The Selection Problem in Distributed Computing Systems with Application to Internet Environments

Dmitry Korzun

Department of Computer Science, Petrozavodsk State University

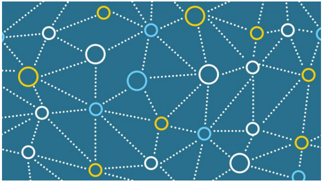
The study is supported by the Ministry of Education and Science of Russia
within project # 2.5124.2017/8.9 of the basic part of state research assignment for 2017–2019.

SMARTY 2018: Stochastic Modeling and Applied Research of TechnologY
September 21 - 25, 2018, Karelian Research Center of RAS
Petrozavodsk, Republic of Karelia, Russia

Individual Strategy in a Distributed System

Components are subject to autonomous and cooperative activity.

- Parallelism
- Synchronization
- Concurrency
- Rational behavior



http://texasurj.com/wp/a-decentralized-internet/

Individual strategy synchronized with the observable situation.

Component (agent) i makes own decision d on time t using strategy f_i:

$$d_t = f_i(t, s, \alpha)$$

based on situation s and individual parameters α,

Rationality (smartness): $\sum_t C(d_t) \to \min$ for cost metric $C(d)$

Selection Problem in Internet System (1/2)

- Selection of the best candidate among several ones

$$f : \left\{ d^{(1)}, d^{(2)}, \dots, d^{(N)} \right\} \to d^*$$

E.g., *multipath routing* when selecting a network path to send the next data packet

- Selection of repeated events in the sequence

$$f : \left\{ \dots, e_{i-1}, \underbrace{e_i, e_{i+1}, \dots, e_{i+n}}_{n+1 \text{ equal elements}}, e_{i+n+1}, \dots \right\} \to (i, n)^*$$

E.g., *data interleaving* when no need to send the same data as subsequent packets

Selection Problem in Internet System (2/2)

- Selection of waiting time t_{i+1} before next access to resource

  $$f : \{\ldots, (t_{i-1}, k_{i-1}), (t_i, k_i)\} \to t_{i+1} = t^*$$

  where k_i is observed expenses on the previous access i.

  E.g., *information update delivery* when longer waiting leads to more update notification losses

- Selection of mapping information fragments d with their consumers u

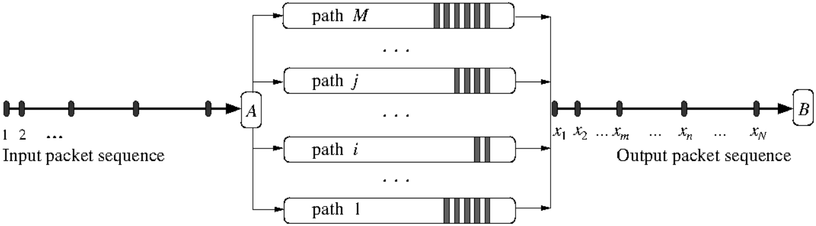
  $$f : D \times U \to \big(\{d_i\}, \{v_j\}\big)$$

  where D and V are information set and agents set, respectively.

  E.g., *task delegation* when appropriate device u is involved into processing of data items $\{d_i\}$.

Multipath Routing of Data Flow

- A source schedules packets $1, 2, \ldots, N$ to split the stream among the different paths $1, 2, \ldots, M$
- Source A applies rate-proportional scheduling, either
  - deterministic (e.g., round-robin),
  - or randomized (e.g., Bernoulli scheme)



D. Korzun, D. Kuptsov, and A. Gurtov, A simulation study of the stochastic compensation effect for packet reordering in multipath data streaming, in The 2015 IEEE European Modelling Symposium (EMS), Oct. 2015, pp. 409414.

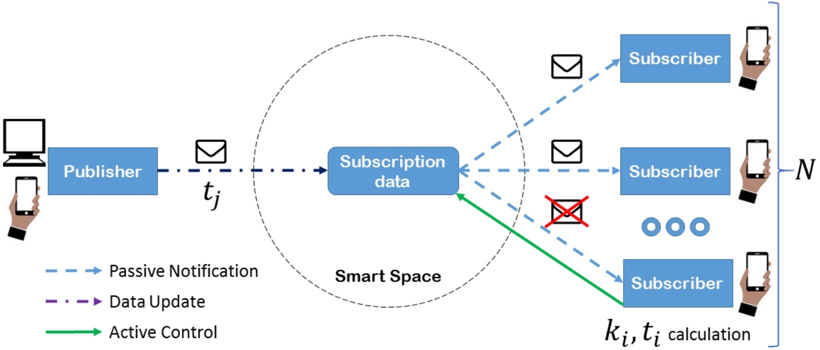
Data Interleaving in Transmission Network

- Dynamic utilization of the channels by considering the repetition in voice bytes and the bytes generated due to silence period in any call
- Whenever a repetition is detected the channel is used for transferring data traffic, i.e., voice and data traffic are interleaved.

Purohit H., Korzun D., Shabaev A., Voronin A. Data Interleaving for Congestion Reduction in Mobile Traffic Transmission. Proceedings of the FRUCT22. Jyvaskyla, Finland, 15-18 May 2018, pp.211-216.

Information Update Delivery

- Let k_i be observed number of losses.
- Then $t_i = \begin{cases} t_{i-1} + \delta & \text{if } k_{i-1} = 0, \\ \alpha t_{i-1} + (1-\alpha)\dfrac{t_{i-1}}{k_{i-1}+1} & \text{if } k_{i-1} > 0. \end{cases}$



Korzun D., Pagano M., Vdovenko A. Control Strategies of Subscription Notification Delivery in Smart Spaces. Distributed Computer and Communication Networks (DCCN2016). Communications in Computer and Information Science, Vol.601, pp.40-51.

Task Delegation / Information Recommendation

- Many information descriptions *d* in a museum collection.
- For a given user *u*, find appropriate information and represent on surrounding devices (local screens, smartphone, etc.)

D.Korzun, S.Yalovitsyna, V.Volokhova. Smart Museum Information Services to Assist Preservation, Transmission and Research in Cultural and Historical Heritage Domain. Baltic DB&IS 2018 Proceedings. 2018.

Semantic Ranking Approach

Recall that agent i makes decision d on time t using strategy f_i:

$$d_t = f_i(t, s, \alpha)$$

based on situation s and individual parameters α,

We apply the semantic network concept as an integration model for describing *knowledge* for s

- history (time-series characteristics)
- participating objects (information and computing resources) and their attributes
- relations between objects (semantics)

Semantic Network

- Semantic network is a directed graph $G = (V, L)$ consisting of nodes (vertices set V) representing domain objects and links (edges set L) representing semantic relations

- Ontology O describes:
  - system of concepts $\{C_i\}_{i=1}^{n}$ (ontology classes) such that any particular node $v \in V$ (ontology class object, instance or individual) belongs to one or more concepts
  - interlinking structure for L, i.e., between which concepts a relation can be and possible types of such relations
  - attributes that $v \in V$ and $l \in L$ may have to reflect additional semantics (e.g., keywords associated with v or l)
  - interpretation for attributes of nodes v and links l as numerical weights (or computation procedures for metrics)

Ranking in Decision Making

- A decision needs to find $k > 0$ the most appropriate information facts
- A fact can be a node $v \in V$, a link $l \in L$, or a connected graph structure s in G
- This data mining can be reduced to the ranking problem when rank values $r_v \geq 0$ or $r_l \geq 0$ are associated with nodes or links
- The higher rank the better is appropriateness of the information
- The rank of a connected graph structure is computed based on ranks of the composed nodes and links

Local Ranking

- Two or more objects are analyzed for similarity based on their content and the similarity of this content

- Rank is computed in respect to some fixed node $u \in V$ and reflects distance of other nodes from u:

$$r_v(u) = 1/\rho(u, v)$$

- Example metric: If u and v have sets K_u and K_v of annotating keywords then the rank reflects the size of overlapping $|K_u \cap K_v|$
  - The larger the number of shared keywords the higher is the similarity

Collaborative Filtering

- Many users (human or computer agents) generate opinions (or provide estimations) about collected object
- Community-based score r_v^* (normalized $0 \le r_v^* \le 1$)
- Scores are combined with other ranking requirements
- Example metric:

$$r_v = \alpha r_v^* + (1 - \alpha) \left(1 - \frac{d_v}{\max_{w \in W} d_w} \right),$$

  - $W \subset V$ is nodes of potential interest for the user,
  - $d_w > 0$ is a physical reachability metric for node w,
  - $0 \le \alpha \le 1$ is a tradeoff parameter between the community scores and reachability.

Structural Ranking

- Rank utilizes the connectivity properties of the semantic network G, similarly as it happens in the PageRank algorithm
- Example metric: Node ranks r_u for $u \in V$ are computed iteratively:

$$r_u^{(i+1)} = \alpha \sum_{l=(u,v)\in L} p_{vu} r_v^{(i)} + (1-\alpha)\pi_u \ ,$$

  - p_{vu} is weight of the link $l = (u, v)$,
  - $0 \leq \alpha \leq 1$ is the damping factor denoting the probability of following the connectivity structure of G,
  - π is a jump probability vector for all $u \in V$.