



Open Source an Internet Exchanges

Oliver Kügow
ok@teamix.de

Agenda

- Wie funktioniert BGP-Routing?
- Was ist ein Internet Exchange?
- Betrieb eines Internet Exchanges
- Beispiel: Nürnberger Internet Exchange
- Typische Services an einem IXP
- Open Source am Exchange
- Netzwerken mit Linux
- OIXPT
- Ausblick

BGP-Routing

- ISP haben Standleitungen zu ihren „Upstream-Providern“. Meist mehrere aus Redundanz-Gründen
- Router des ISPs und die der Upstreams tauschen über BGP Erreichbarkeits-Informationen aus
- Dabei schicken die Router jeweils die lokal bekannten Prefixes (mit dazugehörigen AS-Nummern) an den BGP-Partner
- Der Empfänger der Prefixes errechnet sich aus allen Prefixes (und natürlich anderen BGP-Attributen) den optimalen Pfad (die wenigsten AS-Hops) zu jedem Ziel-Prefix

Errechnen der „besten“ Routen

Prefix	AS-Pfad
1.0.0.0/8	111
2.0.0.0/8	111 222
3.0.0.0/8	111 924 59
40.3.2.0/24	111 3 4242

Prefix	AS-Pfad
1.0.0.0/8	222 111
2.0.0.0/8	222
3.0.0.0/8	222 59
40.3.2.0/24	222 4242

40.3.2.0/24	4242
-------------	------

AS 111

AS 222

AS 4242

AS 123

1.0.0.0/8	111
2.0.0.0/8	111 222
3.0.0.0/8	111 924 59
40.3.2.0/24	111 3 4242
1.0.0.0/8	222 111
2.0.0.0/8	222
3.0.0.0/8	222 59
40.3.2.0/24	222 4242
40.3.2.0/24	4242

Prefix	AS-Pfad
1.0.0.0/8	111
2.0.0.0/8	222
3.0.0.0/8	222 59
40.3.2.0/24	4242
12.1.1.0/16	123

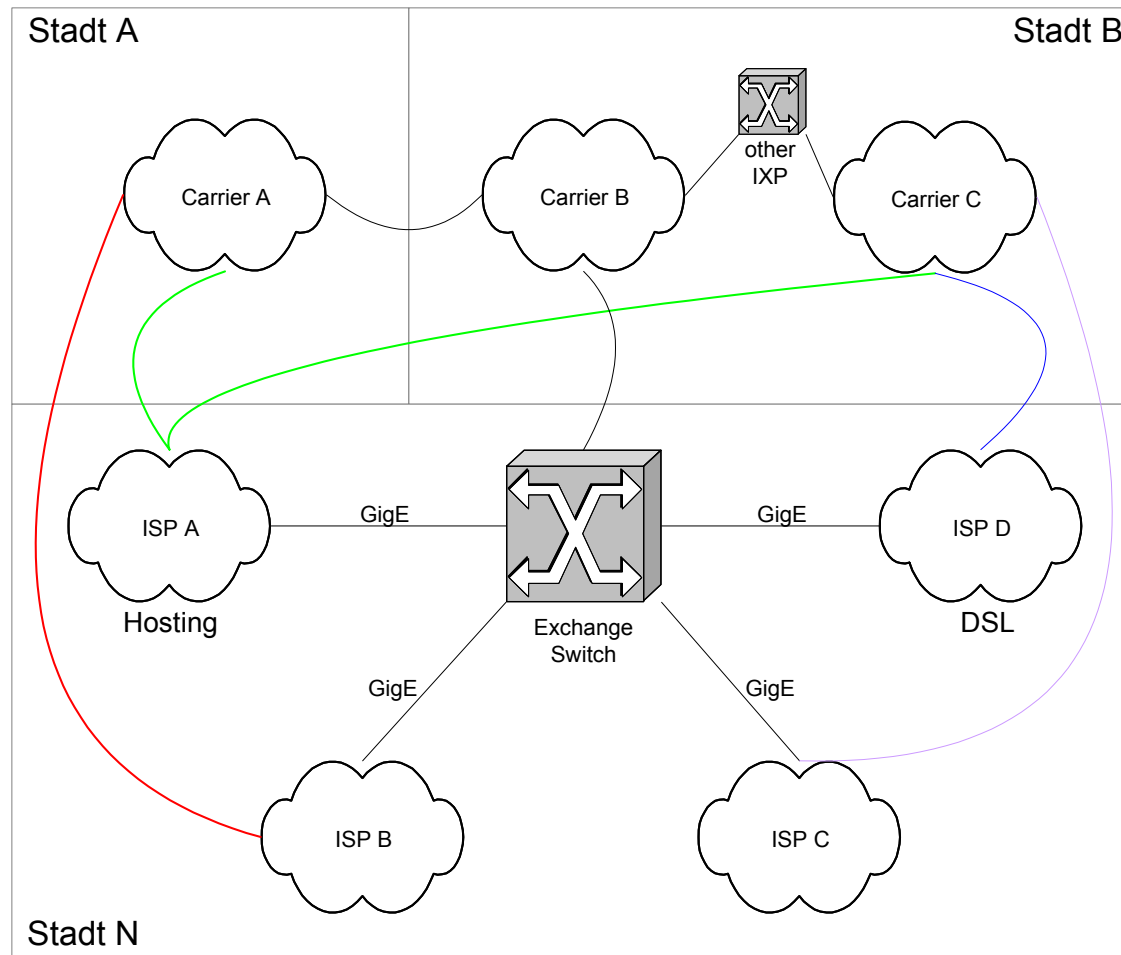
Routing und Kosten

- Daten, die über den Upstream Provider ins „Internet“ übertragen werden verursachen Kosten
- Diese Daten durchqueren das AS des Upstream-Providers, man spricht auch von Transit
- Viele Provider haben „direkte“ Leitungen zueinander, d.h. Daten die zwischen direkt verbundenen ISP ausgetauscht werden sind kostenlos (Peering)
- Früher wurde Peering verstärkt zwischen „gleichberechtigten“ ISPs betrieben. D.h. jeder profitiert gleichermassen vom Datenaustausch

Was ist ein Internet Exchange Point?

- Um Peering zu vereinfachen gibt es IXPs
- Dort „treffen“ sich Leitungen vieler ISPs an einem geteilten Medium
- Meist ein grosser Ethernet-Switch (Layer-2)
- Und fahren Peering-Sessions untereinander
- Also sind alle anderen peering-willigen ISPs direkt (und somit kostenlos) erreichbar
- Der „Rest der Welt“ ist über Upstream erreichbar.

Typischer IXP-Setup



Ziele eines Internet Exchanges

- Kostenreduktion für angeschlossene ISPs (weniger Transit-Kosten)
- Optimierung der Wege im Internet (weniger ASe zwischen Quell und Ziel-AS)
- Erhöhung des Datendurchsatzes (Verbindungen zu einem IXP sind oft schneller als Upstream-Leitungen, da billiger)
- Erhöhung der Redundanz (Oft werden Backup-Transit Vereinbarungen mit befreundeten ISPs über den IXP geschlossen)

Überblick IXPs

- Die 3 grössten in Europa:
 - LINX, London: 38 GBit/s, 145 Members
 - AMS-IX, Amsterdam: 29 GBit/s, 190 Members
 - DE-CIX, Frankfurt: 20 GBit/s, 132 Members
- Mehrere kleine (regionale) in Deutschland:
 - INXS, München, ?? MBit/s, 36 Members
 - B-CIX, Berlin, ?? MBit/s, 17 Members
 - Work-IX, Hamburg, ?? MBit/s, 17 Members
 - E-CIX, Düsseldorf, 290 MBit/s, 11 Members
 - N-IX, Nürnberg, 30 MBit/s, 11 Members
 - Ruhr-CIX, Essen, ?? MBit/s, 3 Members
- Es gibt einen Dachverband der regionalen Austauschpunkte:
DVAP e.V. www.dvap.de

Betrieb eines IXPs

- Politik, Neutralität
Vorsicht: Man steht zwischen den Fronten!
- Verfügbarkeit
Eigentlich nur „Best Effort“, es wird mehr erwartet.
Besonders bei Mehrwertdiensten
- Pricing
Möglichst niedrig, sonst ist Peering teurer als Upstream
- Portspeeds
Möglichst hoch und homogen (kein Paket-Drop!)
- Technisches Management
Out of Band Management nötig! (z.B. Cyclades)

Services an einem IXP

■ Allgemein:

- Alles ausser Peering ist Mehrwertdienst
- Viele Dienste sind an IXPs gut untergebracht, da zentral. Z.B. DNS, News, Mirrors usw.
- Alle angeschlossenen ISPs profitieren, deshalb sollte jeder etwas dazu beitragen (Upstream!)
- **WICHTIG:** Services müssen neutral sein, d.h. **KEINE KONKURENZ** zu den Peers!

Services an einem IXP (2)

- Time-Service: public NTP-Server
- DNS-Service:
 - public recursive Resolver
 - secondary für Peers
 - DNS-Mirrors per Anycast:
.de und f-root Mirrors (in Planung)
- Voice Over IP Terminierung
- File-Mirror Services (z.B. Debian Backports!)

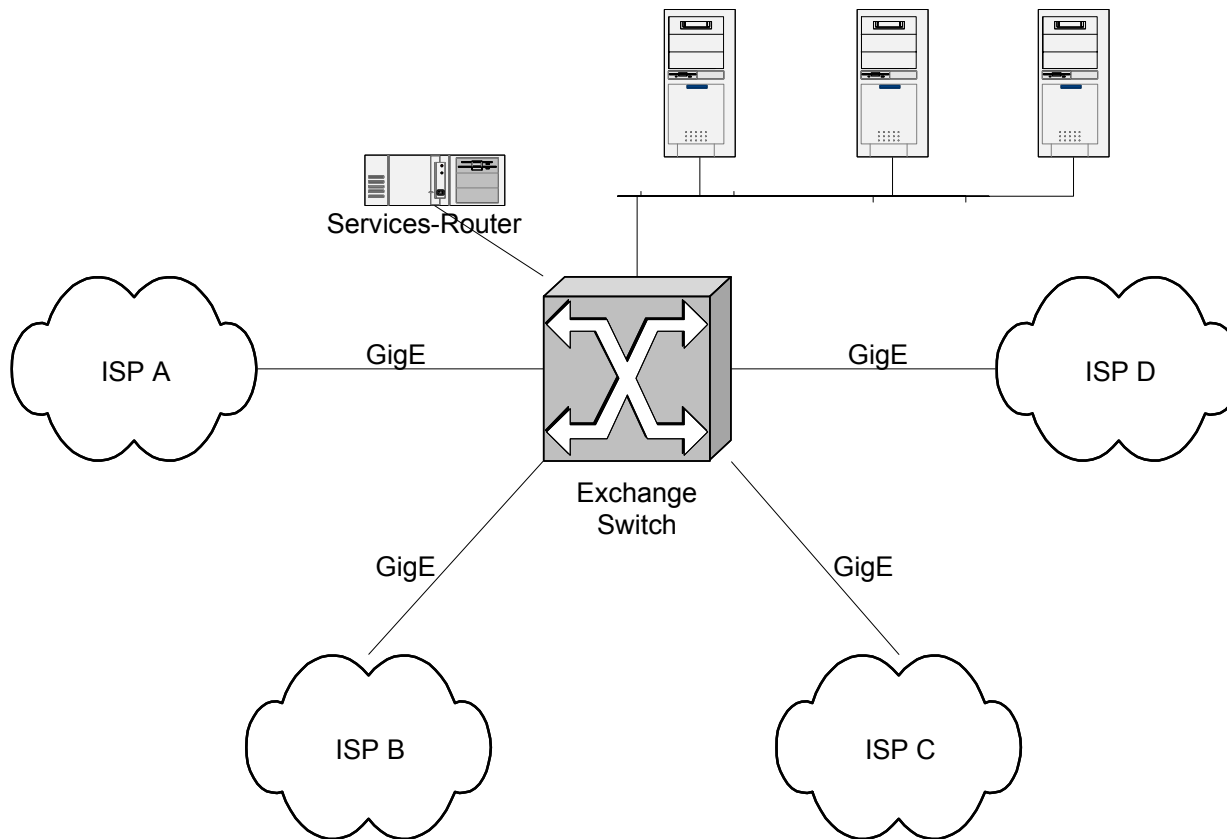
Intermezzo: Debian Backports

- 3 Release Stufen bei Debian:
 - Unstable, Testing Stable
- Problem: Bei stable gibt es keine aktuelle Software, nur Security Updates
- Lösung: Debian Backports
- Backports sind Packages (und deren Dependencies), die aus Unstable genommen werden und neu gegen Libraries aus stable gelinkt werden
- www.backports.org viel aktuelle Software für Woody, Mozilla, Kernel-Packages, SpamAssassin, Postfix, usw!
- Sehr Bandbreitenintensiv, deshalb am Exchange optimal aufgehoben.

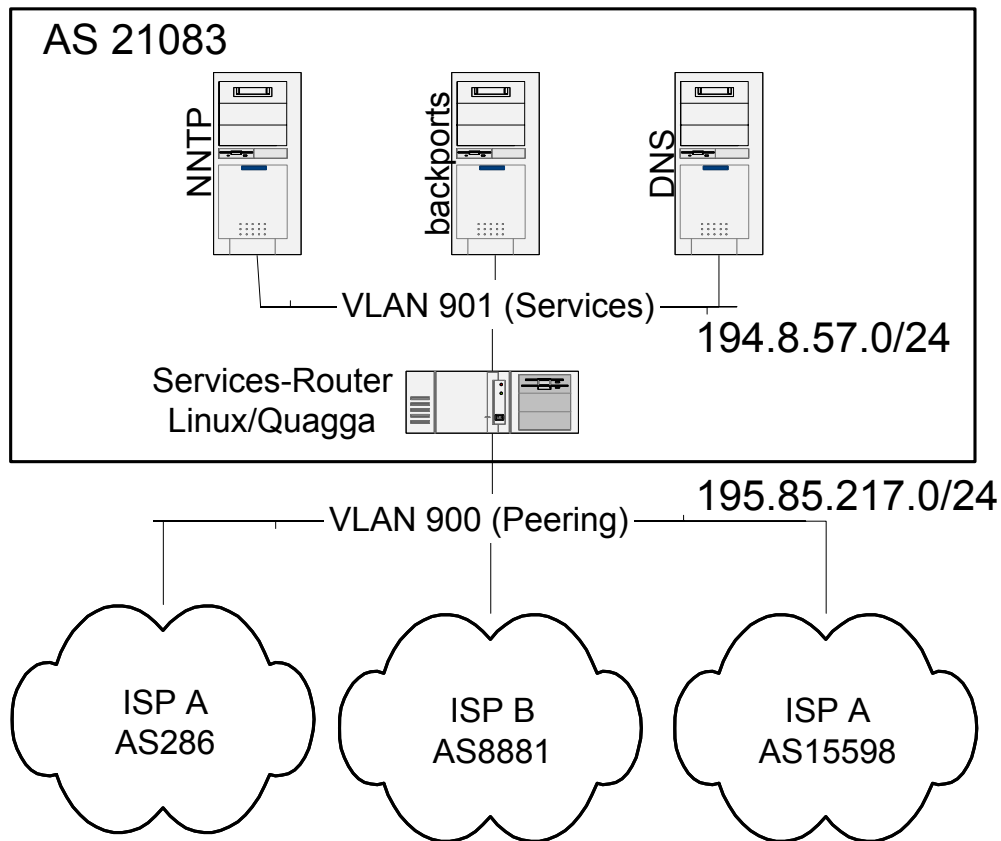
Open Source am N-IX

- Für alles, ausser der Hardware (Switches, Server, Kabel) ☺
- Management:
Out-of-Band Lösung von Cyclades ACS-16
- Backend (Monitoring, Statistik, Verwaltung)
- Frontend Services:
 - Website
 - DNS
 - NNTP
- Alle diese Services sind über einen Linux-Router weltweit erreichbar.

N-IX (physikalisch)



N-IX (logisch)



Services am N-IX erreichbar machen

■ Setup am N-IX:

- Cisco Catalyst, an dem die Teilnehmer direkt mit Gigabit angeschlossen sind
- Getaggte Ports (Vlan-fähig)
- Ausschliesslich Peering am VLAN900
- Private Interconnect über gebuchte VLANs möglich
- N-IX Services laufen über das VLAN901

IP-Adressierung/Routing

- VLAN900 ist Peering-VLAN
195.85.217.0/24
Aus Sicherheitsgründen nicht im globalen Internet geroutet. (Spoofing/DoS/TCP-Resets)
- VLAN901 ist Services-VLAN
194.8.57.0/24
Soll über den Services-Router weltweit announced werden.

Routing-Setup

- Der Services-Router muss Pakete zwischen 2 VLANs schaufeln
- Er lernt von allen Upstreams die optimalen Pfade per BGP
- Und teilt allen Upstreams die über ihn erreichbaren Prefixes (das Services-LAN) mit.

VLAN Support

- Wir setzen Intel E1000 Gigabit Karten ein

```
modprobe e1000
```

- Kernel muss VLANs können

```
CONFIG_VLAN_8021Q
```

```
oder: modprobe 8021q
```

- VLAN-Helper Tool besorgen

```
apt-get install vlan
```

VLAN-Interface konfigurieren

Peering-LAN (da hängen die Upstreams!)

- `vconfig add eth0 900`
- `ifconfig eth0.0900 195.85.217.252`

Services-LAN

- `vconfig add eth0 901`
- `ifconfig eth0.0901 194.8.57.1`

BGP Routing mit Quagga

- Quagga ist ein Routing-Daemon
- Fork des Zebra-Projektes
- Besteht aus mehreren Daemons
- Ein Haupt-Daemon (zebra), der die verschiedenen gelernten Routen in die Kernel-Tabelle einträgt und an andere Routing-Protokolle weiterverteilt.
- Mehrere protokoll-spezifische Daemons, RIPd, OSPFd, ISISd oder BGPd

Grund-Config Quagga

■ /etc/quagga/zebra.conf

```
hostname thunder-quagga
```

```
password geheim  
enable sehr-geheim
```

```
interface eth0.0900
```

```
log /var/log/quagga/zebra.log
```

BGP-Config

■ /etc/quagga/bgpd.conf

```
router bgp 21083
  bgp router-id 195.85.217.235
  network 194.8.57.0/24
  neighbor 195.85.217.1 remote-as 15598
  neighbor 195.85.217.1 description IP-Exchange
  neighbor 195.85.217.1 route-map SERVICES-OUT out
  neighbor 195.85.217.1 route-map IPX in
  neighbor 195.85.217.1 filter-list 2 out
  neighbor 195.85.217.5 remote-as 8881
  neighbor 195.85.217.5 description Versatel
! ... usw, alle anderen Upstreams ...
```

BGP-Config (cont'd.)

```
!  
ip as-path access-list 1 deny .*  
ip as-path access-list 2 permit ^$  
ip as-path access-list 2 deny .*  
ip as-path access-list 5 ^15598  
! .. Weitere access-lists ...  
!  
route-map IXP permit 10  
    match as-path 5  
    set community 21083:15598  
!
```

Check: Announcements vom RIPE

```
quagga> sh ip bgp regex _3333$
```

```
BGP table version is 0, local router ID is 195.85.217.235
```

```
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
```

```
Origin codes: i - IGP, e - EGP, ? - incomplete
```

Network	Next Hop	Metric	LocPrf	Weight	Path
*> 193.0.0.0/21	195.85.217.65	13000	110	0	286 12859 3333 i
*	195.85.217.1			0	15598 3291 3333 i
*	195.85.217.5	0		0	8881 1136 3333 i
*> 193.0.8.0/22	195.85.217.65	13000	110	0	286 12859 3333 i
*	195.85.217.1			0	15598 3291 3333 i
*	195.85.217.5	0		0	8881 1136 3333 i

Check: Pfade zum RIPE

```
quagga> sh ip bgp 193.0.0.0/21
```

```
BGP routing table entry for 193.0.0.0/21
```

```
Paths: (3 available, best #1, table Default-IP-Routing-Table)
```

```
Not advertised to any peer
```

```
286 12859 3333
```

```
195.85.217.65 from 195.85.217.65 (134.222.84.11)
```

```
Origin IGP, metric 13000, localpref 110, valid, external, best
```

```
Community: 21083:286
```

```
Last update: Thu May 6 09:23:31 2004
```

```
15598 3291 3333
```

```
195.85.217.1 from 195.85.217.1 (217.71.108.142)
```

```
Origin IGP, localpref 100, valid, external
```

```
Last update: Mon May 3 13:44:11 2004
```

```
8881 1136 3333
```

```
195.85.217.5 from 195.85.217.5 (195.167.208.97)
```

```
Origin IGP, metric 0, localpref 100, valid, external
```

```
Community: 21083:8881
```

```
Last update: Mon May 3 13:44:36 2004
```

Kernel Routing-Tabelle

- Muss jedes Prefix kennen, da Router in Default-Free Zone

- `ip route list`

```
12.0.0.0/8 via 195.85.217.5 dev eth0:0900 proto zebra equalize
13.0.0.0/8 via 195.85.217.5 dev eth0:0900 proto zebra equalize
15.0.0.0/8 via 195.85.217.1 dev eth0:0900 proto zebra equalize
16.0.0.0/8 via 195.85.217.5 dev eth0:0900 proto zebra equalize
17.0.0.0/8 via 195.85.217.5 dev eth0:0900 proto zebra equalize
.. Viele weitere Prefixes ..
```

- `ip route list | wc -l`
135266

OIXPT (<http://oixpt.sf.net/>)

- Open Internet Exchange Point Toolset
(was dümmeres ist mir nicht eingefallen 😊)
- Viele IXPs haben die gleichen Probleme:
 - Mitgliederverwaltung:
 - Ansprechpartner: Billing, Technik, Notfall
 - Ressourcen: IP-Adressvergabe, VLANs
 - Buchhaltung: Rechnungserstellung
 - Statistik: AS-Path, Port-Statistiken
 - Website:
 - Community/Foren/Mailinglisten/Members-Area
 - Immer ähnliche Inhalte: Technik/News usw. typisches Portal

OIXPT

- Das OIXPT soll nicht nur die technischen Probleme lösen
- Auch generelle Richtlinien und Best-Practices-Guides sollen den Betrieb erleichtern
- Ziel von OIXPT ist, den Ausbau und Neuaufbau der regionalen IXPs zu fördern
- Ein „robusteres“ (dezentraleres) Internet schaffen

OIXPT: Wie weit sind wir?

- Viele gute Ideen sind vorhanden, mehr Feedback von der Community (Users/Peers) gewünscht!
- Mehrere kleine Tools sind bereits fertig (z.B. AS-Path, Looking-Glass)
- Viel Grundlegende Software muss nur richtig zusammengesetzt werden (Webserver, Routing, DB usw.), dafür Anleitungen schreiben
- Im Moment (leider noch) viel Vapor-Ware. Sorry

OIXPT: Was kommt am Ende raus?

- Eine Management-Software für die IXP-Betreiber
 - Leitfäden (technisch) für Betreiber, Peers und potentielle IXP-Betreiber
 - Beispiel-Konfigurationen für div. Router
 - Dokumentenvorlagen wie Anträge, AGBs und Formulare
- => Also ähnliches, wie bei Euro-IX erhältlich, nur: offener, billiger und mehr!

Ausblick: IXP, quo vadis?

Die technische Plattform muss sich ändern:

- Ist schonmal passiert: Jetzt Layer-2 shared Medium auf Ethernet Basis, damals ein Script mit Webmaske, das ATM PVCs „generierte“
- Layer-2 Ethernet skaliert nicht. Der AMS-IX mit 30 Gigabit hat Probleme. ARP (Spoofing, proxy-arp, Sponges) Loops, usw.
- Mehr Quality of Service nötig, da sich auch die Wertigkeit der Exchanges wandelt:

Ausblick: Sinn eines Exchanges

„Mehr“ Dienste als nur „dummes Peering“

- Zentrale Datendrehscheibe in Region
- Vermeidung unnötiger Local-Loops (Multiple Services über 1 Link)
- Dienste, die QOS erfordern (VoIP)
- Media-Distribution (KabelTV/Multicast Inhalte)

Mehr lokale Austauschpunkte

- Etwas Zentralisierung in einer Region ist nötig, sonst machen IXPs keinen Sinn.
- Zuviel Zentralisierung ist schlecht, 85% des deutschen Internet-Traffics laufen über den einen DE-CIX in FFM. Ja. Terror-Gefahr.
- OIXPT soll helfen, IXP leichter zu managen
- Die Einstiegshürde für den „eigenen“ lokalen IXP wird geringer

Fragen??

- Danke fürs Zuhören!
- Wer später noch Fragen hat,
kann sich gerne bei mir melden:

Oliver Kügow

team(ix) GmbH

+49 911 – 30 999 30 (auch VoIP/ENUM!)

ok@teamix.de