Teletraffic Analysis
of
Multicast Networks

Samuli Aalto, Jouni Karvo, Eeva Nyberg, Jorma Virtamo
Networking Laboratory
Helsinki University of Technology

samuli.aalto@hut.fi

Outline

- Introduction to traditional teletraffic theory
  - single link analysis
  - Erlang’s formula
- Teletraffic analysis of unicast access networks
  - Convolution-Truncation algorithm
- Teletraffic analysis of multicast networks
  - Multicast setup
  - Convolution-Truncation type algorithm
  - Variations

Unicast =
point-to-point

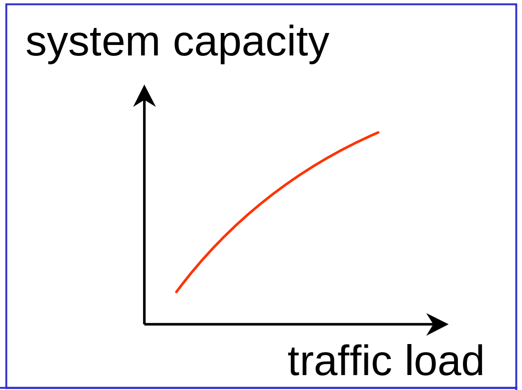
Multicast =
point-to-multipoint

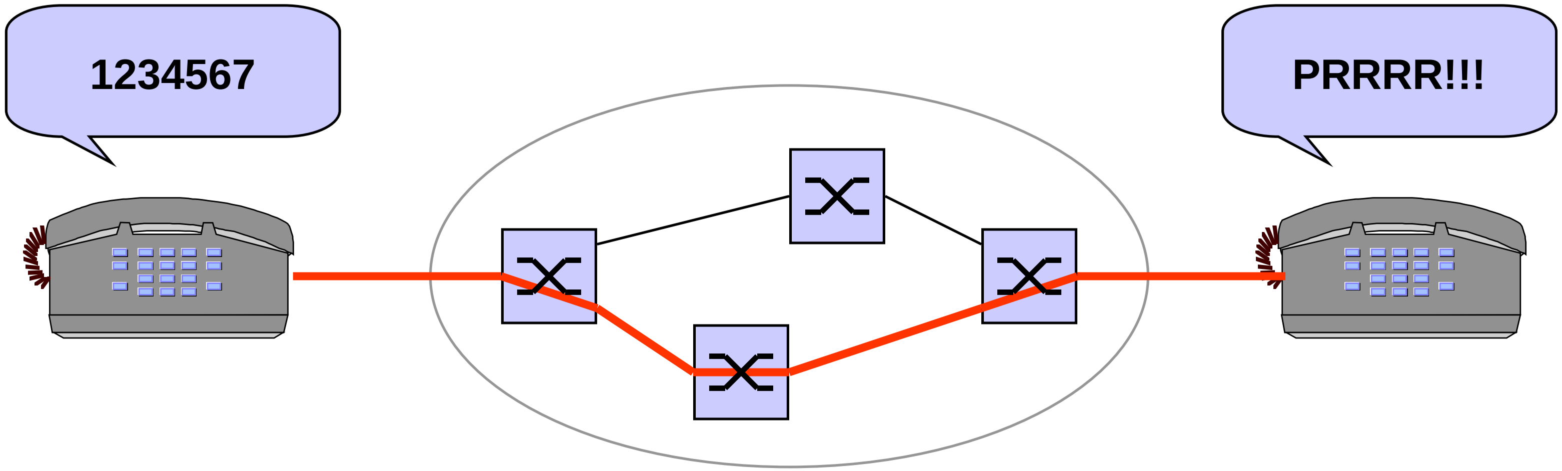
Our contribution

Part I

Introduction to traditional teletraffic theory

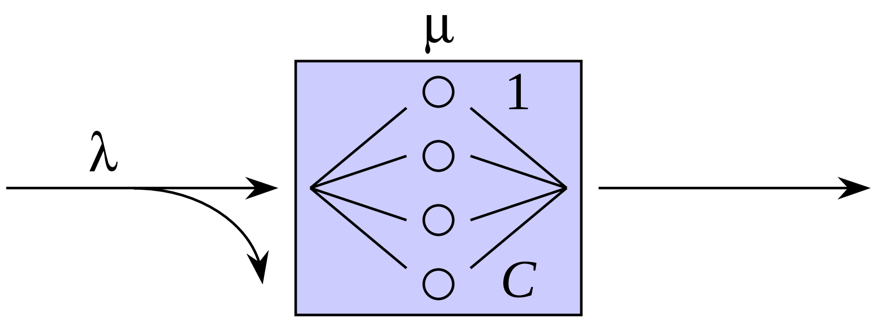
- General purpose: determine **relationships** between
  - quality of service
  - traffic load
  - system capacity
- To describe the relationships quantitatively, **mathematical models** are needed



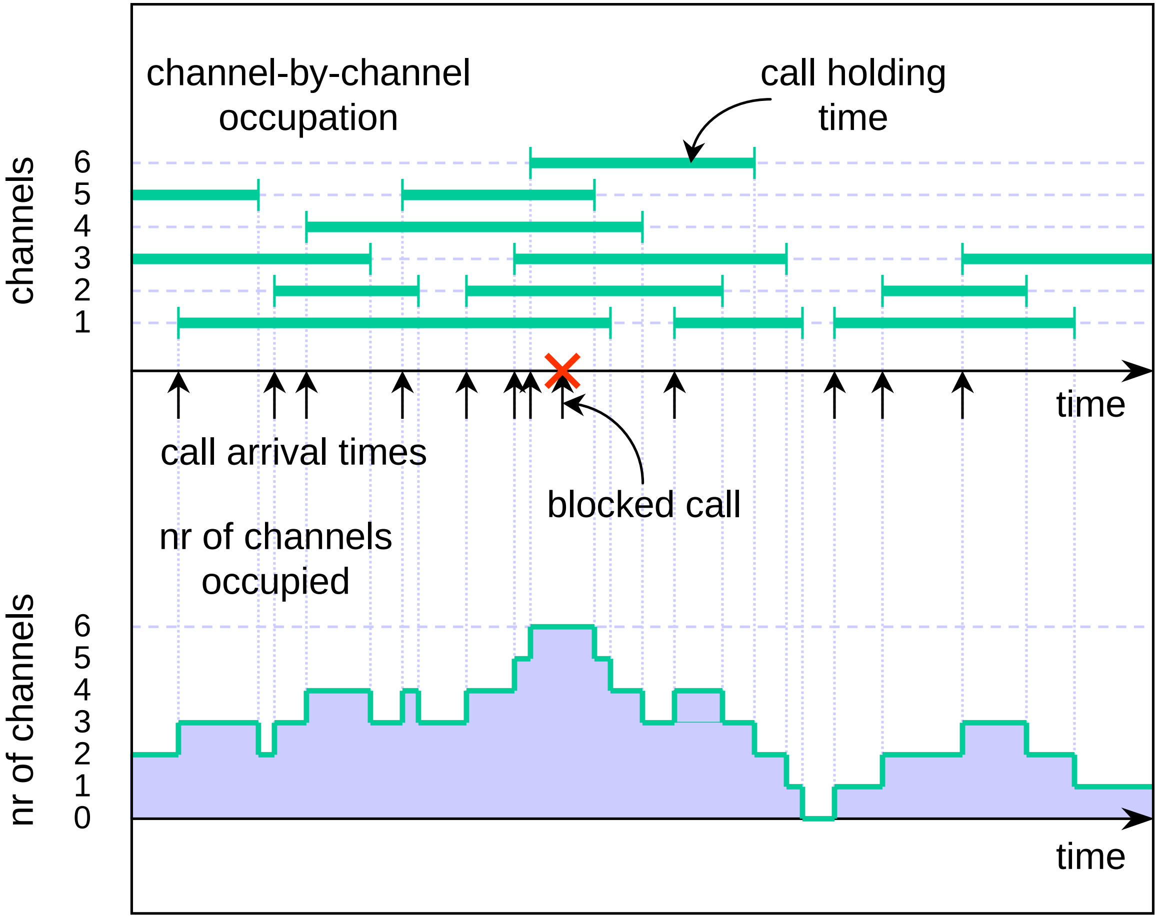


Simple teletraffic model (Erlang's loss model)

- Consider a link between two telephone exchanges
  - there are C **parallel channels** available
  - traffic consists of the ongoing telephone **calls** on the link
  - calls **arrive randomly** at rate λ (calls per time unit)
  - a call occupies one channel in the link for a **random (IID) holding time** with mean $1/\mu$ (time units)
  - blocked calls (arriving in a full system) are lost

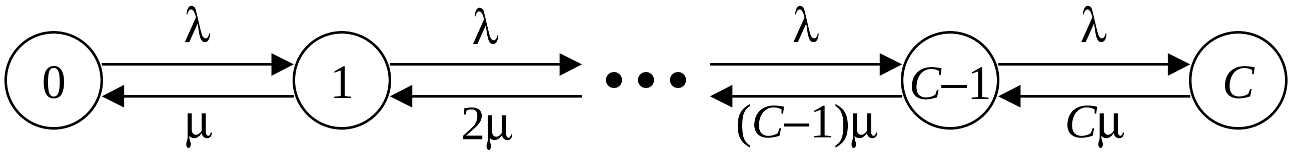


Traffic process



Teletraffic analysis (1)

- $\tilde{X}(t)$ = number of channels occupied at time t
  - under exponential assumptions (Poisson arrivals and exp. holding times), $\tilde{X}(t)$ is a Markov birth-death process (and, thus, reversible)



- Stationary distribution ($a := \lambda/\mu$ = traffic intensity):

$$P\{\tilde{X} = n\} = \frac{\frac{a^n}{n!}}{\sum_{m=0}^{C} \frac{a^m}{m!}} \quad \text{(truncated Poisson distribution)}$$

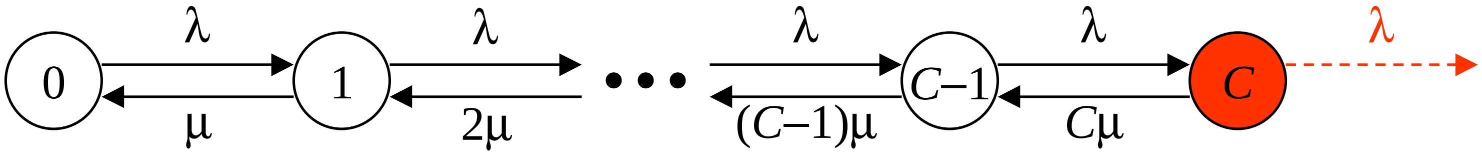
Teletraffic analysis (2)

- **Time blocking** B_t = probability that the system is full

$$B_t := P\{\tilde{X} = C\} = \frac{\frac{a^C}{C!}}{\sum_{m=0}^{C} \frac{a^m}{m!}} \qquad \text{(Erlang's formula)}$$

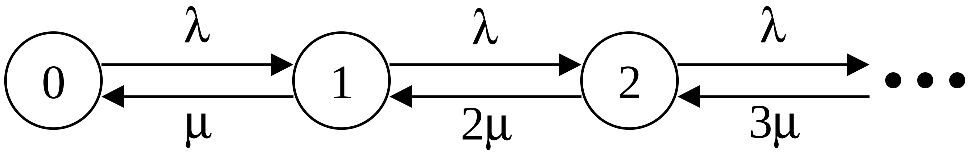
- **Call blocking** B_c = probability that a call is lost

$$B_c := \frac{P\{\tilde{X} = C\}\lambda}{\sum_{n=0}^{C} P\{\tilde{X} = n\}\lambda} = P\{\tilde{X} = C\} = B_t \qquad \text{(PASTA)}$$



Truncation principle

- $X(t)$ = number of channels occupied at time t in a system without capacity constraints ($C = \infty$)
  - under exp. assumptions, $X(t)$ is another Markov birth-death process



- Stationary distribution:

$$P\{X = n\} = \frac{a^n}{n!} e^{-a} \qquad \text{(Poisson distribution)}$$

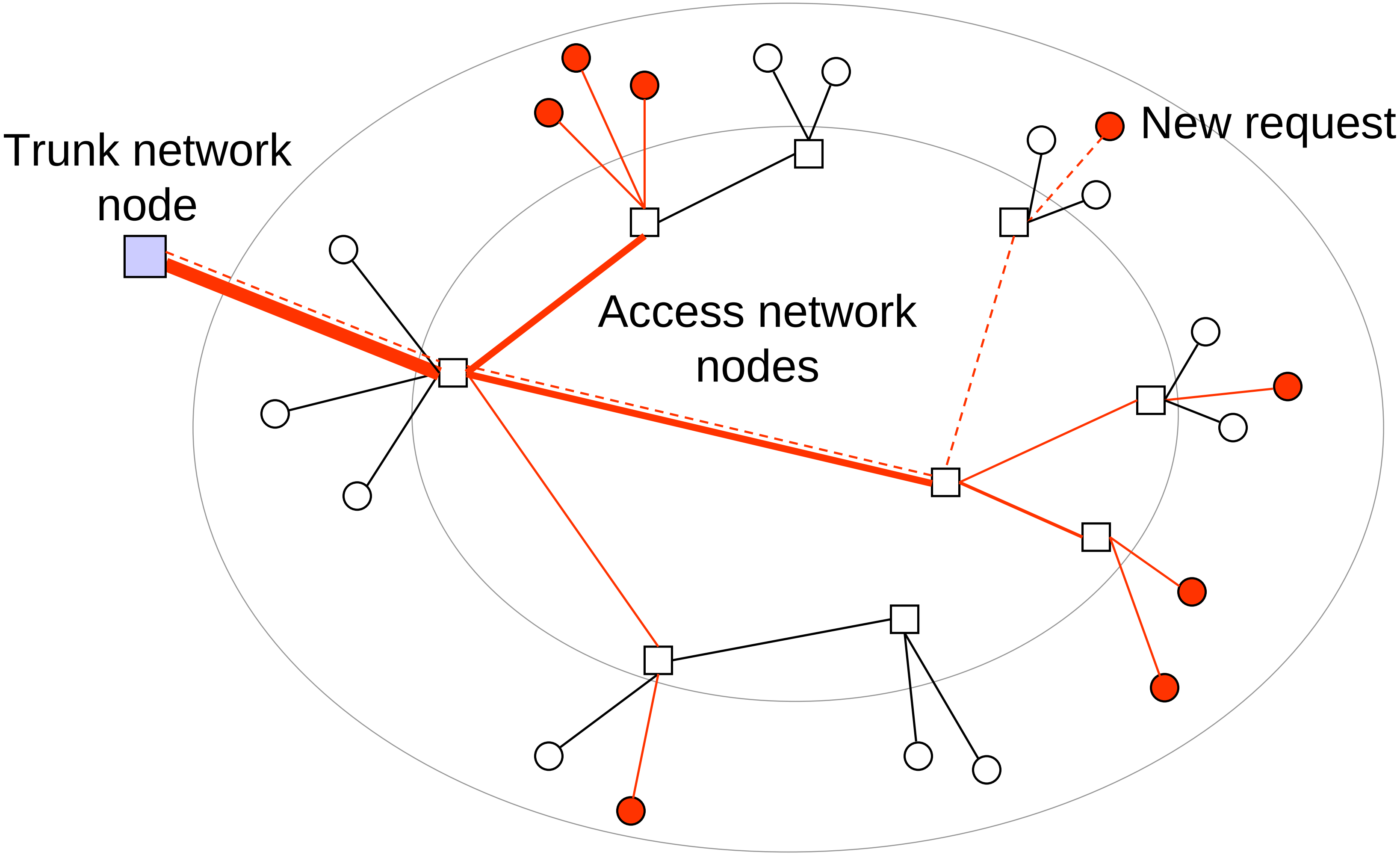
- Thus, **Truncation Principle** applies:

$$P\{\tilde{X} = n\} = \frac{P\{X = n\}}{P\{X \le C\}}, \qquad B_c = B_t = \frac{P\{X = C\}}{P\{X \le C\}} = 1 - \frac{P\{X \le C-1\}}{P\{X \le C\}}$$

Insensitivity

- It is possible to show (e.g., by using the GSMP theory) that the stationary distribution and, thus, the blocking probabilities remain the same even if the holding time distribution is more general than exponential distribution
- So, in this sense, the results are **insensitive** to the holding time distribution

Part II
Teletraffic analysis of unicast access networks

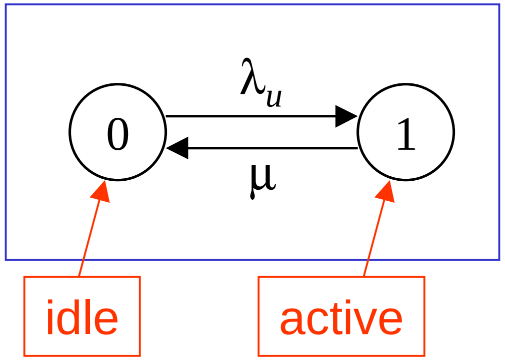


Setup

- Circuit-switched telephone network
- A number of users $u \in U$ physically connected with a unique **trunk network node** by a hierarchial access network with **tree topology**
  - the trunk network node located at the **root node**
  - users located at **leaf nodes**
  - users behave independently
  - unicast connection requests (between the trunk network node and users) arrive randomly
  - random (IID) connection holding times
  - required capacity per link per ongoing unicast connection = 1 unit
- Physical links $j \in J$ with finite capacities C_j

Teletraffic analysis (1)

- Consider first a network **without capacity constraints**
- $Y_u(t)$ = state of user u at time t
  - $Y_u(t) \in \{0,1\}$
  - under exponential assumptions, $Y_u(t)$ is a Markov birth-death process (and, thus, reversible)
  - stationary distribution ($a_u := \lambda_u/\mu$):

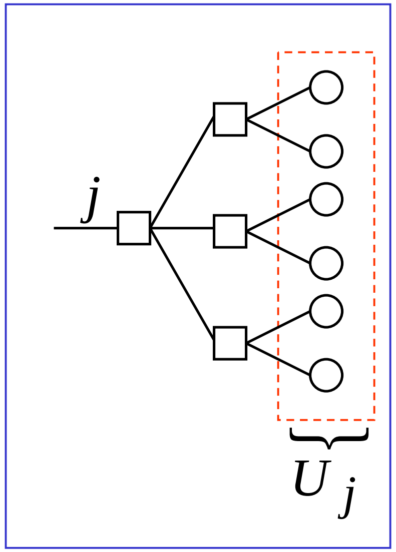


$$\pi_u(y) := P\{Y_u = y\} = \begin{cases} \dfrac{1}{1+a_u}, & y = 0 \\ \dfrac{a_u}{1+a_u}, & y = 1 \end{cases}$$

Teletraffic analysis (2)

- $Y_j(t)$ = state of link j at time t
  - $Y_j(t) \in \{0,1,\ldots,|U_j|\}$

$$Y_j(t) = \sum_{u \in U_j} Y_u(t)$$



- $X(t) = (Y_u(t);\, u \in U)$ = network state at time t
  - $X(t)$ is also a reversible Markov jump process
  - stationary distribution (due to independent users):

$$P\{X = x\} = \prod_{u \in U} P\{Y_u = y_u\} = \prod_{u \in U} \pi_u(y_u)$$

- Thus, a closed form analytical expression exists!

Teletraffic analysis (3)

- $\mathbf{X}$ = network state (**without** capacity constraints)
- $\tilde{\mathbf{X}}$ = network state (**with** capacity constraints)
- $\tilde{\Omega}$ = network state space (**with** capacity constraints)
- $\tilde{\Omega}_u$ = nonblocking states for user u
- B^t_u = time blocking probability for user u

Due to the Truncation Principle!

numerator

$$B_u^t := 1 - P\{\tilde{\mathbf{X}} \in \tilde{\Omega}_u\} = 1 - \frac{P\{\mathbf{X} \in \tilde{\Omega}_u\}}{P\{\mathbf{X} \in \tilde{\Omega}\}}$$

denominator

- Remark: It can be shown that this result is **insensitive** to the holding time distribution, as well as to the idle time distribution

Teletraffic analysis (4)

- In principle, there is a closed form analytical expression both for the numerator and the denominator

- **Problem**: computationally complex
  - worst case: exponential in U (since $|\Omega| = 2^U$)

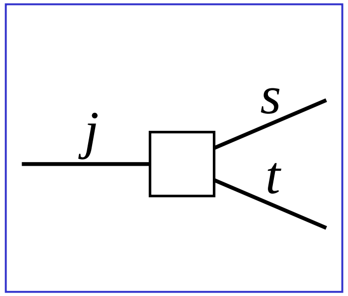
- **Solution**: use a **recursive convolution-truncation algorithm** to calculate the numerator and the denominator
  - always: linear in U

- Remark: Assuming exponential idle times, it can be shown that call blocking (for user u) equals time blocking (for user u) in a modified system, where user u is always idle

Recursive algorithm (1)

- **Convolution**:

$$[f \otimes g](n) := \sum_{m=0}^{n} f(m) g(n-m)$$



- **Key result**:
  - If link *j* has two downstream neighbouring links (*s*,*t*), then

$$P\{Y_j = n\} = \sum_{m=0}^{n} P\{Y_s = m\} P\{Y_t = n-m\}$$

  - In other words,

$$\pi_j(n) = [\pi_s \otimes \pi_t](n)$$

Recursive algorithm (2)

$$P\{\mathbf{X}\in\tilde{\Omega}\}=P\{Y_j\le C_j,\ j\in J\}$$

denominator

numerator

$$P\{\mathbf{X}\in\tilde{\Omega}_u\}=P\{Y_j+1\le C_j,\ j\in R_u;\ Y_j\le C_j,\ j\in J\setminus R_u\}$$

- Q**-functions**:

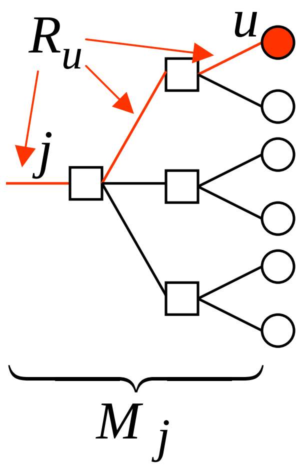
$$Q_j(n):=P\{Y_j=n;\ Y_k\le C_k, k\in M_j\}$$

$$\begin{aligned}Q_j^u(n):=P\{Y_j=n;\ Y_k+1\le C_k, k\in M_j\cap R_u;\\ Y_k\le C_k,\quad k\in M_j\setminus R_u\}\end{aligned}$$

- **Truncations**:

$$T_j f(n):=f(n)1\{n\le C_j\}$$

$$T_j' f(n):=f(n)1\{n+1\le C_j\}$$



Recursive algorithm (3)

- Denominator:

link state

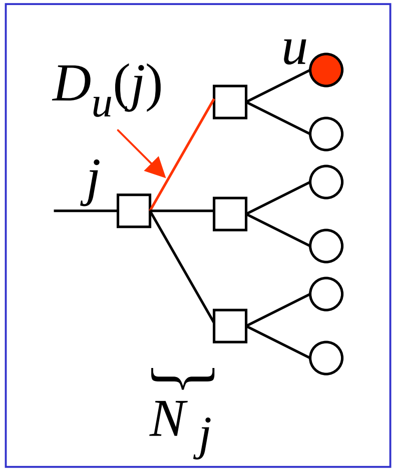
$$P\{\mathbf{X}\in\tilde{\Omega}\}=\sum_{n} Q_J(n)$$

truncation

root link

$$Q_j(n)=\begin{cases} T_j[\pi_j](n), & j\in U\\ T_j[\bigotimes_{k\in N_j} Q_k](n), & j\notin U \end{cases}$$

convolution



- Numerator:

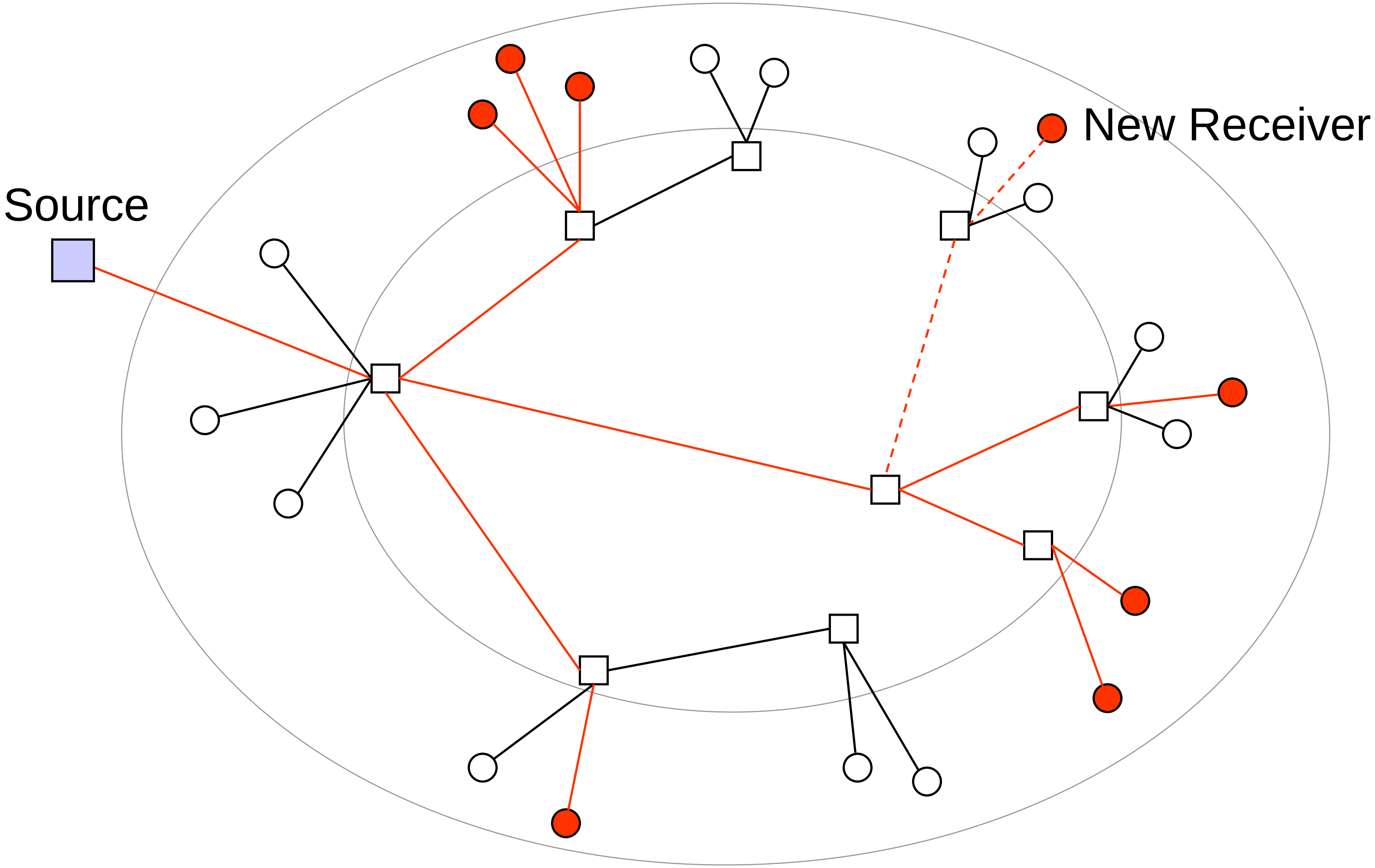
$$P\{\mathbf{X}\in\tilde{\Omega}_u\}=\sum_{n} Q_J^u(n)$$

mod. truncation

route to user *u*

$$Q_j^u(n)=\begin{cases} T_j^{'}[\pi_j](n), & j=u\\ T_j^{'}[Q^u_{D_u(j)}\otimes(\bigotimes_{k\in N_j\setminus R_u} Q_k)](n), & j\in R_u\setminus\{u\} \end{cases}$$

Part III
Teletraffic analysis of multicast networks

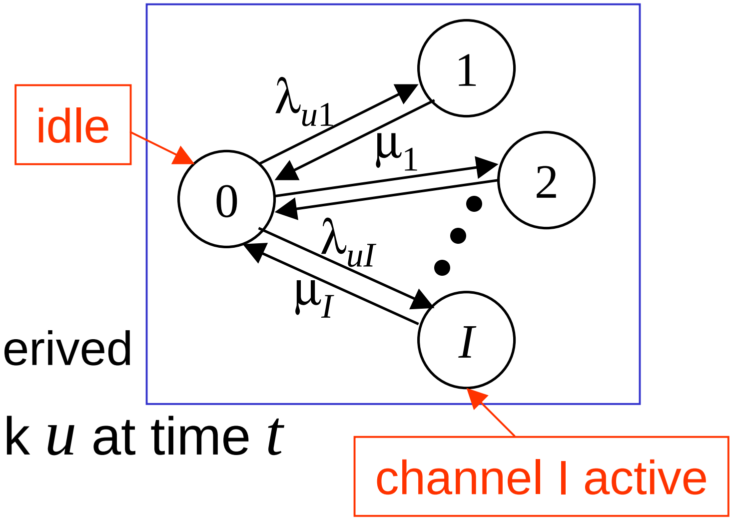


Setup

- Circuit-sw. network, or packet-sw. with strict quality guarantees
- A unique **source** offers a variety of **channels** $i \in I$
  - e.g. audio or video streams
  - required capacity per link per active channel $i = d_i$ units
- Each channel is delivered to users $u \in U$ by a **multicast connection** with **dynamic membership**
  - users behave independently
  - connection requests by users to join channels arrive randomly
  - random (IID) connection holding times
- Each multicast connection uses the same **routing tree**
  - the source located at the **root node**
  - users located at **leaf nodes**
- Physical links $j \in J$ with finite capacities C_j

Teletraffic analysis (1)

- Consider first a network **without capacity constraints**
- $Z_u(t)$ = state of user u at time t
  - $Z_u(t) \in \{0,1,\ldots,I\}$
  - under exponential assumptions, $Z_u(t)$ is a Markov birth-death process
  - stationary distribution $P\{Z_u = i\}$ easily derived
- $\mathbf{Y}_u(t) = (Y_{ui}(t);\, i \in I)$ = state of leaf link u at time t
  - $\mathbf{Y}_u(t) \in \{0,1\}^I$
  - $\mathbf{Y}_u(t)$ is a reversible Markov jump process



idle

channel I active

$$\pi_u(\mathbf{y}) := P\{Y_u = \mathbf{y}\} = \begin{cases} P\{Z_u = 0\}, & \text{if } \mathbf{y} = \mathbf{0} \\ P\{Z_u = i\}, & \text{if } \mathbf{y} = \mathbf{e}_i \\ 0, & \text{otherwise} \end{cases}$$

(0,0,…,0)

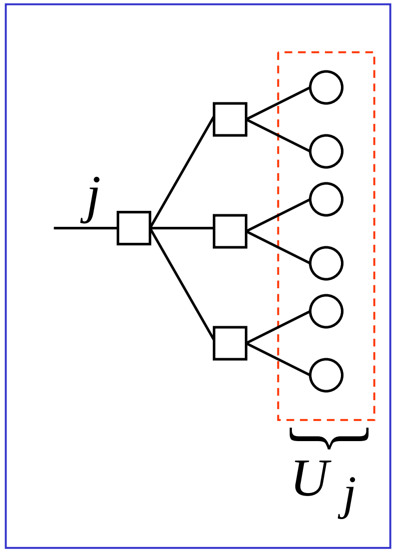
(0, …,1,…,0)

Teletraffic analysis (2)

- $\mathbf{Y}_j(t) = (Y_{ji}(t);\ i \in I)$ = state of link j at time t
  - $\mathbf{Y}_j(t) \in \{0,1\}^I$

$$\mathbf{Y}_j(t) = \bigoplus_{u \in U_j} \mathbf{Y}_u(t)$$

componentwise OR



- $\mathbf{X}(t) = (\mathbf{Y}_u(t);\ u \in U)$ = network state at time t
  - $\mathbf{X}(t)$ is also a reversible Markov jump process
  - stationary distribution (due to independent users):

$$P\{\mathbf{X} = \mathbf{x}\} = \prod_{u \in U} P\{\mathbf{Y}_u = \mathbf{y}_u\} = \prod_{u \in U} \pi_u(\mathbf{y}_u)$$

- Thus, a closed form analytical expression exists!

Teletraffic analysis (3)

- $\mathbf{X}$ = network state (**without** capacity constraints)
- $\tilde{\mathbf{X}}$ = network state (**with** capacity constraints)
- $\tilde{\Omega}$ = network state space (**with** capacity constraints)
- $\tilde{\Omega}_{ui}$ = nonblocking states for user u and channel i
- B^t_{ui} = time blocking probability for user u and channel i

Due to the Truncation Principle!

$$B^t_{ui} := 1 - P\{\tilde{\mathbf{X}} \in \tilde{\Omega}_{ui}\} = 1 - \frac{P\{\mathbf{X} \in \tilde{\Omega}_{ui}\}}{P\{\mathbf{X} \in \tilde{\Omega}\}}$$

numerator

denominator

- Remark: It can be shown that this result is **insensitive** to the holding time distribution, as well as to the idle time distribution

Teletraffic analysis (4)

- In principle, there is a closed form analytical expression both for the numerator and the denominator

- **Problem**: computationally complex
  - worst case: exponential in U (since $|\Omega| = 2^{UI}$)

- **Solution**: use a **recursive convolution-truncation type algorithm** to calculate the numerator and the denominator
  - always: linear in U

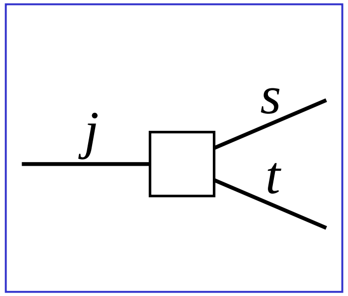
- Remark: Assuming exponential idle times, it can be shown that call blocking (for user u) equals time blocking (for user u) in a modified system, where user u is always idle

Recursive algorithm (1)

- **OR-convolution**:

$$[f \otimes g](\mathbf{y}) := \sum_{\mathbf{u},\mathbf{v}:\mathbf{u}\oplus\mathbf{v}=\mathbf{y}} f(\mathbf{u})g(\mathbf{v})$$

componentwise OR



- **Key result**:
  - If link *j* has two downstream neighbouring links (*s*,*t*), then

$$P\{\mathbf{Y}_j = \mathbf{y}\} = \sum_{\mathbf{u},\mathbf{v}:\mathbf{u}\oplus\mathbf{v}=\mathbf{y}} P\{\mathbf{Y}_s = \mathbf{u}\}P\{\mathbf{Y}_t = \mathbf{v}\}$$

  - In other words,

$$\pi_j(\mathbf{y}) = [\pi_s \otimes \pi_t](\mathbf{y})$$

Recursive algorithm (2)

$$P\{\mathbf{X}\in\tilde{\Omega}\} = P\{\mathbf{d}\cdot\mathbf{Y}_j \le C_j,\ j\in J\}$$

denominator

$$P\{\mathbf{X}\in\tilde{\Omega}_{ui}\} = P\{\mathbf{d}\cdot(\mathbf{Y}_j\oplus\mathbf{e}_i)\le C_j,\ j\in R_u;\mathbf{d}\cdot\mathbf{Y}_j\le C_j,\ j\in J\setminus R_u\}$$

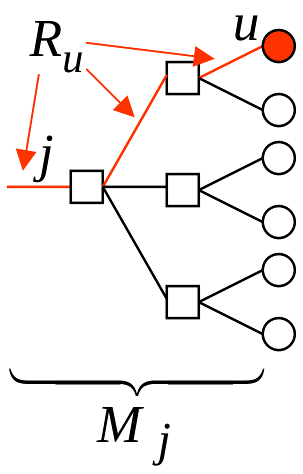
numerator

componentwise OR

- Q-functions:

$$Q_j(\mathbf{y}) := P\{\mathbf{Y}_j=\mathbf{y};\mathbf{d}\cdot\mathbf{Y}_k\le C_k, k\in M_j\}$$

$$\begin{aligned} Q_j^{ui}(\mathbf{y}) := P\{\mathbf{Y}_j=\mathbf{y};\mathbf{d}\cdot(\mathbf{Y}_k\oplus\mathbf{e}_i)\le C_k, &\ k\in M_j\cap R_u \\ \mathbf{d}\cdot\mathbf{Y}_k\le C_k, &\ k\in M_j\setminus R_u\} \end{aligned}$$



- **Truncations**:

$$T_j f(\mathbf{y}) := f(\mathbf{y})1\{\mathbf{d}\cdot\mathbf{y}\le C_j\}$$

$$T_j^i f(\mathbf{y}) := f(\mathbf{y})1\{\mathbf{d}\cdot(\mathbf{y}\oplus\mathbf{e}_i)\le C_j\}$$

Recursive algorithm (3)

- Denominator:

link state

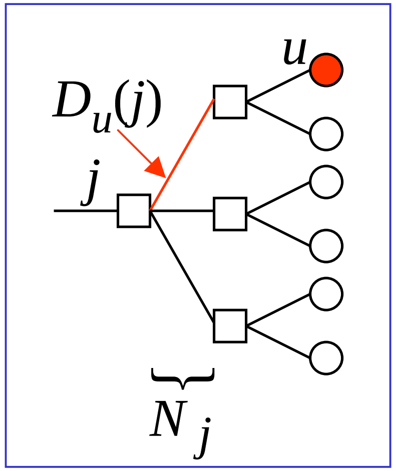
$$P\{\mathbf{X}\in\tilde{\Omega}\} = \sum_{\mathbf{y}} Q_J(\mathbf{y})$$

truncation

root link

$$Q_j(\mathbf{y}) = \begin{cases} T_j[\pi_j](\mathbf{y}), & j\in U \\ T_j[\bigotimes_{k\in N_j} Q_k](\mathbf{y}), & j\notin U \end{cases}$$

OR-convolution



- Numerator:

$$P\{\mathbf{X}\in\tilde{\Omega}_{ui}\} = \sum_{\mathbf{y}} Q_J^{ui}(\mathbf{y})$$

mod. truncation

route to user *u*

$$Q_j^{ui}(\mathbf{y}) = \begin{cases} T_j^i[\pi_j](\mathbf{y}), & j=u \\ T_j^i[Q^{ui}_{D_u(j)} \otimes (\bigotimes_{k\in N_j\setminus R_u} Q_k)](\mathbf{y}), & j\in R_u\setminus\{u\} \end{cases}$$

Variations

- Single link analysis (Karvo, Virtamo, Martikainen & Aalto, 1997-1998)
  - starting point
- Network wide analysis (Nyberg, Virtamo & Aalto, 1999)
  - all channels handled individually (as in this presentation)
  - first convolution-truncation type algorithm
  - **or-convolution** needed
  - “background” unicast traffic possible to be taken into account by modifing the truncation operators
- Multi-class case (Aalto, Karvo & Virtamo, 2000)
  - class = group of statistically indistinguishable channels
  - **combinatorial convolution** needed (instead of or-convolution)
- Multi-layer case (Karvo, Aalto & Virtamo, 2000-2001)
  - layered coding of audio/video streams
  - **max-convolution** needed (instead of or-convolution)

THE END