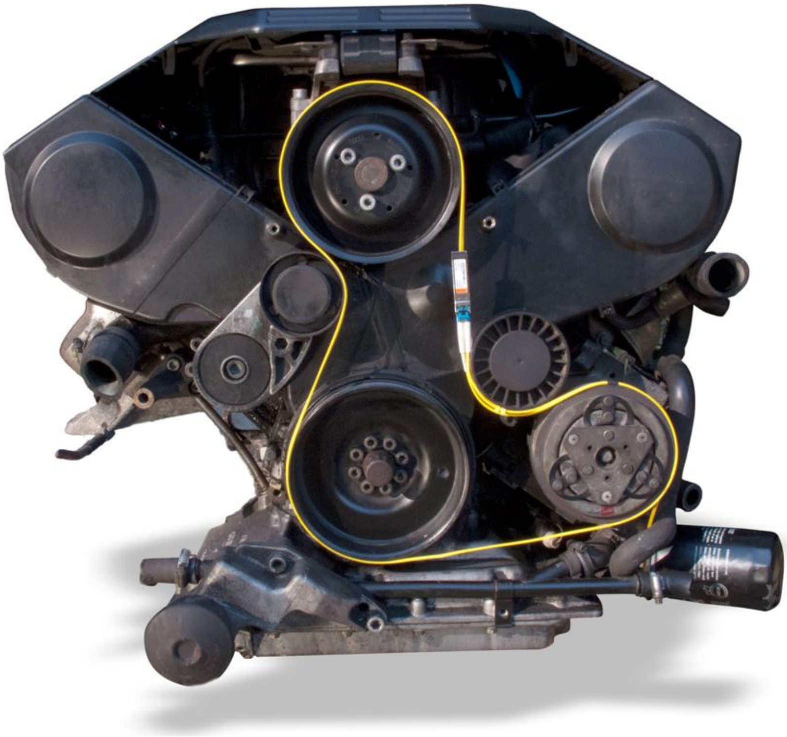
# Transmission 3

five theses for a longer, faster, easier, more reliable & scalable network



flexOptix

Welcome to DeNOG #3 and my presentation Transmission 3 – five theses for a longer, faster, easier, more reliable & scalable network. The theses are:

1. Wouldn't it be fast if we would live in a world where we can transmit 100 000 000 000 single state changes over a single fibre cable ?
2. Wouldn't it be better if we could communicate over a distance of 170km on a naked fibre without any boosters at a speed of 10Gbps ?
3. Wouldn't it be easy if we would have instant access to all kind of gear which we do not have to hand or forgot to order in advance, but our customer still needs the link now and not tomorrow!
4. How reliable would your network be if it would tell you next week there will be failure on your link to your upstream provider ?
5. How well would you sleep if you would know that you could scale up your network instantly if you need more bandwidth ?

Human mankind is already living in the 2 nd part of the Ethernet. Now this here is transmission 3. come to DeNOG #3 for your better future.

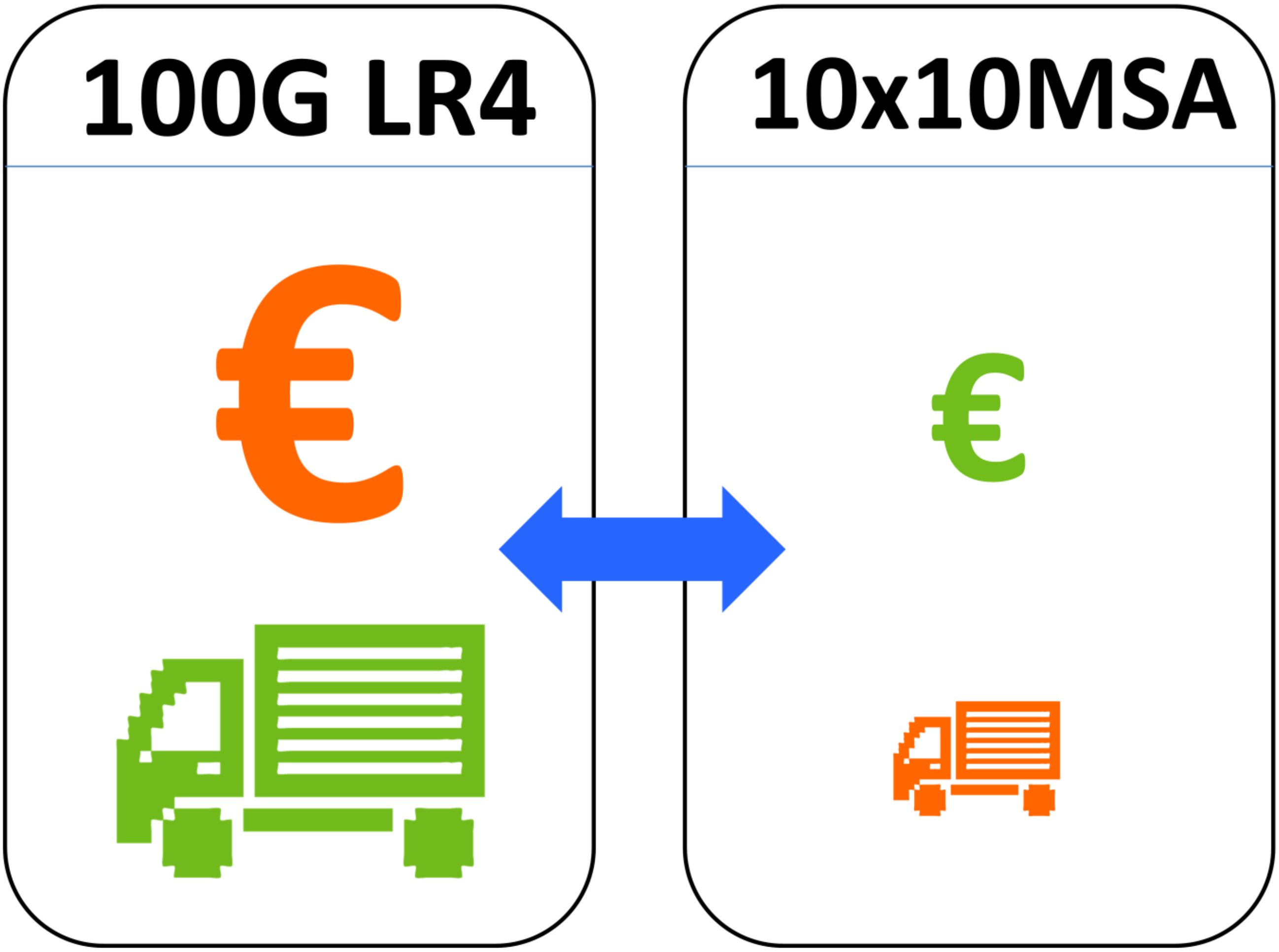
DISCLAIMER: this speech is for technical people only. It is about a collection of optical and non-optical technologies which I have seen, used and worked on in the last 12 month. I think it is worth to mention them because you can use them to make your network longer, faster, easier, more reliable & scalable.

# faster

Thesis 1: Faster

It addresses the current state of 100G Ethernet optics, esp. the two hot types 100G Base-LR4 and 10x10MSA to achieve 100G Ethernet. As we are in a ISP-world we do not cover 100G Base-SR10 which fails by design.

The discussion here won‘t cover the formfactor because this discussion will be ongoing and is more a philosophy question for the system vendors. At the end a higher port density for high speed connections with the lowest power consumption will will. There is a good presentation from Jörg Ammon (Brocade) at the More IP Event (http://more-ip-event.net/assets/pdfs/2_Joerg_Ammon-Brocade-ams-ix-100-gbe_.pdf) with the title „100 GIGABIT ETHERNET AND BEYOND – WHAT NEW TECHNOLOGY WILL MEAN FOR MORE IP“. Jörg covers the different formfactors and developments very good.



100G-BaseLR4:
************
Pricing is very high because of two facts:

1. There is a need to convert the signal from the electrical pane which runs on 10 x 10G to 4 x 25G for the optical side. This is done in the so called gearbox which is expensive
2. The optical components operate on serial 25G which is cutting edge at the moment and low quantities in the assembly line. This leads to a high price

A lot of transceiver manufacturers jumped on this band wagon and already have some products available (since the beginning of 2011).

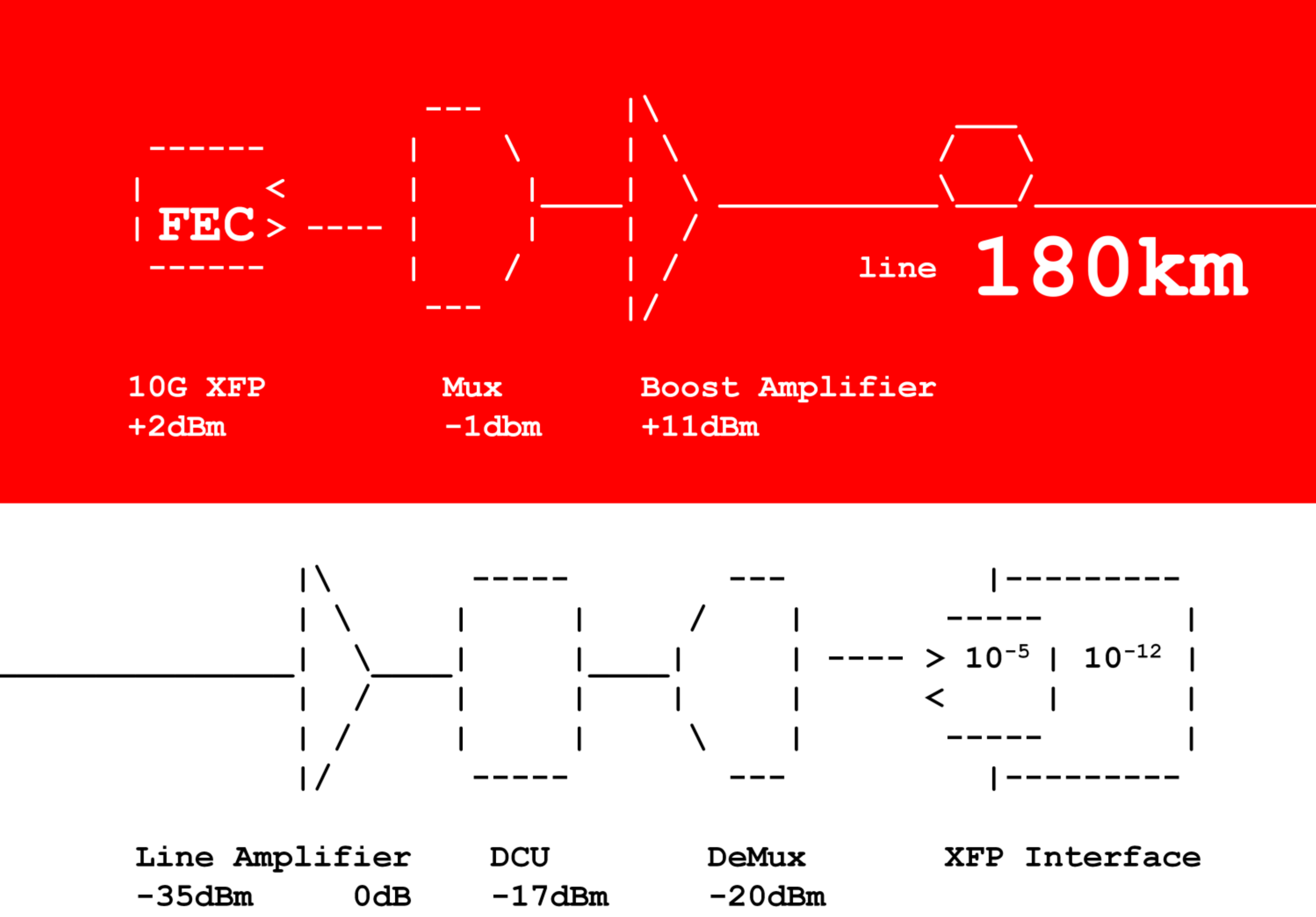
10x10MSA:
*********
Pricing is way better than for 100G-BaseLR4 because

1. There is no need for a gearbox because
2. The optical components operate with 10G only

Souring today is a big issue. There are only a couple of transceiver manufacturer in this MSA working group. The main force of this working group is organized by end user companies, e.g. Google, Facebook, etc.

Thesis 2: Longer

80km distance must not be the end for your 10G link. There are solutions which make it easy to extend this without mayor drawbacks.



FEC

line 180km

10G XFP
+2dBm

Mux
-1dbm

Boost Amplifier
+11dBm

$10^{-5}$ | $10^{-12}$

Line Amplifier
-35dBm 0dB

DCU
-17dBm

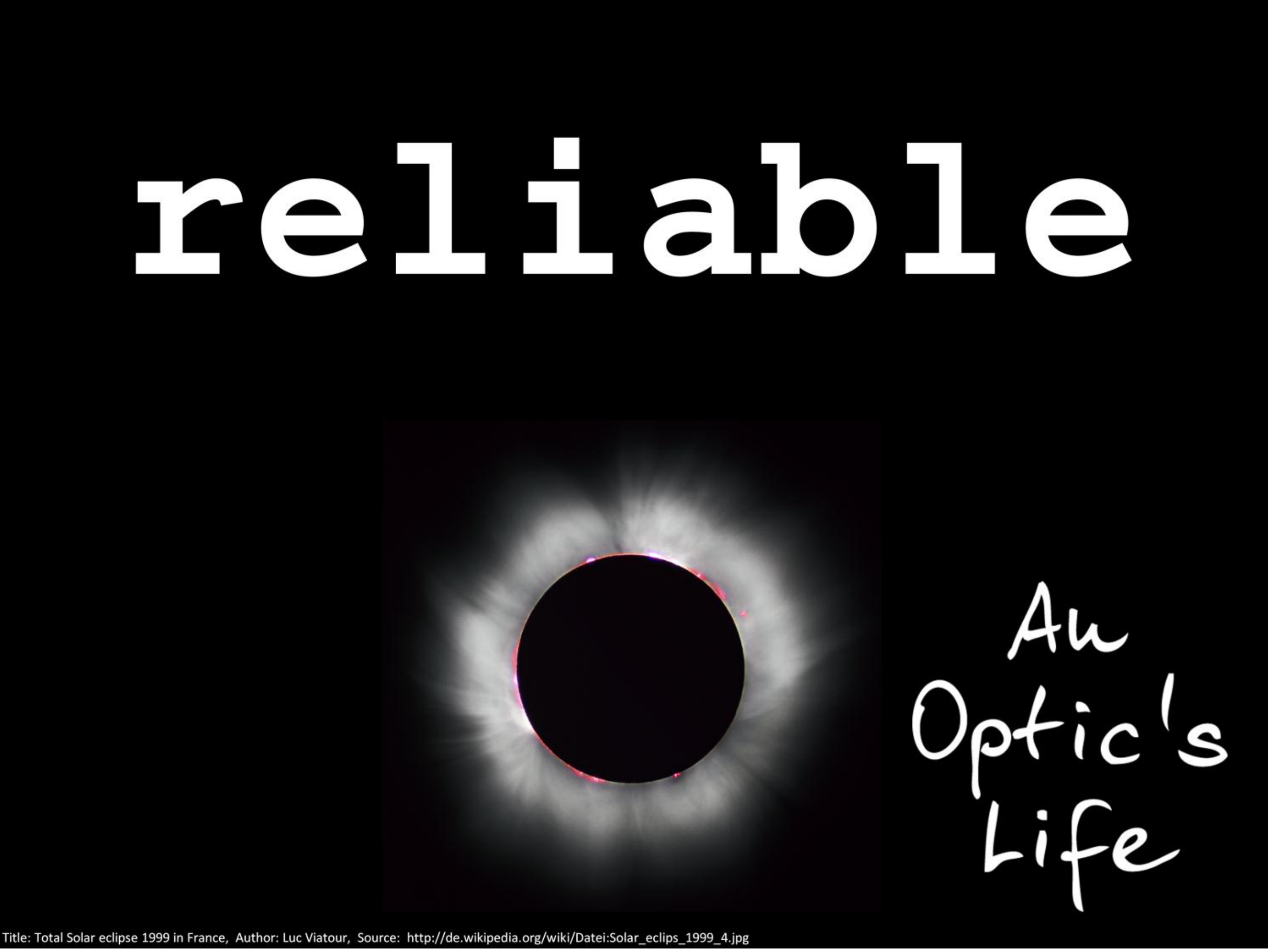
DeMux
-20dBm

XFP Interface

FEC – Forward Error Correction does add additional payload the datastream. So the overall datarate is higher, e.g. 10G over 709.G ends up with datarate of 10.7 Gbps. Regular Ethernet has 10.3125 Gbps. The receiver uses this addidtional payload to check and correct the transmitted data if there are biterrors during the transmission.
DeMux / Mux: optical wavelength Multiplexer to add several wavelenths – either CWDM oder DWDM - on one fiber. There is no influence of the different waves with eachother.
Boost Amplifier: adds up to 20dB to the signal and the noise. The Signal to Noise Ratio is very important, otherwise the receiver can't differenciate between the actual signal and the noise.
Line / Pre Amplifier: can add on other 35dB to the signal. There are two different technologies, either EDFA or Raman Amplifiers. The last one has the benifit to amplify only the signal and keeps the noice low. This is done via backwards pumping to stimulate the line (fiber) itself to amplify the signal (i am not a physician….so this explanation is quite rare)
DCU – Dispersion Compensation Unit can be done via a special dotted fiber (DCF) or grating. This is needed for distances beyond 80km because the signal will be stretched in time due to some physical laws and rules of fiber optics and the media. With the DCU this time-streteching will be compensated.

**Result: the interessting part of this setup is that the 180km are completely passive without any active gear in line. This is very nice because the failure points will be both ends and not between. So troubleshooting will be easy in the case of a failure. At the receiving interface (right bottom corner of the slide) the signal has a Bit Error Rate of $10^{-5}$ which is for Ethernet not sufficient but in this case it is not a problem because the XFP takes care with ist FEC capability to boost this BER to $10^{-12}$. This BER is seen by the router/switch interface and $10^{-12}$ is ok for Ethernet.**

The power levels show the output power of each device. At the Line Amplifier also the input power iof -35dBm is mentioned. This leads to a overal attenuation of the 180km line of 46dB (0,25dB/km).



reliable

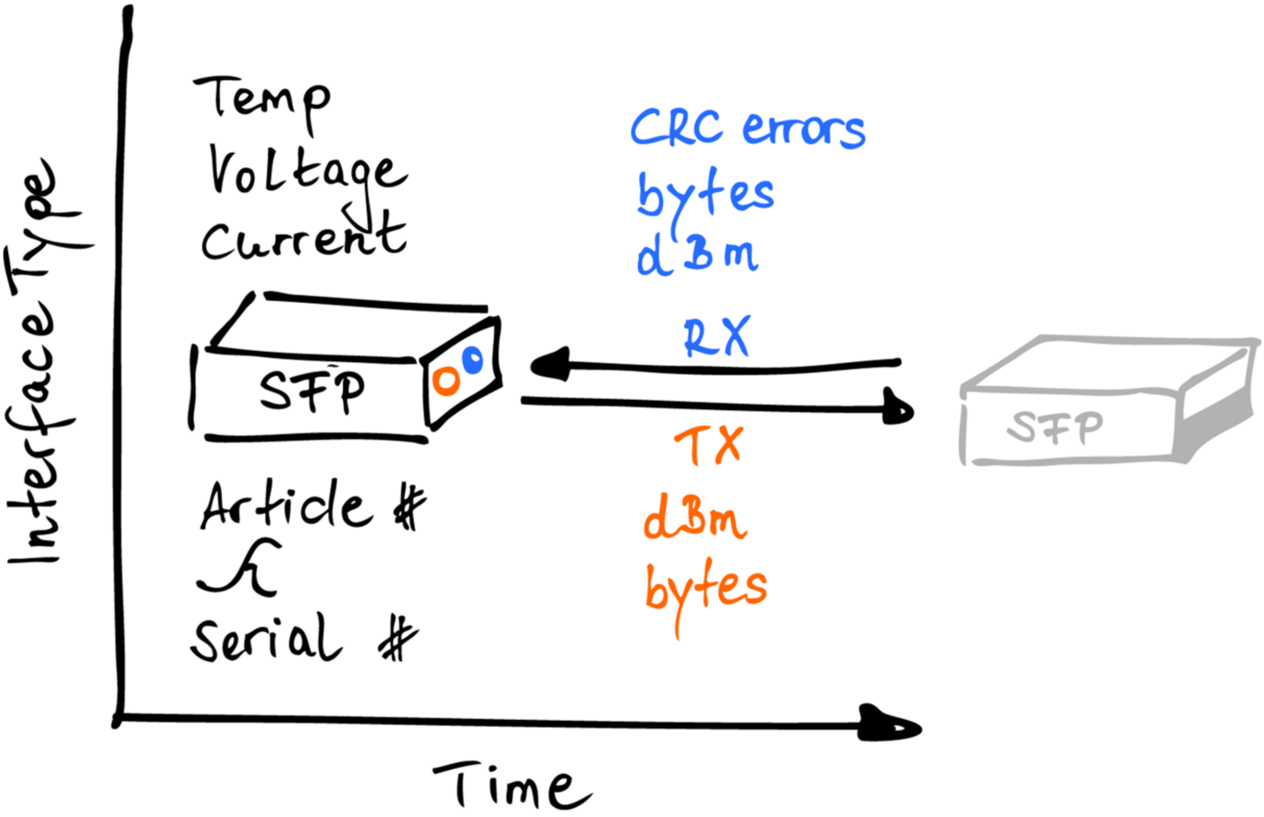
An Optic's Life

Title: Total Solar eclipse 1999 in France, Author: Luc Viatour, Source: http://de.wikipedia.org/wiki/Datei:Solar_eclips_1999_4.jpg

Thesis 3: more reliable

**The goal of the project “An optic’s life” is, to predict the time when a transceiver will reach its real end-of-life-time based on the actual setup in the datacenter / colocation.**

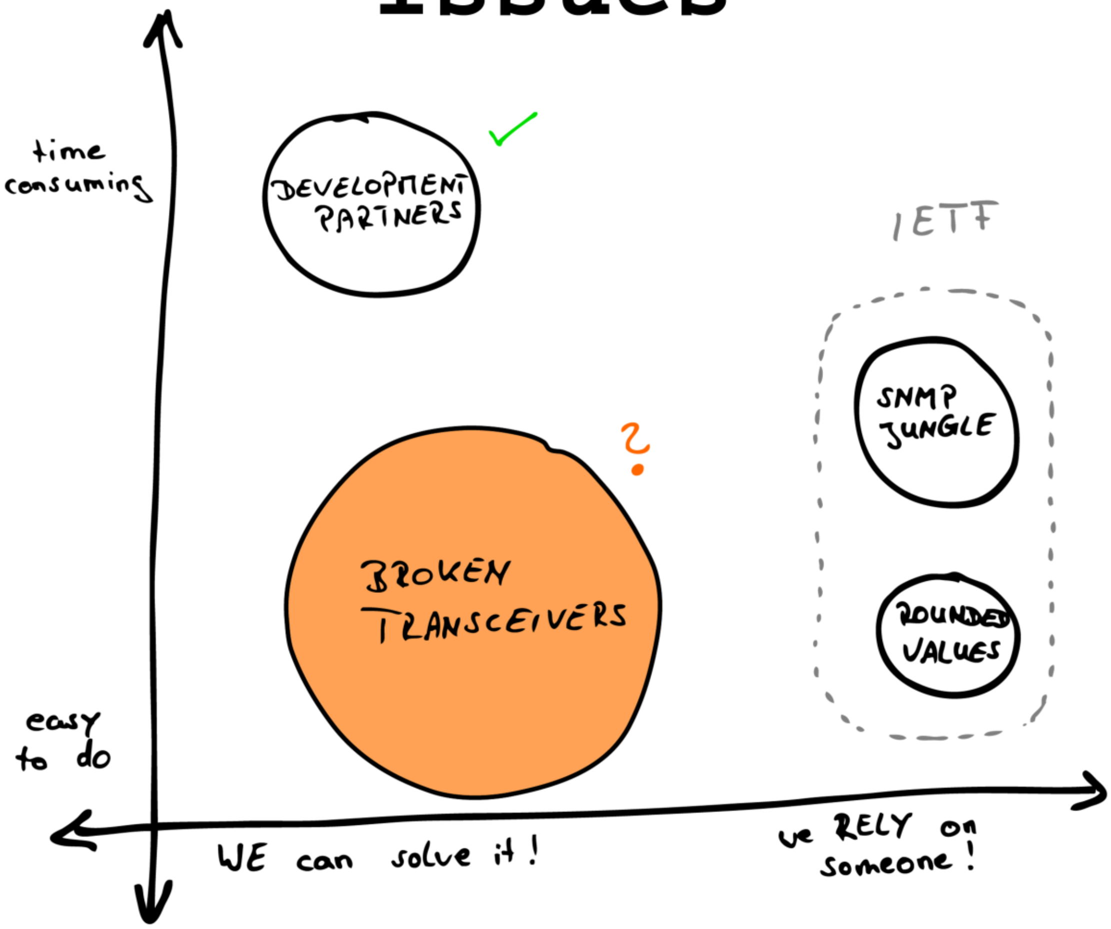
# Metrics



**we need 11 metrics in total!**

- Interface metric (amount of TX und RX bytes) is needed to identify the peer interface. When we know the peer interfaces TX powerlevel value (dBm) we can do a better scoring on the RX powerlevel of the analysed transceiver.
- RX & TX powerlevel, Temperature, Current and Voltage are the transceivers own values. These values will play a main role in the prediction process
- The Serial# of each transceiver helps us to identify the transceiver within the network. An optic might be swapped within its operational lifetime with on other transceiver
- The Article# is very helpful to get parameters like the wavelength or the transceiver‘s supported distance. These values might have an impact on the prediction algorithm (still to be defined)
- CRC errors on the interface are a very good indicator  for a corrupt transmission (either the sending instance mixed it up, the fibre or the receiver). The more optics with CRC erros we can identify for training set the better we can adjust the prediction algorithm later on.
- All 10 metrics will change over time. These differentiated timelines while end up in a transceiver type specific pattern.

# Issues



SNMP jungle & Rounded values: we can‘t take care about this but there is work going on in the IETF MIB that will improve the data.

e.g. Power and Energy Monitoring MIB draft-ietf-eman-energy-monitoring-mib-00 (http://tools.ietf.org/html/draft-ietf-eman-energy-monitoring-mib-00)

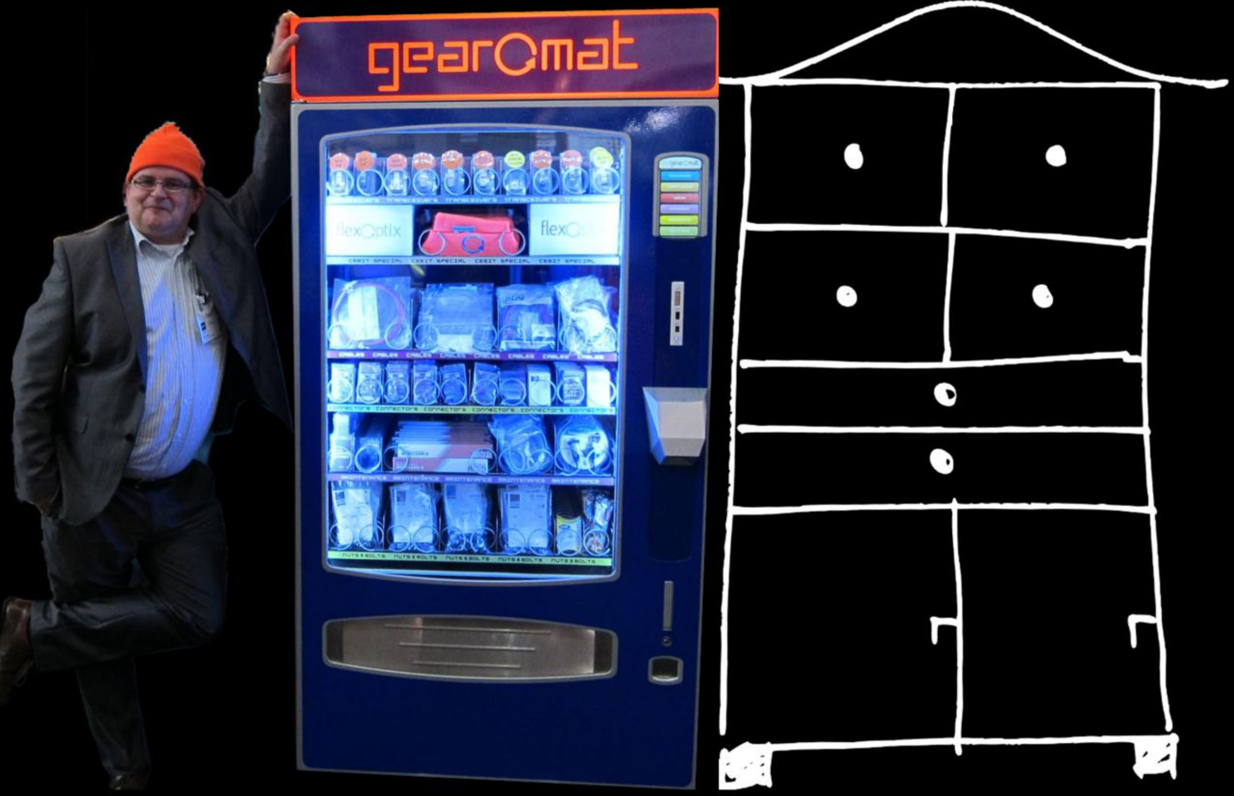
Broken transceivers:this is a task for the ISP community. The more companies join the better the qualitity of the research will be.

Development partners: is almost solved. Currently we are developing a research project with the Knowledge Engineering Group at the University of Darmstadt, which will lead to two funded full time researchers.

# easier

Thesis 4: easier

Gearomat - developed by a bunch of cowboys



Gearomat Prototype as Point of Sales vending machine in the datacenter was presented at Cebit 2011.
After some research we figured out that a „attachment-unit“, build like a cupboard with several drawer, might be of interesst as well. So a technician or the purchasing department can order all kind of stuff and gear and let it shipped directy to the datacenter. When all parts are collected / commissioned the technician can get all parts directly and onsite at the datacenter for its installation. This will save a lot of time and reduces stress because every part can be tracked.

scalable

Thesis 5: more scalable

We want to reuse transceivers because they have a mayor costive influence in these days. This leads to a flexible usage.

DWDM SFP+
tunable
80km fixed

10G BIDI

SFP+ DWDM as tunable will be available in mid 2012 (23dB Powerbudget). Still to be tested because the power consuption seems to be higher than 1,5W.

SFP+ DWDM with 14 or 23dB are tested in running in several setups since a while

SFP+ and XFP BIDI will help you to double your currenty fibre capacity

In combination with a X2 to SFP+ Adapter (soon coming XENPAK to SFP+ Adapter) this is a powerful and scalble setup for existing and soon replaced gear (e.g. Ciscos 6500 which is quite often installed)

# RIPE63 – Foosball Tournament

flexOptix
The Transceiver Specialists.

Thomas

thomas.weible@flexoptix.net
mobil: +49 176 27053733
Adresse: gleich neben Frankfurt

See you at RIPE63 and challenge yourself at the table soccer tournament.

The ideas for transmission 3 arose during a lot of sessions at the customers of Flexoptix GmbH and their data centers.

Pictures where taken & modified by Faheem. Thanks to Autoverwertung Fröhlich.