, so dass eine Serviceklasse mit geringer Priorität nicht die anderen Serviceklassen anderer Queues beeinflussen kann.
- ▶ Wenn ein Dienst nicht die gesamte Bandbreite verwendet, die für ihn reserviert ist, dann sollte diese Kapazität anderen Diensten zur Verfügung stehen, bis der eigentliche Dienst diese Kapazitäten wieder benötigt. Alternativ soll die Bandbreite für diesen Dienst absolut begrenzt werden.
- ▶ Ein schneller Algorithmus, der hardwaremäßig implementiert werden kann, muss für diese Strategie existieren. Nur dann kann diese Strategie auch auf Switches eingesetzt werden, die mit hoher Geschwindigkeit arbeiten. Algorithmen, die nur softwareseitig implementiert werden können, eignen sich lediglich für langsame Switches.

Eine intelligente Ausnutzung der in einem konvergenten Netzwerk vorhandenen Bandbreiten und die Garantie für angemessene Service-Qualitäten der einzelnen Anwendungen setzt eine durchgängige Switching-Policy voraus, die nicht nur blind bestimmten Daten die Vorfahrt anderen gegenüber einräumt, sondern ein sinnvolles Bandbreitenmanagement für das gesamte Netzwerk realisiert. So ist es möglich, bei auftretenden Überlasten die Störungen im Betrieb gering und lokal begrenzt zu halten. Aus diesem Grund haben wir für unseren Vergleichstest Switches gefordert, die CoS-Queuing-Mechanismen sowie Bandbreitenmanagement unterstützen und damit erlauben, nicht nur bestimmten Applikationen Vorfahrt anderen gegenüber einzuräumen, sondern auch durch die Einräumung von Mindestbandbreiten die Service-Qualität niedriger Prioritäten und somit das Funktionieren der entsprechenden Applikationen zu garantieren und dafür zu sorgen, dass Dienste nicht mehr senden als vorgesehen.

### Info

#### Das Testfeld

##### Core-Switches

- ▶ Extreme Networks Alpine 3804
- ▶ Hewlett-Packard HP ProCurve Switch 5308XL
- ▶ MRV OptiSwitch Master

##### Edge-Switches

- ▶ Cisco 2950G-24
- ▶ D-Link DES-3326S
- ▶ Extreme Networks Summit 48si
- ▶ Hewlett-Packard HP ProCurve Switch 5308XL
- ▶ MRV OptiSwitch Master

Die Hersteller von Switches verwenden oft eigene Namen für die CoS-Queuing-Strategien oder ändern die eigentliche Strategie nach ihren Vorstellungen ab. Oft werden auch verschiedene Strategien miteinander kombiniert, um die Ergebnisse zu verbessern. Die ursprünglichen Queuing-Strategien sind:

- ▶ First-In First-Out (FIFO),
- ▶ Strict/Rate-Controlled-Priority-Queuing (PQ),
- ▶ Fair-Queuing (FQ),
- ▶ Weighted-Fair-Queuing (WFQ),
- ▶ Weighted-Round-Robin-Queuing (WRR), Class-Based-Queuing (CBQ) und
- ▶ Deficit-Weighted-Round-Robin-Queuing (DWRR).

Wie diese Verfahren im einzelnen arbeiten, stellen wir aus Platzgründen in einem separaten Grundlagenartikel in diesem Heft ab Seite 48 dar.

## Das Real-World-Labs-Testszenario

Wir wollten wissen, wie gut Queuing-Mechanismen heute in aktuellen LAN-Switches arbeiten und ob es mit ihnen möglich ist, die gewünschte Policy zu realisieren. Aus diesem Grund haben wir ein Feld von Class-of-Service-fähigen LAN-Switches in unseren Real-World Labs an der FH Stralsund einer umfangreichen Testprozedur unterzogen. Um unseren Test im Vorfeld sauber zu strukturieren, haben wir zur Definition unserer Test-Spezifikation, die wir an alle einschlägigen Hersteller gesandt haben, wieder auf unser Modellunternehmen HighFair zurückgegriffen.

Unser Modellunternehmen HighFair möchte neben den klassischen Datenapplikationen und Voice-over-IP weitere Real-Time-Applikationen in ihr Unternehmensnetz integrieren. Ein geeigneter Vergleichstest sollte evaluieren, welche Switches für diese Aufgaben auch unter entsprechender Last geeignet sind. Dabei galt es, verschiedene CoS-Queuing-Mechanismen, wie Strict-Priority-Queuing, Weighted-Fair-Queuing oder Weighted-Round-Robin, auf ihre Eignung für das geplante Szenario zu untersuchen.

Folgende Dienste sollten im LAN der HighFair integriert werden:

- ▶ Videokonferenzen (Video-over-IP, bidirektional, unicast),
- ▶ Voice-over-IP (Call-Center),
- ▶ SAP-Anwendungsdaten sowie
- ▶ übrige Datenanwendungen und Updates.

Um die möglichst absolute Störungsfreiheit der Kommunikations- und Arbeitsprozesse in unserem Modellunternehmen zu garantieren, ist eine vierstufige Daten-Priorisierung sowie eine intelligente Queuing-Policy erforderlich. Gefordert haben wir für die Edge-Switches neben der Datenpriorisierung ein intelligentes Bandbreitenmanagement, das es ermöglicht, von vorneherein eine Überlastung des Backbones zu vermeiden. Aber auch die Core-Switches sollten mindestens zwei CoS-Queuing-Mechanismen beherrschen, so dass die Implementierung der gewünschten und notwendigen Switching-Policy im gesamten Netzwerk möglich ist. Aus diesem Szenario ergaben sich im Detail die folgenden Anforderungen an die Teststellungen.

Teststellung Core-Switch:

- ▶ Layer-3-Ethernet-Switch,
- ▶ mindestens 5 Gigabit-Ethernet-Ports (Multimode-LWL mit SC-Stecker),
- ▶ Datenpriorisierung nach IEEE 802.1p/Q auf Layer-2,
- ▶ Diffserv-Datenpriorisierung nach RFC 2474 oder
- ▶ Type-of-Service-Datenpriorisierung nach RFC 791 und/oder 1349 auf Layer-3 sowie
- ▶ mindestens 2 CoS-Queuing-Mechanismen, wie Strict-Priority-Queuing, Weighted-Fair-Queuing oder Weighted-Round-Robin, die softwareseitig konfiguriert werden können.

Teststellung Edge-Switch:

- ▶ Layer-3-Ethernet-Switch,
- ▶ mindestens 24 Fast-Ethernet-Ports und 2 Gigabit-Ethernet-Ports (Multimode-LWL mit SC-Stecker oder Kupfer),
- ▶ Datenpriorisierung nach IEEE 802.1p/Q auf Layer-2,
- ▶ Diffserv-Datenpriorisierung nach RFC 2474 oder
- ▶ Type-of-Service-Datenpriorisierung nach RFC 791 und/oder 1349 auf Layer-3,
- ▶ mindestens 2 CoS-Queuing-Mechanismen, wie Strict-Priority-Queuing, Weighted-Fair-Queuing oder Weighted-Round-Robin, die softwareseitig konfiguriert werden, sowie
- ▶ Bandbreitenmanagement mit einstellbaren Maximalbandbreiten je Priorität.

Als Testverfahren haben wir Messungen nach RFC 2544 (One-to-Many, Many-to-One, Many-to-Many) festgelegt, die die Parameter Performance, Packet-Loss, Latency und Jitter ermitteln. Analysiert wird dann das unterschiedliche Verhalten der Systeme in den verschiedenen CoS-Queuing-Modi.

Unsere Testspezifikation haben wir an alle einschlägigen Hersteller geschickt und diese eingeladen, an unserem Real-World-Labs-Test in unseren Labs an der FH Stralsund teilzunehmen. Letztendlich folgten die Firmen Cisco, CompuShack, D-Link, Extreme Networks, Hewlett-Packard und MRV unserer Einladung. CompuShacks »GIGALine Xologic 2600M« unterstützte entgegen unserer Test-Spezifikation nur zwei Prioritäten, weshalb wir den Switch aus dem Testfeld herausnehmen mussten. Das Feld der Core-Switches bildeten Extreme Networks »Alpine 3804«, Hewlett-Packards »HP ProCurve Switch 5308 XL« und MRVs »OptiSwitch Ma-

## Info

## Datenverluste mit Strict-Priority-Switching Layer-3 in Prozent

	Ausgang	ToS 6	ToS 4	ToS 2	ToS 0	Verlust gesamt
<b>Extreme Alpine 3804</b>						
<b>Sollwert</b>	<b>100</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>
Extreme FS 64 Byte BS1	100	0,00	0,00	0,00	0,00	0,00
Extreme FS 512Byte BS1	100	0,00	0,00	0,00	0,00	0,00
Extreme FS 1518 Byte BS1	100	0,00	0,00	0,00	0,00	0,00
Extreme FS 64 Byte BS100	100	0,00	0,00	0,00	0,14	0,03
Extreme FS 512Byte BS100	100	0,00	0,00	11,74	11,97	5,87
Extreme FS 1518 Byte BS100	100	0,00	50,85	55,95	55,00	40,05
<b>Sollwert</b>	<b>200</b>	<b>0,00</b>	<b>0,00</b>	<b>100,00</b>	<b>100,00</b>	<b>50,00</b>
Extreme FS 64 Byte BS1	200	0,00	0,00	100,00	100,00	49,75
Extreme FS 512Byte BS1	200	0,00	0,00	99,99	99,99	49,75
Extreme FS 1518 Byte BS1	200	0,00	0,00	99,99	99,99	49,75
Extreme FS 64 Byte BS100	200	0,00	0,14	100,00	100,00	49,78
Extreme FS 512Byte BS100	200	0,00	12,10	87,92	99,99	49,75
Extreme FS 1518 Byte BS100	200	0,00	56,71	57,00	86,28	49,75
<b>Sollwert</b>	<b>400</b>	<b>0,00</b>	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>	<b>75,00</b>
Extreme FS 64 Byte BS1	400	0,00	100,00	100,00	100,00	74,81
Extreme FS 512Byte BS1	400	0,00	100,00	100,00	100,00	74,81
Extreme FS 1518 Byte BS1	400	0,00	99,99	99,99	99,99	74,81
Extreme FS 64 Byte BS100	400	0,00	100,00	100,00	100,00	74,81
Extreme FS 512Byte BS100	400	0,00	100,00	100,00	100,00	74,81
Extreme FS 1518 Byte BS100	400	0,00	99,99	99,99	99,99	74,81
<b>HP ProCurve 5308XL</b>						
<b>Sollwert</b>	<b>100</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>
HP FS 64 Byte BS1	100	0,00	0,00	0,00	0,00	0,00
HP FS 512Byte BS1	100	0,00	0,00	0,00	0,00	0,00
HP FS 1518 Byte BS1	100	0,00	0,00	0,00	0,00	0,00
HP FS 64 Byte BS100	100	0,00	0,00	0,00	0,11	0,03
HP FS 512Byte BS100	100	0,00	0,00	0,00	0,00	0,00
HP FS 1518 Byte BS100	100	0,00	0,00	0,00	0,00	0,00
<b>Sollwert</b>	<b>200</b>	<b>0,00</b>	<b>0,00</b>	<b>100,00</b>	<b>100,00</b>	<b>50,00</b>
HP FS 64 Byte BS1	200	28,70	26,41	44,57	100,00	49,67
HP FS 512Byte BS1	200	0,00	0,00	100,00	100,00	49,75
HP FS 1518 Byte BS1	200	0,00	0,00	100,00	100,00	49,75
HP FS 64 Byte BS100	200	28,45	27,81	43,56	100,00	49,71
HP FS 512Byte BS100	200	0,00	66,24	67,30	66,41	49,74
HP FS 1518 Byte BS100	200	0,00	67,92	66,78	65,08	49,70
<b>Sollwert</b>	<b>400</b>	<b>0,00</b>	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>	<b>75,00</b>
HP FS 64 Byte BS1	400	64,89	62,22	72,73	100,00	74,77
HP FS 512Byte BS1	400	0,00	100,00	100,00	100,00	74,81
HP FS 1518 Byte BS1	400	0,00	100,00	100,00	100,00	74,81
HP FS 64 Byte BS100	400	64,69	62,93	72,21	100,00	74,77
HP FS 512Byte BS100	400	0,00	100,00	100,00	100,00	74,81
HP FS 1518 Byte BS100	400	0,00	100,00	100,00	100,00	74,81
<b>MRV OptiSwitch Master</b>						
<b>Sollwert</b>	<b>100</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>
MRV FS 64 Byte BS1	100	0,01	0,01	0,01	0,01	0,01
MRV FS 512Byte BS1	100	0,02	0,02	0,02	0,02	0,02
MRV FS 1518 Byte BS1	100	0,09	0,08	0,08	0,08	0,08
MRV FS 64 Byte BS100	100	0,01	0,00	0,01	0,12	0,03
MRV FS 512Byte BS100	100	0,07	33,18	52,30	52,27	34,12
MRV FS 1518 Byte BS100	100	0,47	72,84	83,68	83,42	59,51
<b>Sollwert</b>	<b>200</b>	<b>0,00</b>	<b>0,00</b>	<b>100,00</b>	<b>100,00</b>	<b>50,00</b>
MRV FS 64 Byte BS1	200	0,01	0,01	99,99	100,00	49,75
MRV FS 512Byte BS1	200	0,02	0,02	99,99	100,00	49,76
MRV FS 1518 Byte BS1	200	0,07	0,08	99,99	100,00	49,79
MRV FS 64 Byte BS100	200	0,01	0,14	99,99	100,00	49,78
MRV FS 512Byte BS100	200	0,02	50,46	52,73	96,84	49,77
MRV FS 1518 Byte BS100	200	0,13	72,77	83,67	83,38	59,69
<b>Sollwert</b>	<b>400</b>	<b>0,00</b>	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>	<b>75,00</b>
MRV FS 64 Byte BS1	400	0,01	99,99	100,00	100,00	74,81
MRV FS 512Byte BS1	400	0,02	99,99	100,00	100,00	74,82
MRV FS 1518 Byte BS1	400	0,09	100,00	100,00	100,00	74,83
MRV FS 64 Byte BS100	400	0,01	99,99	100,00	100,00	74,81
MRV FS 512Byte BS100	400	0,02	99,99	100,00	100,00	74,82
MRV FS 1518 Byte BS100	400	0,08	100,00	100,00	100,00	74,83

Vergleich der Messergebnisse mit den theoretisch zu erwartenden Sollwerten. FS = Frame-Size in Byte, BS = Burst-Size in Frames, alle Verlustangaben in Prozent. Werte, die deutlich von den Sollwerten abweichen sind hervorgehoben. Die Ergebnisse des Strict-Priority-Switchings auf Layer-2 sowie ein Überblick über alle Sollwerte unter [www.networkcomputing.de/v-test-switches-ergebnisse-2004.pdf](http://www.networkcomputing.de/v-test-switches-ergebnisse-2004.pdf).



## Info

**Real-Time-Switches in Network Computing**

- ▶ Klassengesellschaft im Netz – die Theorie, in NWC 4/04, S. 48.
- ▶ Real-Time-Core- und -Edge-Switches, Klassengesellschaft im Netz, Teil 2 Edge-Switches, für NWC 5/04 vorgesehen.
- ▶ Pilottest Class-of-Service-Queuing: Vorfahrt im Netz, in Special Konvergenz, S. II ff. in NWC 20/03.

ster«. Cisco hat ihren ursprünglich zugesagten Core-Switch – vermutlich auf Grund von Vor-Tests in ihrem Hause – unmittelbar vor ihrem Testtermin in unseren Real-World Labs aus dem Rennen genommen. Als Edge-Switches standen dann Cisco »2950G-24«, D-Links »DES-3326S«, Extreme Networks »Summit 48si«, Hewlett-Packards »HP ProCurve Switch 5308XL« und MRVs »OptiSwitch Master« zur Verfügung.

**Die Core-Switches im Test**

Messtechnisch sind die einzelnen CoS-Queuing-Verfahren zum Teil schlecht auseinander zu halten, da sie faktisch unter entsprechenden Lasten zu einem ähnlichen Verhalten der Systeme führen. Dieser Fakt ist aber auch nicht weiter problematisch, da für einen möglichst störungsfreien Netzwerkbetrieb das konkrete Switching-Verhalten der Systeme und nicht die dahinter stehenden Mechanismen und Theorien entscheidend sind. Konkret haben wir zwei Policies isoliert und messtechnisch untersucht. Zunächst sollten die Switches eine Strict-Priority-Policy umsetzen. Hier kam es vor allem darauf an, dass die Daten der höchsten Priorität unter allen Umständen weitergeleitet werden sollten. Als zweites TestszENARIO haben wir dann mit Bandbreitenmanagement gearbeitet. Hier sollten die Systeme den Datenströmen aller vier Prioritäten Maximalbandbreiten garantieren, um einem Zusammenbruch der Anwendungen niedrigerer Priorität bei Überlast vorzubeugen.

Aus den Ergebnissen von Performance-Messungen wie den von uns durchgeführten ist gut zu erkennen, ob, und wenn ja, in welchem Bereich, das jeweilige System Schwierigkeiten hat. Arbeitet der so belastete Switch korrekt, muss er in allen Fällen gemäß den »Class-of-Service-Regeln« die niedrig priorisierten Daten zugunsten der hoch priorisierten verwerfen. Ein Datenverlust in der höchsten Priorität dürfte bei allen Tests theoretisch nicht vorkommen, nur so würde der jeweilige Switch die fehlerfreie Übertragung der am höchsten priorisierten Echtzeitanwendung, beispielsweise einer Video-Konferenz, garantieren.

Für die Switches sind pro Zeiteinheit um so mehr Header-Informationen auszuwerten, um so kleiner die einzelnen Datenrahmen sind. Ein Switch wird also zuerst Probleme mit 64-Byte-Datenströmen bekommen, wenn er bei der internen Verarbeitungsgeschwindigkeit an seine Grenzen stößt. Bei großen Datenrahmen können je nach Design dagegen schneller Probleme mit

dem Speichermanagement beziehungsweise mit dem Mangel an überhaupt verfügbarem Speicher entstehen.

**Strict-Priority-Switching**

Da im Core-Switch-Test die Last auf dem Switch von vier Gigabit-Ethernet-Ports auf einen Gigabit-Ethernet-Port des jeweiligen Switches gesendet wurde, kam es bei einer maximalen Last von 100 Prozent auf den Eingangsports zu einer vierfachen Überlastung des Ausgangsports. Dadurch ist es natürlich normal, dass die Switches im Test viele Frames nicht übertragen konnten und somit viele Frames verloren. Anhand der Verteilung der einzelnen Prioritäten oder der einzelnen resultierenden Bandbreiten konnten wir dann erkennen, ob und wenn ja wo der jeweilige Testkandidat Probleme hatte.

In einer ersten Messreihe haben wir zunächst Strict-Priority-Switching gefordert und untersucht. Hierzu haben wir mit unseren Smartbits-Lastgeneratoren auf vier Gigabit-Ethernet-Eingang-Ports gesendet und diese Datenströme auf einen Gigabit-Ethernet-Ausgangsport adressiert. Hierbei haben wir Datenströme in den vier Layer-2-Prioritäten – VLAN 7, 5, 3 und 1 nach IEEE 802.1p/Q – erzeugt. Die Eingangslast wird hierbei schrittweise erhöht, so dass die Last an den Eingangsports 25, 33,33, 50 und 100 Prozent betrug, was einer Last am Ausgangsport von 100, 133, 200 und 400 Prozent entspricht. Die Datenströme bestanden aus konstant großen Frames von jeweils 64, 128, 512, 1024 und 1518 Byte. Als Burst-Size haben wir einmal 1 Frame und danach 100 Frames verwendet. Die selbe Messreihe haben wir dann mit Layer-3-Priorisierung nach DiffServ wiederholt und Datenströme der Prioritäten DSCP 48, 32, 16 und 0 generiert. Für die Ergebnisse haben wir die für CoS wichtigen Parameter Frame-Loss, Latency und Jitter ausgewertet. Im Mittelpunkt unserer Analysen steht dabei wegen seiner Bedeutung für die Übertragungsqualität das Datenverlustverhalten.

Verhält sich ein Switch anforderungsgerecht, dann verliert er bei 100 Prozent Last am Ausgangsport noch keine Daten. Bei 133 Prozent Last sollte er dann Totalverlust der niedrigsten Priorität erzeugen, die anderen Streams sollten ohne Verluste ankommen. Bei 200 Prozent Last

sollte der Switch dann die Daten der beiden niedrigen Prioritäten komplett verlieren und die beiden hohen Prioritäten ungehindert passieren lassen. Bei Vollast ist dann ein Totalverlust aller Prioritäten mit Ausnahme der höchsten erforderlich, damit die höchste Priorität noch verlustfrei verarbeitet werden kann. Dieses fehlerfreie Verhalten gilt sowohl für die Layer-2- als auch für die Layer-3-Priorisierung. Generell beherrschten alle drei Core-Switches die vierstufige Priorisierung. Allerdings hatten alle drei Systeme unter bestimmten Bedingungen Schwierigkeiten ihre Vorgaben zu erfüllen. Dabei kam es bei allen drei Core-Switches zu Datenverlusten in den niedrigeren Prioritäten bei einer Burst-Size von 100 Frames, obwohl diese gar nicht erforderlich gewesen wären.

Extreme Networks Alpine-3804 hatte Probleme bei der Layer-2-Messung mit einer Burst-Size von 100 Frames und einer Framegröße von 1518 Byte. Hier kamen die Daten der höchsten Priorität zwar ungehindert ans Ziel, so dass die Höchstlast-Ergebnisse hier nicht weiter auffällig sind. Allerdings verlor der Switch bereits bei 100 Prozent Last am Ausgangsport, als noch gar keine Datenverluste erforderlich gewesen waren, gut 40 Prozent aller Daten. Dabei verteilten sich die Datenverluste relativ gleichmäßig zu jeweils 50 bis 55 Prozent auf die VLAN-Prioritäten 5, 3 und 1. Bis auf geringfügige Verluste in der niedrigsten Priorität verliefen die übrigen Layer-2-Messungen des Extreme-Switches unauffällig. Ein Blick auf die Ergebnisse unserer Messungen mit Layer-3-Priorisierung zeigt ein ähnliches Bild. Hier ist die Schwäche des Core-Switches sogar noch ausgeprägter. Bei der Messung mit 1518-Byte-Paketen und einer Burst-Size von 100 Paketen verlor der Switch bei 200 Prozent Last fast 50 Prozent aller Datenrahmen, wobei auch hier lediglich die Daten der höchsten Priorität ungehindert das System passierten. Hier arbeitete der Switch unsauber, er warf rund 57 Prozent der zweit höchsten Priorität, die eigentlich noch ungehindert das System hätte passieren sollen. Die Daten der beiden niedrigen Prioritäten, die der Extreme-Switch dagegen vollständig zu Gunsten der höheren Prioritäten hätte verwerfen sollen, wurden dagegen nur um 57 beziehungsweise gut 86 Prozent reduziert. Auch die Layer-3-Messung mit 512-Byte-Paketen und Burst-Size 100 führte

## Info

**Datenverluste mit Bandbreitenmanagement in Prozent**

	Ausgang	Prio 7	Prio 5	Prio 3	Prio 1
<b>Sollwert</b>	400	90,00	80,00	70,00	60,00
<b>Extreme</b>	400	91,23	81,67	77,33	49,75
<b>FS1518</b>		-1,23	-1,67	-7,33	10,25
<b>Extreme</b>	400	90,33	88,28	88,28	33,10
<b>FS1518_64</b>		-0,33	-8,28	-18,28	26,90
<b>HP</b>	400	90,22	79,39	69,88	60,38
<b>FS1518</b>		-0,22	0,61	0,12	-0,38
<b>HP</b>	400	90,38	77,89	67,69	63,93
<b>FS1518_64</b>		-0,38	2,11	2,31	-3,93
<b>MRV</b>	400	89,97	79,95	69,92	60,38
<b>FS1518</b>		0,03	0,05	0,08	-0,38
<b>MRV</b>	400	89,98	67,78	55,42	86,89
<b>FS1518_64</b>		0,02	12,22	14,58	-26,89

noch zu deutlichen Datenverlusten, so waren bei 100 Prozent Last in den beiden niedrigeren Prioritäten noch Verluste von fast 12 Prozent zu verzeichnen. Auch bei 200 Prozent Last arbeitete der Switch nicht sauber und erreichte noch eine Verlustrate von 12 Prozent in der zweithöchsten Priorität, die unbeschadet hätte passieren sollen.

Auch Hewlett-Packards Procurve-Switch-5308-XL hatte Probleme bei den Messungen mit einer Burst-Size von 100 Frames. So verlor der HP-Switch bei 200 Prozent Last am Ausgangsport und 1518-Byte-Paketen mit Ausnahme der höchsten Priorität gleichmäßig zwischen rund 65 und gut 67 Prozent der Daten, obwohl der Switch in der zeithöchsten Priorität eigentlich gar keine Daten verlieren sollte. Bei den Messungen mit den kleineren Datenrahmen ist mit abnehmender Rahmengröße eine zunehmende Tendenz zur Priorisierung zu erkennen. Allerdings verlor das HP-System auch noch bei der Messung mit 64 Byte großen Datenpaketen und 200 Prozent Last über 44 Prozent in der zweithöchsten Priorität, die eigentlich ungehindert hätte passieren sollen. Dagegen hat der Switch auch noch Daten der beiden niedrigen Prioritäten weitergeleitet, die das System eigentlich hätte vollständig verwerfen sollen. Die Messungen mit Layer-3-Priorisierung deckten dann noch eine zweite Achillesferse des HP-Switches auf: 64-Byte-Pakete gingen bei 200 Prozent Last am Ausgangsport und einer Burst-Size von 1 selbst in der höchsten Priorität schon zu über 28 Prozent verloren. Bei den Layer-3-Messungen mit Burst-Size 100 zeigte sich dann die schon mit Layer-2-Priorisierung festgestellte Schwäche bei größeren Frames. Bei Höchstlast, Burst-Size 1 und auch 100 sowie 64-Byte-Frames und Layer-3-Priorisierung verlor der Switch dann fast 65 Prozent in der höchsten Priorität.

Der dritte Core-Switch im Bunde, MRVs Optiswitch-Master, hatte Probleme bei unseren Layer-2-Messungen mit einer Burst-Size von 100 Frames und mittelgroßen bis großen Frames. So waren schon bei 100 Prozent Last am Ausgangsport in der obersten Priorität geringfügige Datenverluste von unter 1 Prozent nachweisbar. In den übrigen Prioritäten verwarf der Switch aber auch hier schon deutlich zu viele Daten, so dass das System beispielsweise bei der Messung mit 1518-Byte-Frames fast 60 Prozent der Daten verlor, obwohl hier noch gar keine Überlast den Switch zu Datenverlusten zwang. Bei 200 Prozent Last hat der große MRV-Switch dann zwar nicht mehr Daten verloren, als erforderlich war. Allerdings mussten sich die zweithöchste und die dritthöchste Priorität die Verlustraten hier teilen, was beispielsweise bei der Messung mit 1518-Byte-Rahmen einen Datenverlust von fast 73 Prozent in der zweithöchsten Priorität bedeutete, obwohl in dieser Priorität keine Verluste hätten vorkommen dürfen. Die Messungen des MRV-Systems mit Layer-3-Priorisierung brachten praktisch identische Ergebnisse wie auf Layer-2.

## Switching mit Bandbreitenmanagement

Auch in unserer zweiten Core-Switch-Messreihe haben wir mit unserem Smartbits-Lastgenerator auf vier Gigabit-Ethernet-Eingangs-Ports gesen-

det und diese Datenströme auf einen Gigabit-Ethernet-Ausgangsport adressiert. Hierbei haben wir Datenströme in den vier Layer-2-Prioritäten VLAN 7, 5, 3 und 1 nach IEEE 802.1p/Q erzeugt. Die Eingangslast wird hierbei schrittweise erhöht, so dass die Last an den Eingangsports 25, 33,33, 50 und 100 Prozent betrug, was einer Last am Ausgangsport von 100, 133, 200 und 400 Prozent entspricht. Die Datenströme bestanden aus konstant großen Frames, zunächst haben wir für die VLAN-Prioritäten 7 und 1 1518- und für die VLAN-Prioritäten 5 und 3 64-Byte-Pakete verwendet. Dann haben wir die Messungen mit 1518-Byte-Paketen in allen Prioritäten wiederholt. Die Burst-Size betrug hierbei 1 Frame. Als Maximalbandbreiten haben wir für die höchste Priorität VLAN 7 10 Prozent, für VLAN 5 20 Prozent, für VLAN 3 30 Prozent und für VLAN 1 40 Prozent gefordert. Für die Ergebnisse haben wir dann die für CoS wichtigen Parameter Frame-Loss, Latency und Jitter ausgewertet. Im Mittelpunkt auch dieser Analysen steht wegen seiner Bedeutung das Datenverlustverhalten.

Bei der Priorisierung mit Bandbreitenmanagement sind zwei grundsätzlich unterschiedliche Verhaltensweisen der Switches zu unterscheiden: Rate-Limited-Switching und Weighted-Switching. Beim Rate-Limited-Switching garantiert der Switch den entsprechend konfigurierbaren Bandbreitenanteilen der einzelnen Prioritäten nicht nur eine Mindestbandbreite, er »deckelt« quasi auch die Durchsätze, indem er übrige Bandbreiten, die ein Dienst, dem sie zur Verfügung stehen, derzeit nicht benötigt, auch nicht für andere Dienste verfügbar macht. Eine solche Funktionalität ist insbesondere für den Edge-Bereich je nach Policy unverzichtbar, weil es so möglich ist, der Entstehung von Überlasten bereits in der Netzwerkperipherie vorzubeugen. Switches im Core-Bereich sollten dagegen die Mindestbandbreiten für die ihnen zugeordneten Dienste reservieren. Wenn diese Dienste die ihnen zustehenden Bandbreiten aber nicht benötigen, dann sollten andere Dienste freie Bandbreiten über die ihnen selbst zustehenden hinaus nutzen können. Ansonsten wird die insgesamt im Core-Bereich zur Verfügung stehende Bandbreite unnötig verringert. Ein solches Verhalten kann aber auch im Rahmen der Policy erwünscht sein. Um den verschiedenen Mechanismen gerecht zu werden, haben wir lediglich das Verhalten der Systeme bei Volllast gewertet, da dann unabhängig vom verwandten Mechanismus die gleichen Maximalbandbreiten für die jeweilige Priorität eingehalten werden müssten.

Extreme Networks Alpine-3804 blieb bei den Messungen mit homogenen Frame-Größen in den beiden hohen Prioritäten mit Abweichungen von den konfigurierten Sollwerten von rund 1,2 und 1,7 Prozent relativ dicht an den geforderten Vorgaben. In den beiden niedrigen Prioritäten waren die Abweichungen dann schon deutlich höher. Um rund 7,3 Prozent zu hohe Verluste und über 10 Prozent zu niedrige Verluste lassen es dann schon an Präzision fehlen. Bei der Verwendung heterogener Frame-Größen waren die Unter- und Überschreitungen der Sollwerte dann in den hohen Prioritäten noch deutlich größer.

Eine Abweichung von fast 27 Prozent in der niedrigsten Priorität macht das Verhalten des Systems schon recht unberechenbar. Strikt war dagegen die Einhaltung der Vorgaben für die höchste Priorität. Hier war bei der Verwendung von heterogenen Datenrahmen eine Abweichung von den vorgegebenen 10 Prozent um 0,33 Prozent festzustellen.

Hewlett-Packards Procurve-Switch-5308XL arbeitete in dieser Disziplin deutlich präziser als Extremes Core-Switch. Mit einer maximalen Abweichung in beiden Messungen von 0,38 Prozent in der höchsten und knapp 4 Prozent in der niedrigsten Priorität erfüllte er seine Aufgabe recht gut. Die Ergebnisse des MRV-Optiswitch-Master im Bandbreitenmanagement-Test schwankte dann in der niedrigen Priorität sehr. Vorbildlich war die Einhaltung der Vorgaben bei der Verwendung von homogenen Frame-Größen sowie in der höchsten Priorität bei der Verwendung heterogener Frame-Formate. Mit Höchstwerten von fast 27 Prozent Abweichung von den Vorgaben in der niedrigen Priorität und mit heterogenen Frame-Größen arbeitete der MRV-Switch dann ähnlich unpräzise wie der Alpine-3804.

## Fazit

Insgesamt kamen die Core-Switches im Testfeld relativ gut mit den Vorgaben zurecht. So lange die Burst-Size 1 betrug war die Switch-Welt weitgehend in Ordnung, deutliche Datenverluste in der höchsten Priorität gehören hier der Vergangenheit an. Trotzdem leisteten sich alle drei Core-Switches mehr oder weniger ärgerliche Schwächen.

Extremes Alpine-3804 patzte bei den Messungen mit einer Burst-Size von 100 Frames und großen sowie auf Layer-3 auch mittelgroßen Datenrahmen. Wirklich verlässlich verarbeitete er unter entsprechender Last nur die Daten der höchsten Priorität. Auf Grund dieses Verhaltens

liegt die Vermutung nahe, dass die Pufferspeicher zu klein sind oder das Speichermanagement überfordert war. Auch bei der Verarbeitung von heterogenen Frame-Größen im Bandbreitenmanagement-Test konnte der Extreme-Core-Switch nur bedingt überzeugen.

Auch MRVs Optiswitch-Master zeigte Probleme bei der Verarbeitung großer Datenrahmen. Dies gilt insbesondere für die Messungen mit einer Burst-Size von 100 Frames. Hinzu kommt, dass die mit dem MRV-Switch ermittelten Werte nur erreicht werden konnten, weil die MRV-Experten ihr System unter Zuhilfenahme von versteckten Kommandos, die dem normalen Benutzer nicht zugänglich sind, optimiert hatten. Die erweiterten Setup-Möglichkeiten sollen aber nach Aussage von MRV in einer kommenden Firmware-Version integriert sein. Unter alleiniger Verwendung der im Handbuch enthaltenen Konfigurationsmöglichkeiten wären die Testergebnisse deutlich schlechter ausgefallen. Wenn MRV-Switches entsprechend eingesetzt werden sollen, dann ist daher bis auf weiteres dringend zu empfehlen, dass ein MRV-Experte auch die Konfiguration der Systeme übernimmt.

Hewlett-Packards Procurve-Switch-5308XL priorisierte insbesondere bei einer Burst-Size von 100 Frames teils unsauber und verwarf dann massiv Daten der Prioritäten, die in den jeweiligen Messungen unbeschadet das System hätten passieren sollen. Die Messergebnisse deuten wie schon bei Extreme Networks Core-Switch auf eine Überlastung der Pufferspeicher oder Speichermanagementprobleme hin. Speziell im Layer-3-Betrieb kam dann noch eine Schwäche bei der Verarbeitung von kleinen Datenpaketen hinzu, die auf Probleme in der Verarbeitungsgeschwindigkeit hindeuten. Dabei haben die Ebene-3-Messungen gezeigt, dass der HP-Switch für das Layer-3-Mapping noch zu viel Zeit benötigt und deshalb hier sogar Probleme bei einer Burst-

Size von 1 hatte. Positiv hervorzuheben ist an dieser Stelle, dass HP auf Grund der Ergebnisse unserer CoS-Switch-Tests im vergangenen Jahr, bei denen der damals getestete Procurve-5304XL Strict-Queing überhaupt nicht beherrschte, nachgebessert und die aktuelle Software entsprechend überarbeitet hat. Dass HP im Bereich Bandbreitenmanagement über mehr Erfahrung verfügt, zeigte der Procurve-Switch-5308XL im Bandbreitenmanagement-Test, wo er deutlich präziser als der Mitbewerber arbeitete und sich auch durch die Verwendung heterogener Frame-Größen nicht aus dem Takt bringen ließ, so dass es für HP dieses Mal für einen knappen ersten Platz reicht.

Insgesamt hat unser Core-Switch-Vergleichstest gezeigt, dass die Hersteller inzwischen gelernt haben, die höchste Priorität bei Überlasten der jeweiligen Policy entsprechend bevorzugt zu behandeln. Die Einhaltung der Vorgaben für die zweithöchste Priorität wird dann schon mehr zur Glücksache, keiner der im Testfeld befindlichen Core-Switches hat hier wirklich zur vollen Zufriedenheit gearbeitet. Dabei mehren sich die Probleme deutlich bei einer Burst-Size von 100 Frames. Gerade diese Disziplin ist aber für Core-Switches ein in vielen Unternehmensnetzen häufig auftretendes Phänomen, dem sie gewachsen sein sollten. Noch größere Bursts sollten Netzwerkverantwortliche durch ein geeignetes Bandbreitenmanagement im Edge- wie im Core-Bereich allerdings vermeiden. Wichtig ist es die Burst-Größen mit Hilfe von intelligentem Bandbreitenmanagement immer so klein zu halten, dass die Pufferspeicher nicht überlastet werden.

Kritisch ist generell die Verwendung von mehr als zwei Prioritäten, dies hat auch der vorliegende Test deutlich gezeigt. Die Daten höchster Priorität wurden praktisch durchgängig bevorzugt behandelt. Eine weitere Differenzierung – wir hatten vier Stufen gefordert, acht sehen die zu Grunde liegenden Standards vor – kann schnell zur Glücksache werden. Um diese Situation zu verbessern sind einerseits die Hersteller gefragt. Probleme wie zu knappe Pufferspeicher sollten die Hersteller möglichst bald angehen.

Andererseits sind aber auch die IT-Verantwortlichen in den Unternehmen und die Systemintegratoren gefragt. Ein mehrstufiges Priorisieren, wie es in konvergenten Netzen notwendig ist, erfordert ein wirklich intelligentes Bandbreitenmanagement, das durchgängig im gesamten Netzwerk realisiert wird und dafür sorgt, dass die vorhandenen und entsprechend reservierten Ressourcen nicht überschritten werden. Nur ein intelligentes Rate-Limiting schützt den Core-Bereich des Netzwerks wirklich effektiv. Allerdings ist die Implementation einer solchen Policy nicht gerade trivial. Sie setzt voraus, dass die Lasten, die ein Netzwerk verarbeiten muss, recht detailliert bekannt sind. Dann muss eine exakte Policy erstellt werden. Um diese zu realisieren sind alle Switches im Netz entsprechend zu konfigurieren. Wenn dann auch noch die Switches tun, was ihre Hersteller versprechen, dann klappt es auch mit der Priorisierung und der Kommunikation im konvergenten Netz.

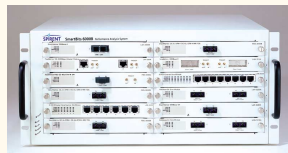
Prof. Dr. Bernhard G. Stütz, [ dg ]



## Info

### So testete Network Computing

Als Lastgenerator und Analysator haben wir in unseren Real-World Labs einen »Smartbits 6000b Traffic Generator/Analysor« von Spirent Communications eingesetzt. Das in dieser Konfiguration rund 250 000 Euro teure System ist mit der Software »SmartFlow« ausgestattet und mit 24 Fast-Ethernet-Ports, fünf Gigabit-Ethernet-LWL-Ports sowie vier Gigabit-Ethernet-Kupfer-Ports bestückt. Alle Ports können softwareseitig als Lastgeneratorkanal und/oder als Analysatorkanal eingesetzt werden. Die Class-of-Service-Eigenschaften der Switches im Testfeld haben wir in verschiedenen Testreihen gemäß RFC 2544 (siehe: [www.ietf.org/rfc/rfc2544.txt](http://www.ietf.org/rfc/rfc2544.txt)) gemessen. In diesen Tests haben wir zunächst die Priorisierung auf Layer-2 nach IEEE 802.1p/Q und dann die Priorisierung auf Layer-3 nach Diffserv



beziehungsweise ToS untersucht. In unseren Testscenarien »Real-Time-Core- und -Edge-Switches« haben wir verschiedene priorisierte Datenströme von einem oder mehreren Eingangsport zu einem oder mehreren Ausgangsport gesendet. Hierbei haben wir Fast-Ethernet- sowie Gigabit-Ethernet-Ports verwendet. Die die Priorisierung festlegenden Bits haben wir im Header der Datenrahmen mit drei Bits nach IEEE 802.1p und auf Layer-2 und nach ToS beziehungsweise Diffserv auf Layer-3 festgelegt. Durch eine gezielte Überlastung der Switches in diesen Tests ist es möglich, das genaue Datenverlustverhalten sowie weitere Testparameter wie Latency oder Jitter zu ermitteln, das Leistungspotential der untersuchten Switches zu analysieren und deren Eignung für bestimmte Einsatzszenarien zu prüfen.