On Resource Sharing and Careful Overbooking for Network Virtualization

Markus Fiedler

Blekinge Institute of Technology (BTH)

School of Computing (COM)

Karlskrona, Sweden

www.bth.se

BLEKINGE INSTITUTE OF TECHNOLOGY

markus.fiedler@bth.se

ITC-SS 18, Karlskrona, May 2008

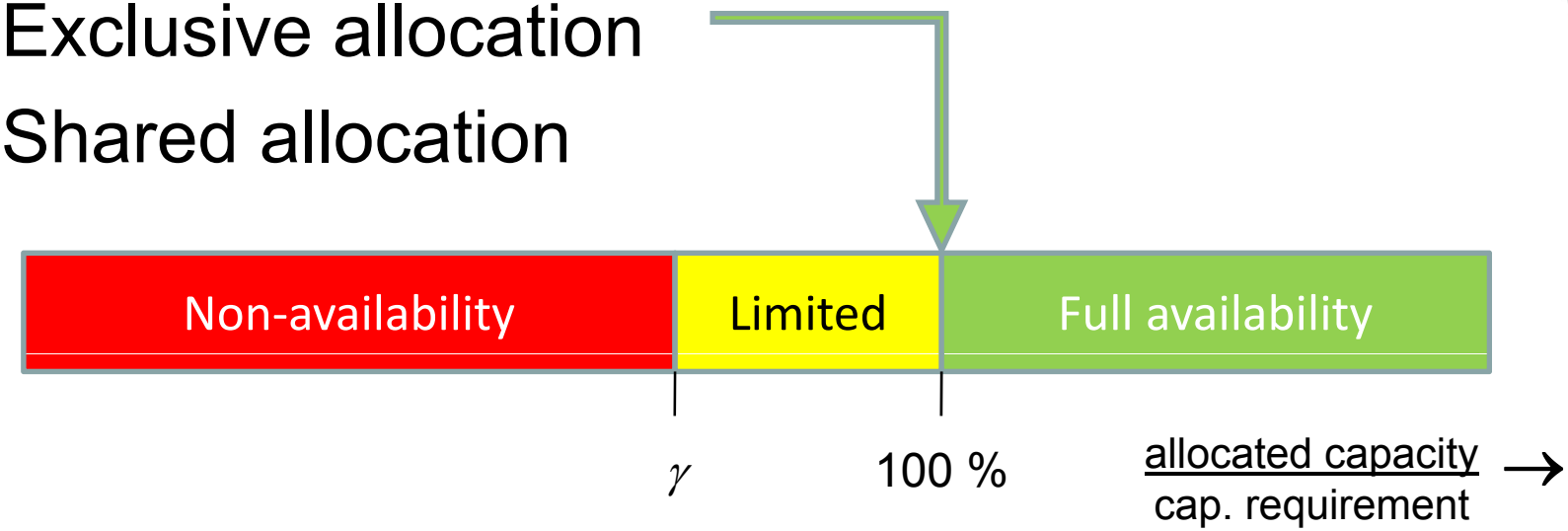
ITC-SS 20, Hoi An, May 2009

Motivation

- Network Virtualization
  - ”… simultaneous operation of multiple logical networks or overlays on a single physical platform” [1]
- Issues of resource sharing and scalability
- Overbooking
  - Integrate extra user without upgrading the capacity – under which circumstances?
  - Two SLA: regarding full and limited availability

Resource Allocation Per User

- Exclusive allocation
- Shared allocation

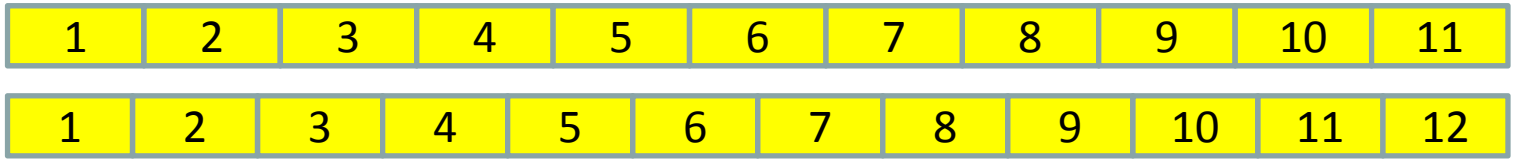


- We are not talking of systems with unlimited capacity ☺

Full vs. Limited Availability

- Assume capacity for 10 exclusive users, γ = 80 %
- **Full availability** (100 %) – desired degree: **1 –** δ

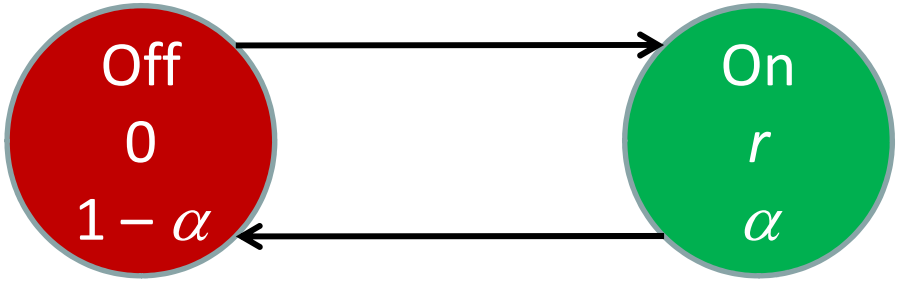
- **Limited availability** (≥ 80 %) – desired degree: **1 –** ε



- **Non-availability** (< 80 %)

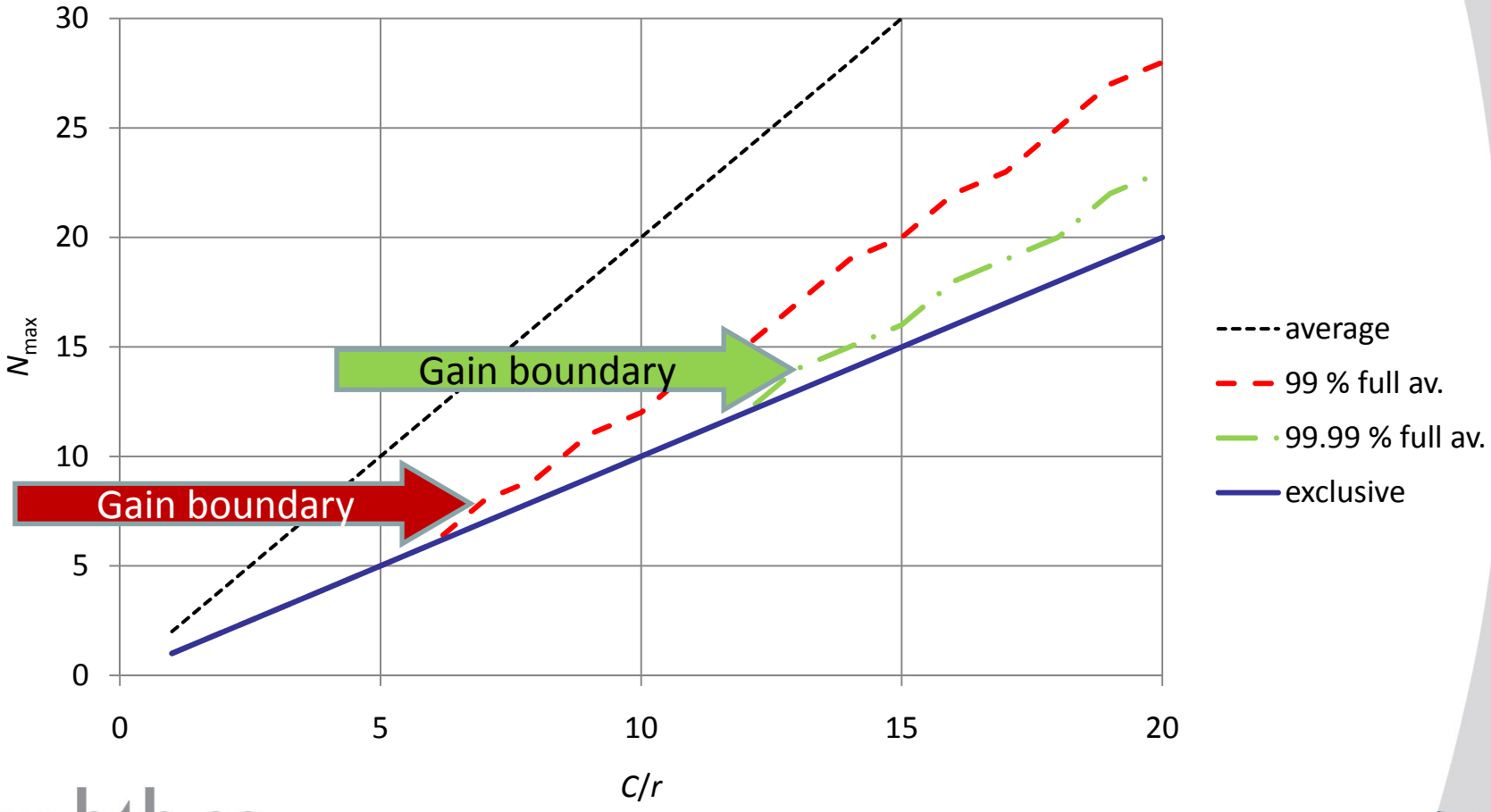
User Model

- On-off
  - peak resource request r
  - activity level α
  - geometrically distributed phases

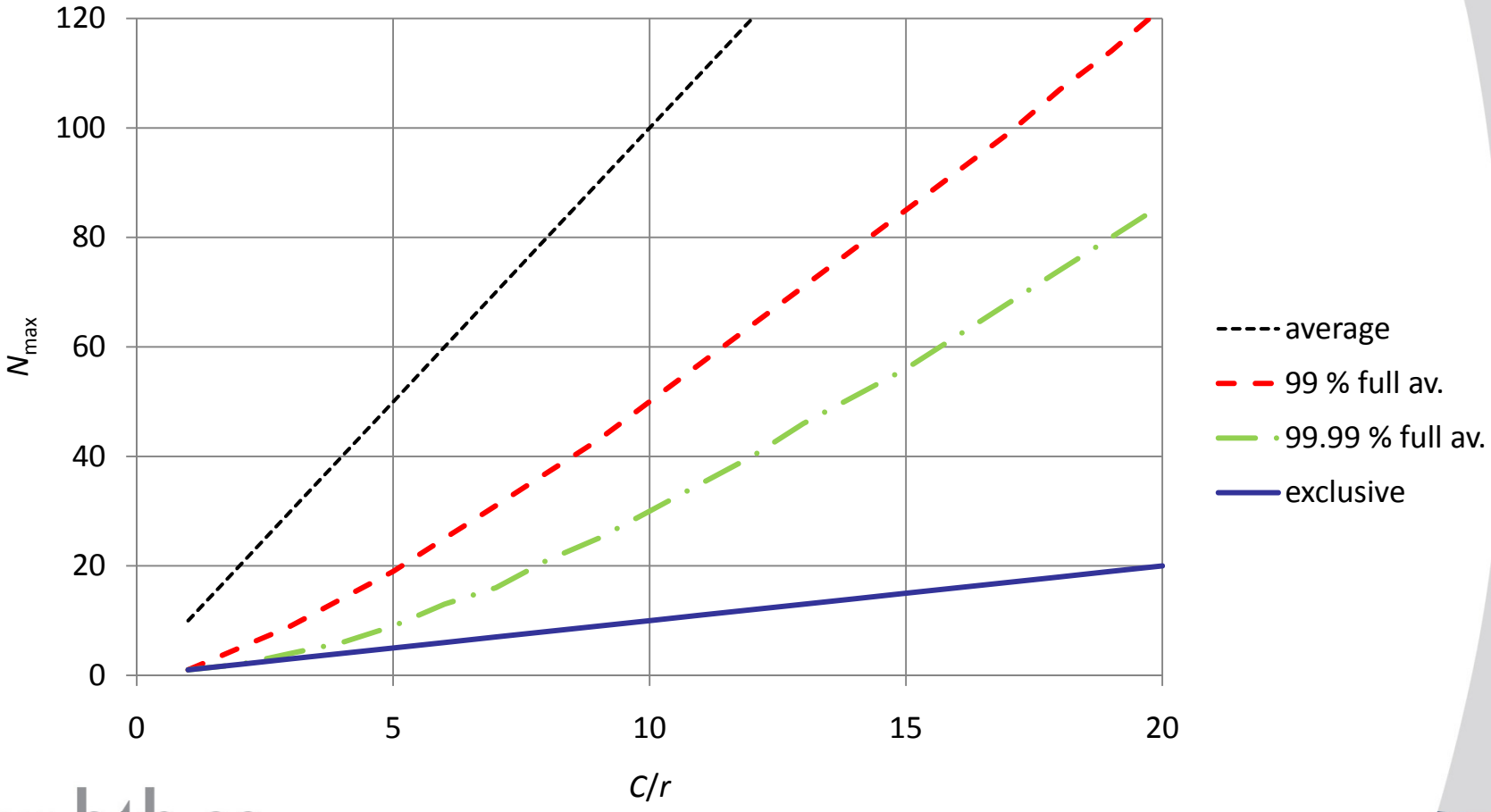


- Independent from each other

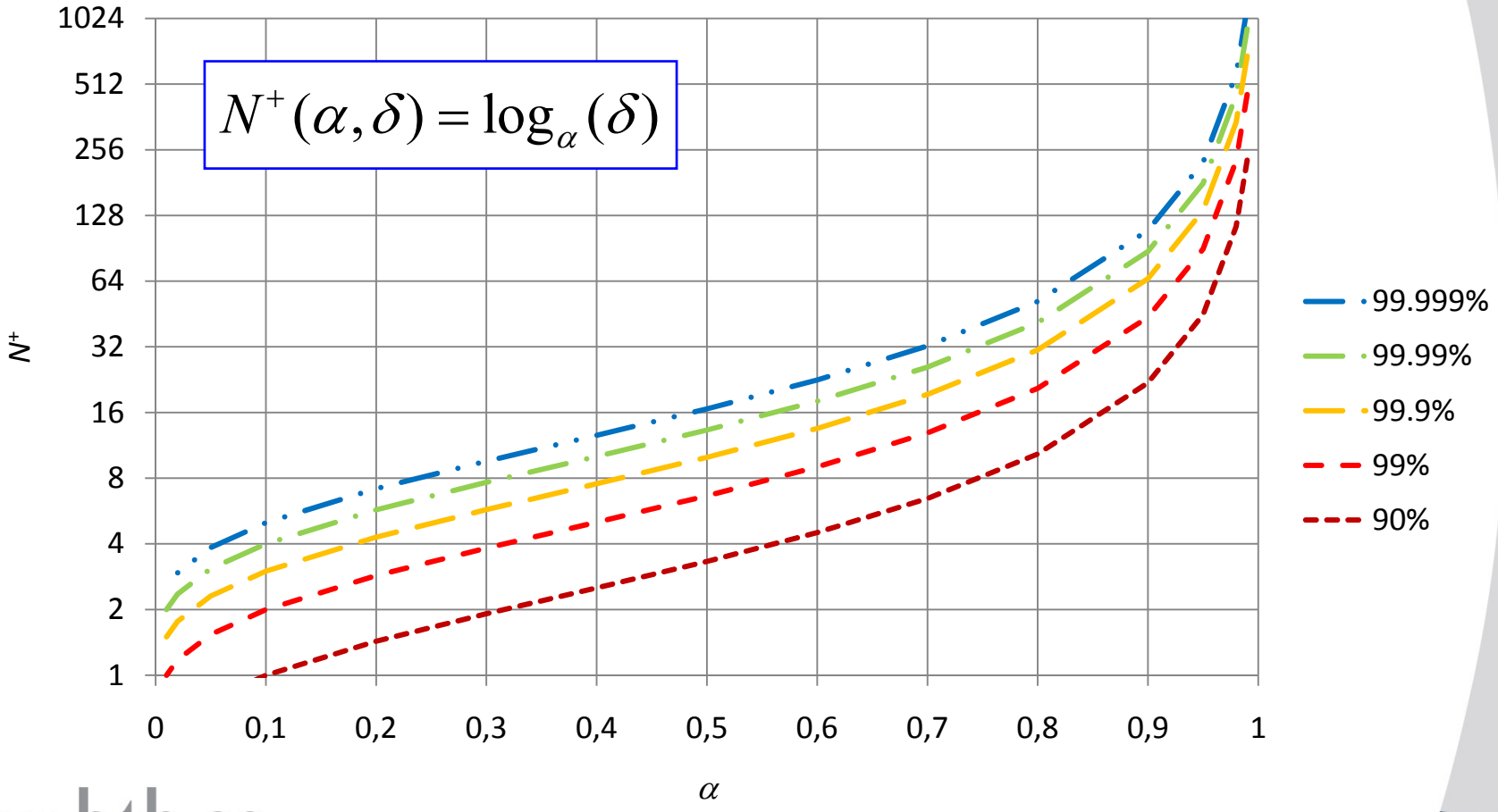
Maximal Number of Users (α = 50 %)



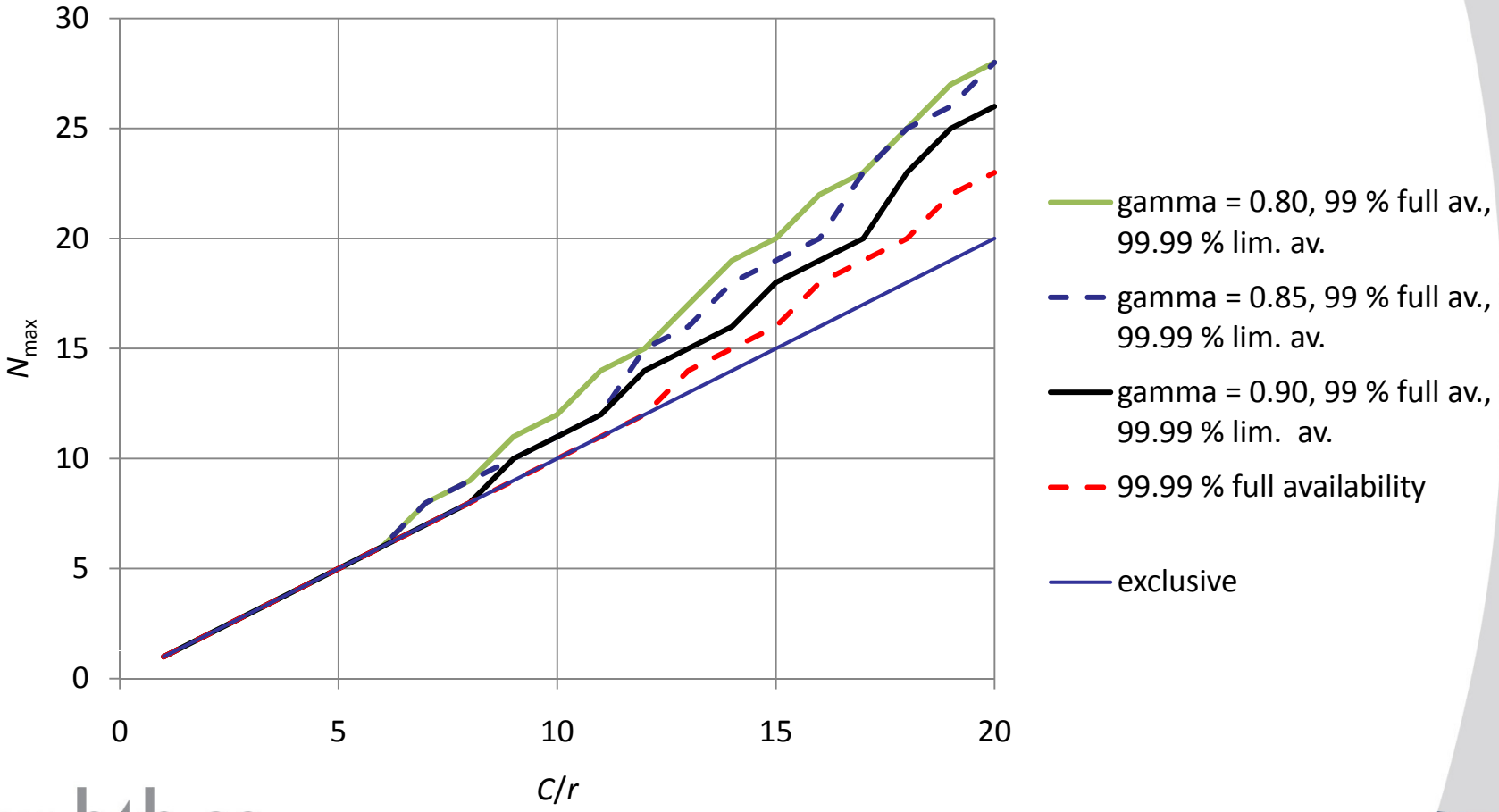
Maximal Number of Users (α = 10 %)



Gain Borderline (One Extra User)



Different Allocation Strategies



Optimal Share γ^* at Gain Borderline

$$\gamma^* = \frac{N^+_{\text{full}}(\alpha,\delta)}{N^+_{\text{lim}}(\alpha,\varepsilon)} = \frac{\lg\delta}{\lg\varepsilon}$$

- Rule-of-thumb: ratio of ”number of nines”
- Examples:

$1-\delta$	$1-\varepsilon$	γ^*
0.99	0.999	2 / 3
	0.9999	1 / 2
0.999	0.9999	3 / 4

Users With Different Activity Factors

- Gain borderline for full availablity:

$$N_2^+(N_1,\alpha_1,\alpha_2,\delta)=\frac{\log\delta}{\log\alpha_2}-N_1\frac{\log\alpha_1}{\log\alpha_2},\begin{cases}0<\alpha_1\le 1\\ 0<\alpha_2<1\end{cases}$$

  - No gain if some activity factor approaches 100 %
  - Best when both activity factors are rather small

Conclusions

- Considered two SLA for full and limited availability
  - There may be room for one additional customer
  - Key: non-permanent activity
  - Some simple formulae available, providing insight in how the capacity sharing works
- Inspiration for this study: overbooking in the context of ATM Virtual Circuits and Paths
  - Problems re-appear, performance models still useful

Outlook

- Measurements in real NV environments
  $\Rightarrow$ Refined user models
- Simulations
- Cross-validations of the findings
- Coordination between resource consumers

Thank you for listening –
welcome to discussing ☺