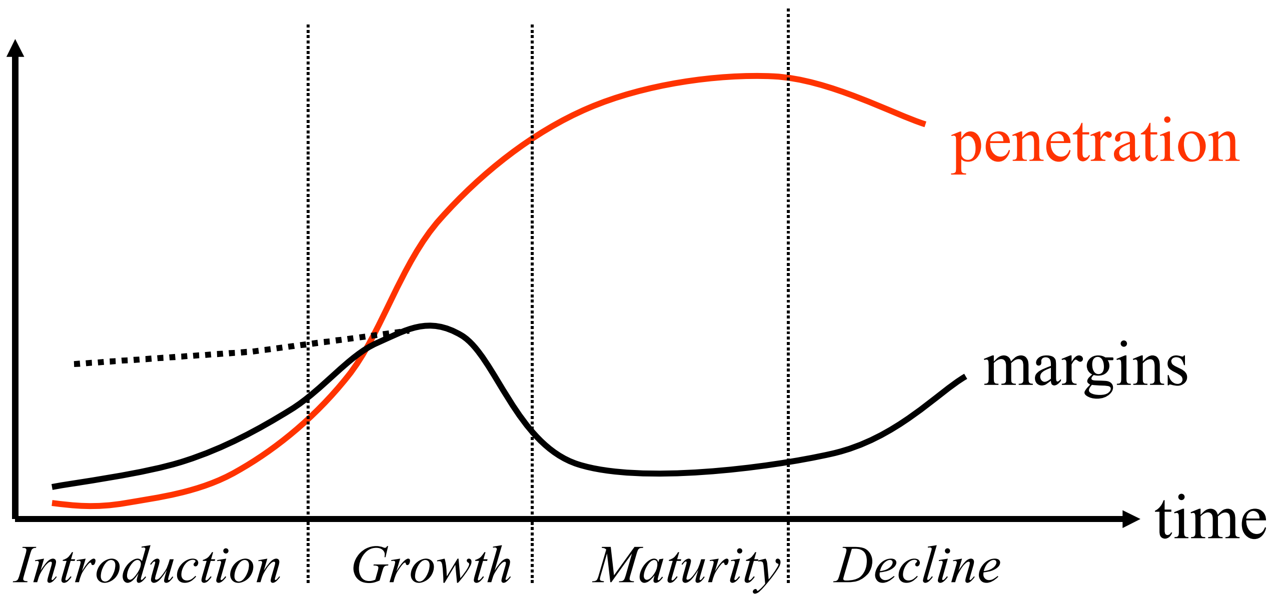
Pricing – part 2

S-38.041 Networking Business

Service life cycle phases

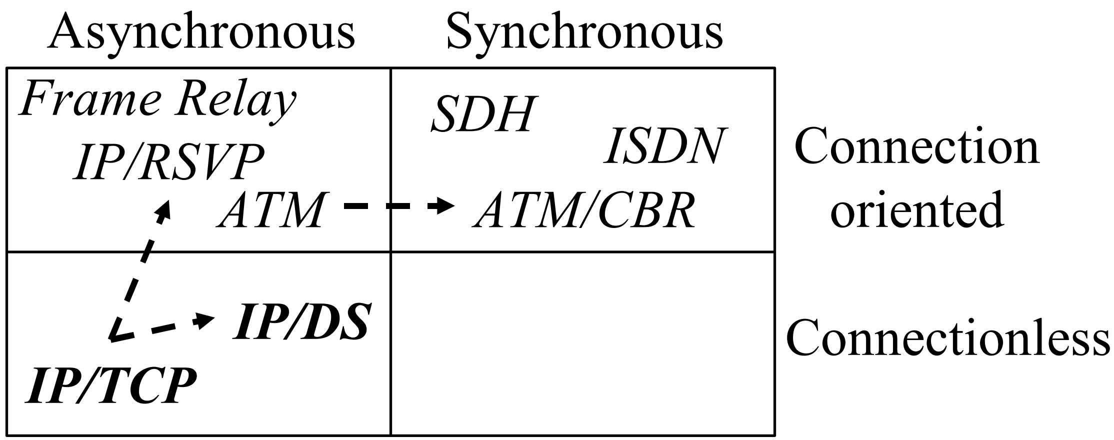
Impact on pricing



- <u>Introduction</u>: early adopters, skimming vs. aggressive growth
- <u>Growth</u>: increasing demand, little competition, high margins
- <u>Maturity</u>: differentiation pressure, tough competition, low margins
- <u>Decline</u>: cost cutting, harvesting niche segments, high margins

Backbone services

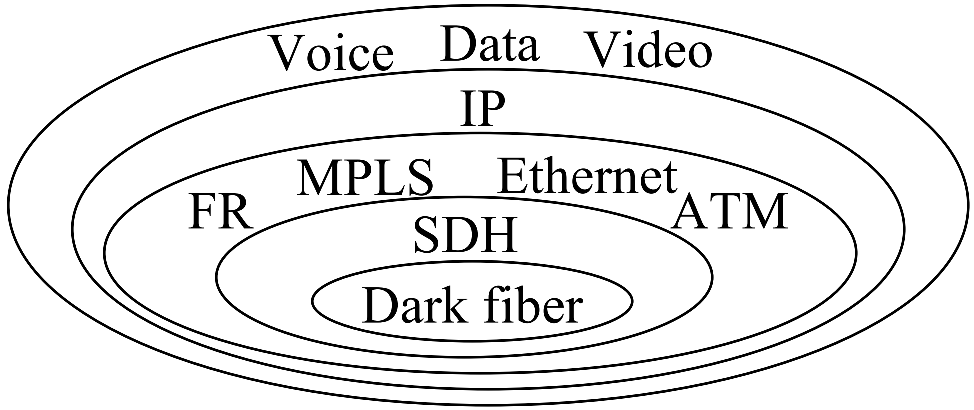
Impact of IP



- Growth of IP traffic involves evolution
  - from inelastic to elastic applications (e.g. video streaming inelastic → elastic)
  - from guaranteed services to best-effort (the fundamental nature of IP is best-effort)
  - from deterministic to statistical multiplexing (ref. *effective bandwidth*)
  - from bottleneck control to over-dimensioning
  - from layer 2 VPN to layer 3 IP VPN
- Key issue: demand vs. supply of backbone capacity?

Backbone services

Wholesale of capacity between pre-defined similar end-points



- Customers are other operators or individual firms
- Portfolio of services
  - point-to-point vs. multipoint
  - basic (dark fiber) vs. value-added (managed IP router service)
  - voice vs. data vs. video
- ATM being gradually replaced by Ethernet and MPLS
- Pricing based on Service Level Agreements (SLA) and traffic parameters (peak rate, mean rate, data loss probability, max delay, mean delay, etc)

Backbone services

Service Level Agreement (SLA)

- **Service level agreement**: a documented result of a negotiation between a customer and a provider of a service that specifies the levels of availability, performance, operation and other attributes of the service

- **Static SLA management**: SLA contract is made between two human parties and its terms cannot be changed without human intervention

- **Dynamic SLA management**: SLAs are negotiated and contracted automatically using some signaling procedures

- **SLA trading**: dynamic SLA management where information on service provisioning, routing, and pricing are exchanged between providers

Backbone services

SLA evolution scenario

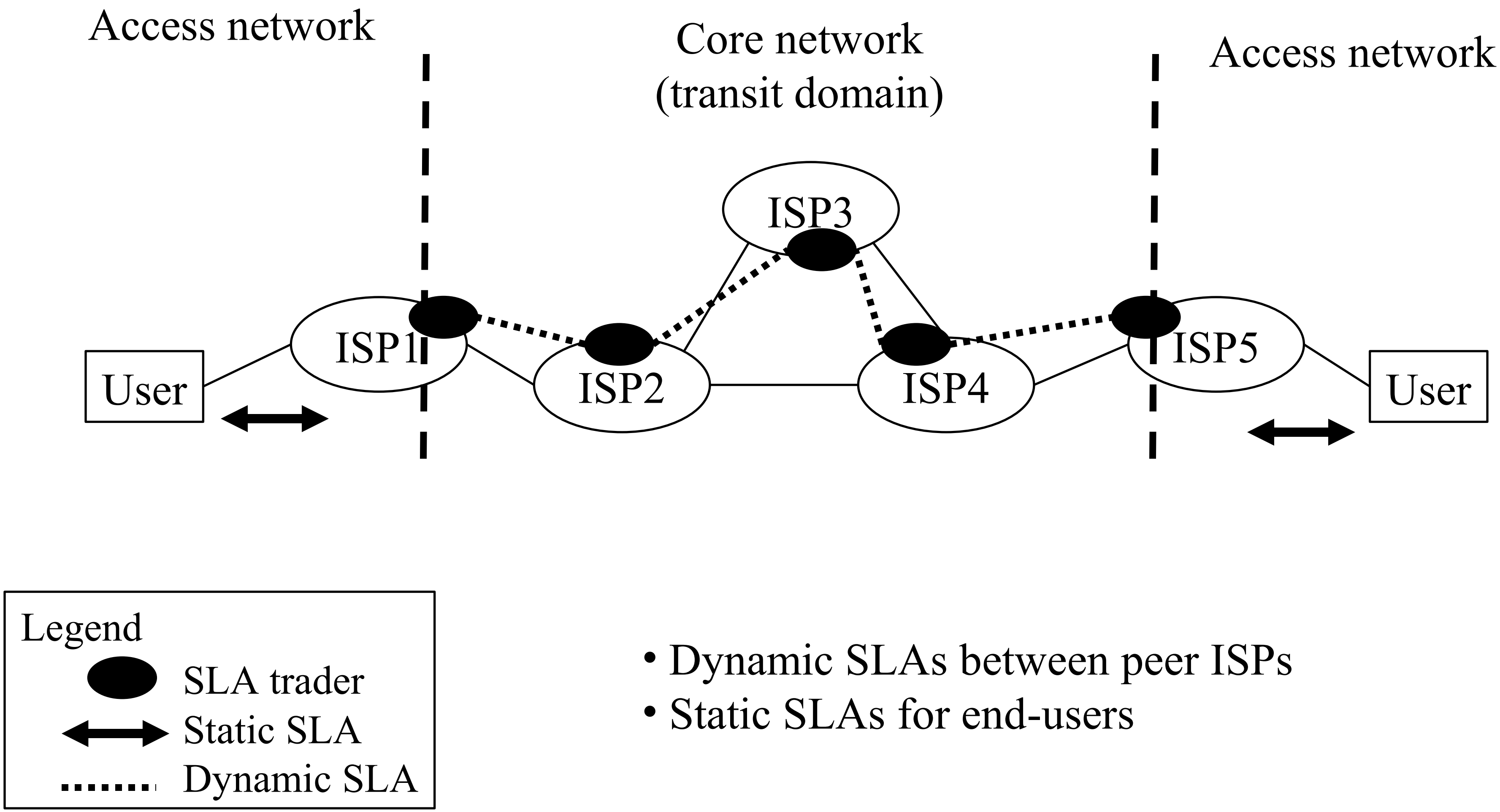
1. Static SLA management in telecom networks and dedicated data networks
2. Static SLA management in IP-based best effort networks
3. Static SLA management in IP diffserv (DS) networks ?
4. Dynamic SLA management in IP DS networks ?

DS has the following SLA characterictics

- Large *traffic aggregates* (as opposed to ATM SVC)
- Typical traffic aggregates are VoIP, WWW, specific routes
- Aggregates appear as *Traffic Conditioning Agreements* (TCA)
- Traffic flows through DS domains (via *ingress/egress nodes*)
- Standardized *Per-Hop-Behaviors* (PHB) for e2e pricing?
  - *Expedited Forwarding* (EF)
  - *Assured Forwarding* (AF)

Backbone services

SLA traders



- Dynamic SLAs between peer ISPs
- Static SLAs for end-users

Backbone services

Summary of SLA trading

- SLA trading has not been tested in real deployments
- SLA trading suits best for large networks and ISPs
- Transition from static to dynamic SLA trading is a major management challenge
- Based on simulation results, SLA trading can improve network utilization by up to 40% compared to a traditional, shortest-path routed inter-domain network
- The residual bandwidth pricing strategy is a suitable candidate for SLA trading since it ensures that prices increase with SLA or link load

Internet access services

Congestion control

- The end-to-end bottleneck may occur at different points
  - In dedicated access:
    - Increase the dedicated per subscriber access speed (e.g. ADSL)
    - Push bandwidth sharing closer to subscribers (e.g. HomePNA)
  - In shared access/backbone/server: apply *congestion control*
- The level of congestion needs to be optimized
  - Too much congestion $\Rightarrow$ *negative network externality*
  - Too little congestion $\Rightarrow$ waste of network capacity
- Options for congestion control
  - Over-dimensioning (wasting of network capacity)
  - Call admission control, e.g. RSVP blocking (latest customers suffer)
  - Automatic flow control, e.g. TCP (all customers suffer)
  - Human fairness control, e.g. HomePNA (local group discipline)
  - *Congestion pricing* (maximal social surplus?)

Internet access services

Congestion pricing - theory

- Congestion price is two-part: normal + externality, $p+p_E$
  - Social surplus maximization
    (1) $max\ \sum_j u_j\ (x_j,y)-c(k)$, where $y=\sum_i x_i\ /k$, k=total fixed capacity
    $\Rightarrow p_E=-(1/k)\sum_j du_j(\underline{x}_j,y)/dy$, where $\underline{x}_j$=socially optimal demand
  - Individual maximization of surplus for consumer i
    (2) $max[u_i(x_i,y)-p_E x_i] \Rightarrow x_i=\underline{x}_i$, if number of users is large
  - Social and individual optima are the same, Nash equilibrium!
  - Congestion price converges to optimal price via tatonnement: network determines p_E using step (1) and publishes it, then each consumer i solves step (2) to find $\underline{x}_i$, and so on
  - u_j are unknown $\Rightarrow$ network must vary p_E until finding equilibrium
  - y is unknown to consumers $\Rightarrow$ consumers estimate it via congestion
- Congestion pricing suits for expensive bottlenecks like radio
- Congestion pricing facilitates automatic optimal capacity planning via the customer feedback loop

Internet access services

Congestion pricing - practice

- Time-of-day pricing (e.g. fixed-price tickets in Internet Café)
- Pricing per application & traffic type
  - Types pre-defined using diffserv, e.g. www, VoIP, etc
  - Automatic traffic classification and resource re-allocation
- Pricing per user’s willingness-to-pay
  - Price-driven separation of service classes (e.g. Paris Metro Pricing)
  - Priority service classes based on relative quality (e.g. via diffserv)

- Note that flat-rate pricing well reflects the operator’s large share of fixed cost, but cannot efficiently tackle the problem of temporary congestion!

Congestion pricing

Example: Time-of-day pricing

- Assume utility for consumer i : $u_i(x_1^i, x_2^i)$, in which the x variables imply the amount of usage for peak-hour and off-peak-hour, respectively.
- By denoting capacity limits for both peak-hour (t=1) and off-peak-hour (t=2) periods with C, we end up with a maximization problem:

$$\max_{x_1^i, x_2^i} \sum_{i=1}^{N} u_1(x_1^i, x_2^i) \quad s.t. \sum_{i=1}^{N} x_t^i \leq C_t \quad t = 1,2$$

- This leads into a Lagrangian optimization problem (from the perspective of social planner), in which we can now denote the Lagrangian constants with symbols p (for comfort in the interpretation):

$$L = \sum_{i=1}^{N} u_1(x_1^i, x_2^i) - p_1\left(\sum_{i=1}^{N} x_1^i - C_1\right) - p_2\left(\sum_{i=1}^{N} x_2^i - C_2\right)$$

- It is easy to see that based on the first-order conditions of the Lagrangian formula above, we end up with the same solution (given the price vector) by solving the consumer's problem for each i,

$$\max_{x_1^i, x_2^i} u_i(x_1^i, x_2^i) - p_1 x_1^i - p_2 x_2^i \quad \forall i$$

only if we can balance the Lagrangian constants (=prices) so that the capacity is in full use (i.e. Kuhn-Tucker conditions).

- This requires tatonnement (slight adjustments of price so that the consumption and capacity are balanced). Note that if e.g. the peak-hour utility is higher on average, in equilibrium its price must be higher, too.

Content services

Private vs. public goods

<u>Private good</u> (e.g. candy bar)

- You consume one, there is one less for others - *depletetable*
- If consumed – no one else can - *excludable*
- Marginal cost > 0
- Price = marginal cost. Achieved on ideal market when supply = demand

<u>Public good</u> (e.g. radio broadcast)

- *Nondepletable* – when used by one, the same amount is available to others.
- *Nonexcludable* – Use by one does not exclude others from using the good.
- Marginal cost ≈ 0
- Price ≈ 0 → fixed cost is not recovered → taxation, non-usage based fees

Content services

Evolution examples

- Best-effort IP service: Initially public good → Flat monthly fee → Congestion → Private good externality.
- Telephone call: In PSTN and over radio interface = private good (”candy bar”) → price/unit.
- Value-added IP service, e.g. VoIP: Initially usage fee. CPU and memory getting cheaper (Moore’s law) → Marginal cost of new customer ≈ 0 → Flat-rate.
- Digital Content: Marginal cost ≈ 0 → Copyright and IPR control enable both private and public goods. Copyright violations, e.g. peer-to-peer traffic → development of digital rights management (DRM) or bundling with other private goods!

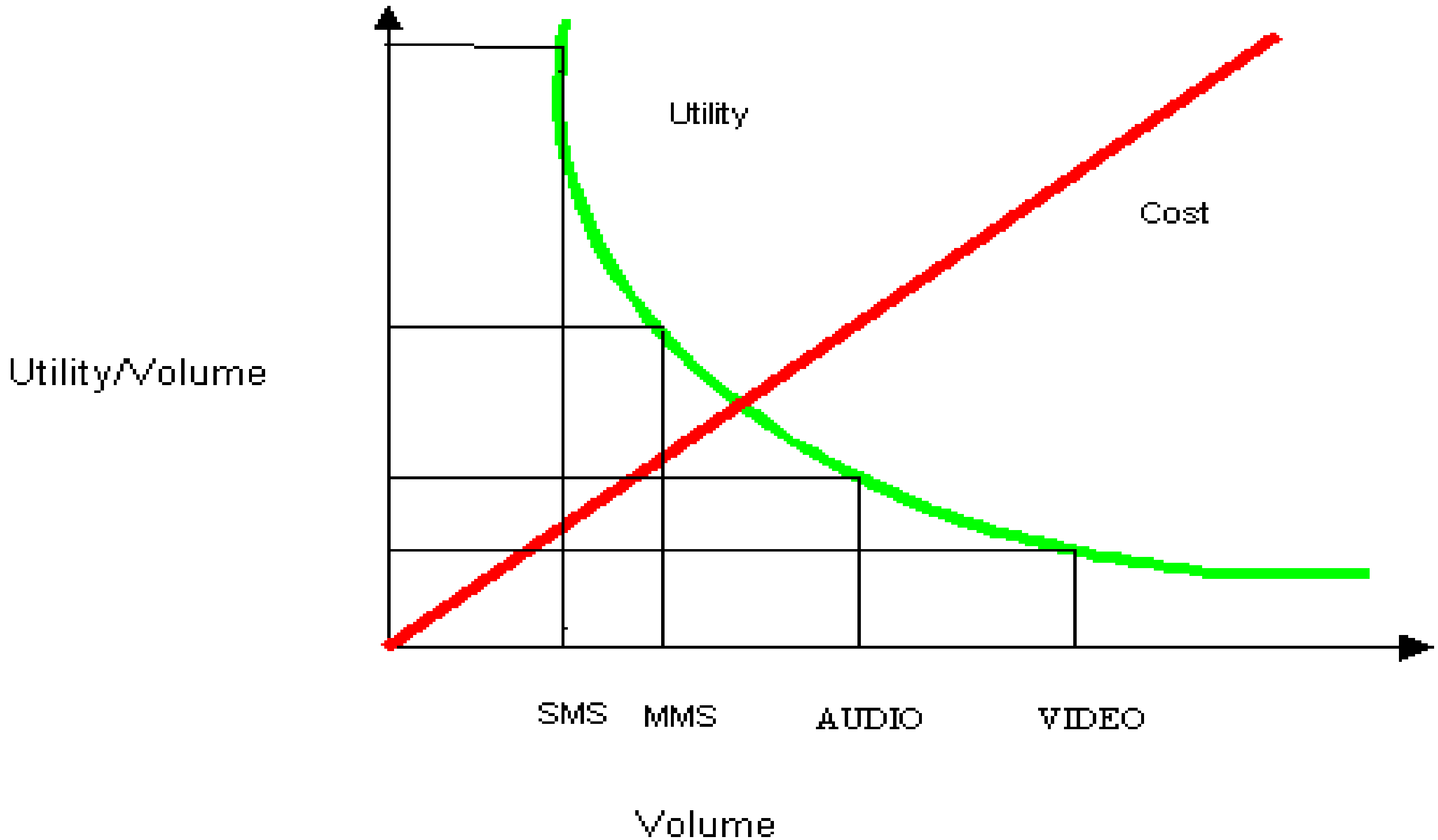
Service bundling

Vertical vs. horizontal bundling in GSM

- Vertical bundling
  - Bundling of access with content
  - For instance weather report over SMS
- Horizontal bundling
  - Bundling of access services (e.g. multiple radios, circuit vs. packet-switched, voice vs. data)
  - Bundling of vertically bundled services (e.g. weather report over SMS vs. WAP)
- Bundling enables
  - Cross-subsidies and service differentiation
  - Value-based pricing, i.e. flexible testing of subscriber's willingness-to-pay

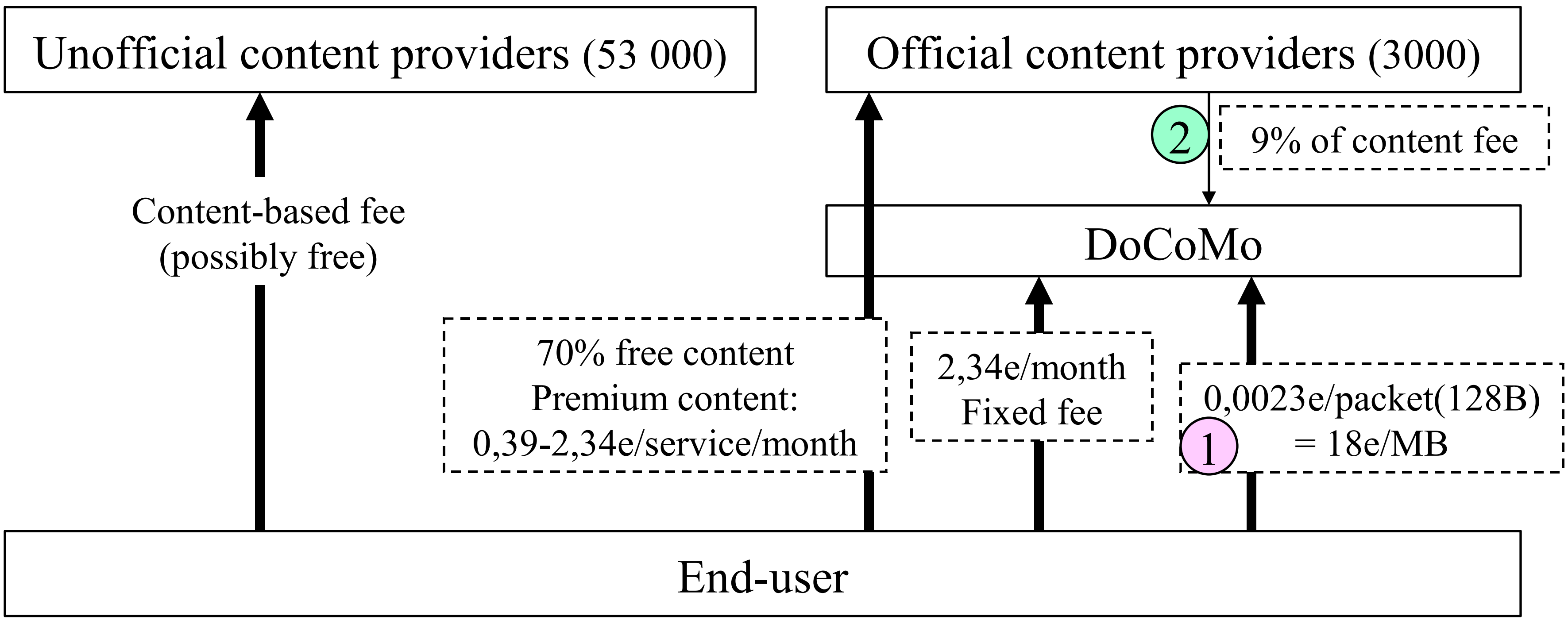
Service bundling

Roll-out of new services



- Cross-subsidies enable early roll-out of still non-profitable services
- Operator can also take risk of new handsets via handset subsidies

Case: DoCoMo i-mode pricing



1. Accounts for 87% of the i-mode ARPU
2. Accounts for less than 1% of the I-mode ARPU

Source: Sandro Grech, 2003 (prices 2002)

Pricing of telecom equipment

- Traditionally pricing is based on hardware capacity (e.g. switching centers, routers, base stations), which hides software R&D costs → pressure to price software
- Capacity pricing is adapted per type of capacity
  - GSM MSC switching capacity (number of simultaneous calls)
  - GSM HLR storage capacity (number of subscribers)
  - GSM BTS radio transmission capacity (number of TRXs)
  - IP router capacity (bits/sec, packets/sec, number of ports, etc)
  - Server transaction capacity (SMS/sec, locations/sec, etc)
- Growing exploitation of general purpose operating systems and hardware (e.g. Unix) in network elements is likely to gradually un-bundle the pricing of software and hardware