Parallel Implementation of Hyperpath based Railway Assignment

- an Application to Tokyo Metropolitan Rail Network -

Hideki YAGINUMA

yaginuma@civil.t.u-tokyo.ac.jp

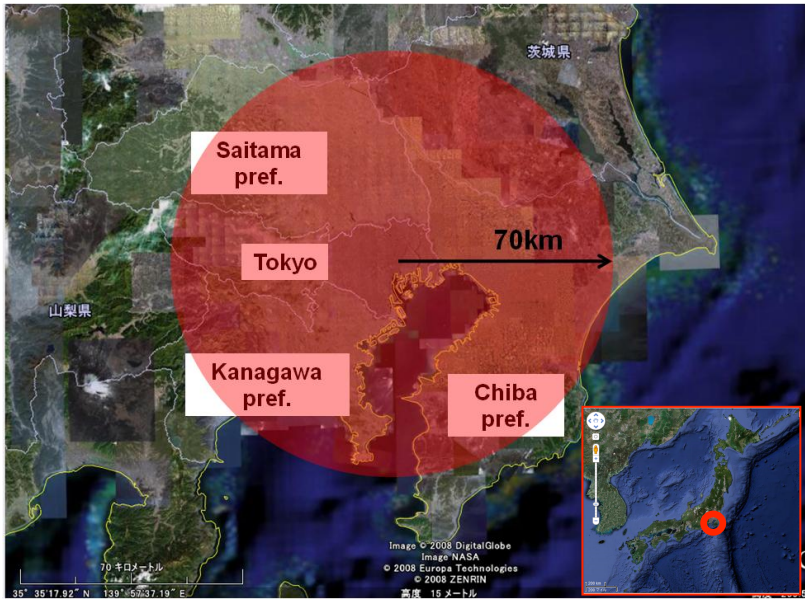
Department of Civil Engineering

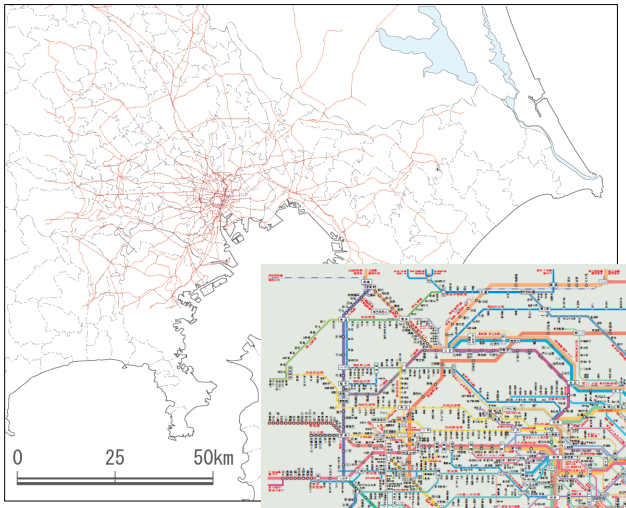
The University of Tokyo

1. Urban railway network in TMA

Tokyo Metropolitan Area (TMA) is a huge capital area and has highly dense railway network.

- **Capital area**
  - ✓ Size: The radius of 70km from the city center (Tokyo station).
  - ✓ Population: About 35 million.

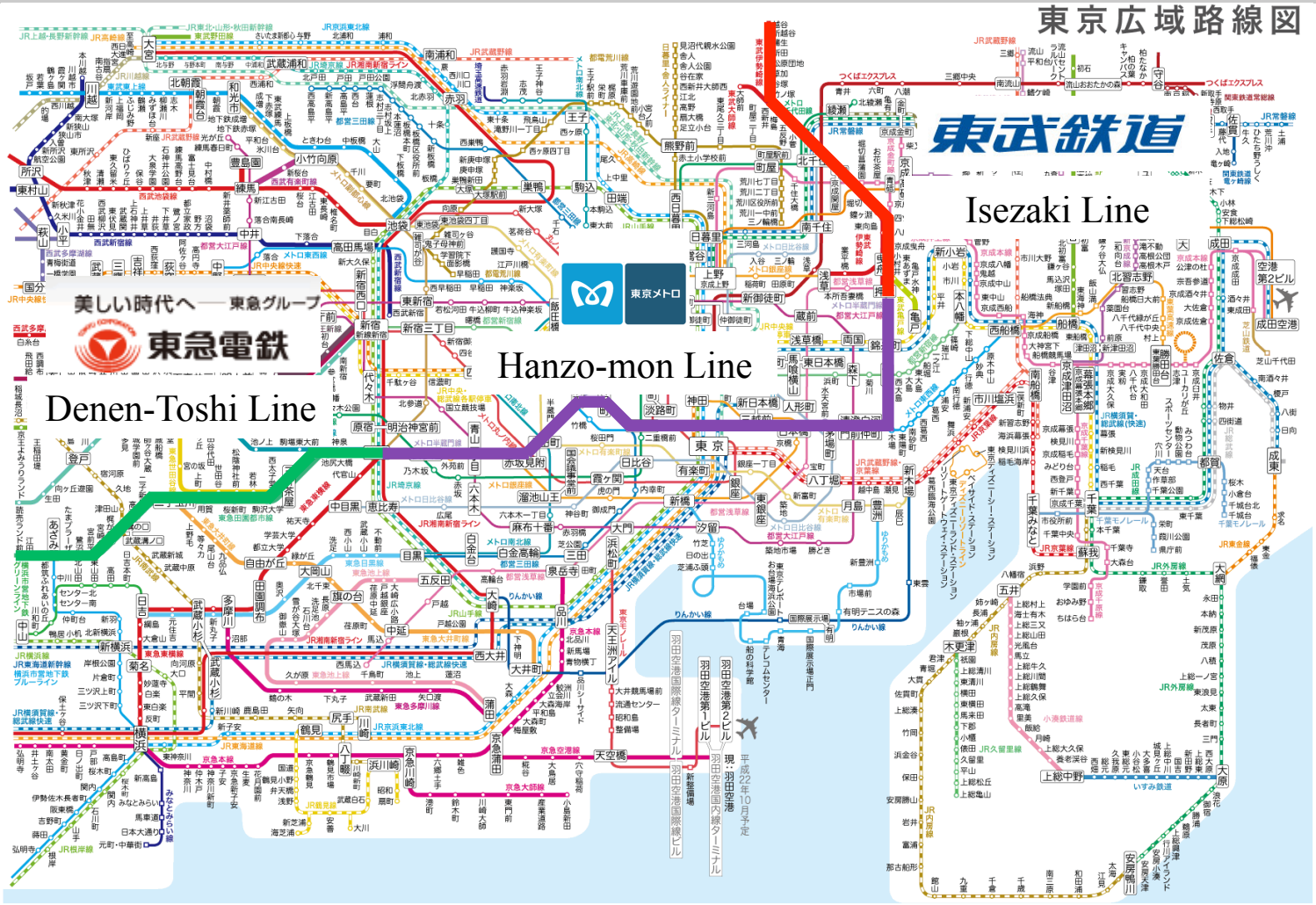




- **Railway Network**
  - ✓ Size: About 130 lines and 1800 stations.
  - ✓ Total number of daily passenger is 40 million.

The largest urban railway network in the world.

1. Mutual direct operation in TMA

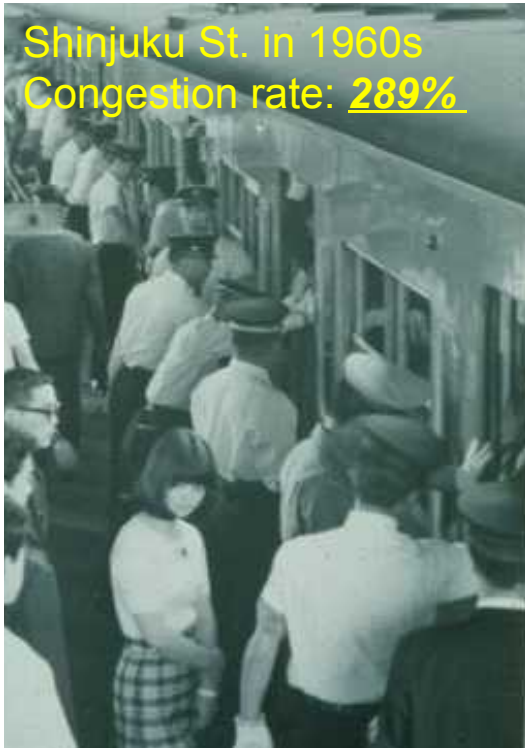


<u>Mutual direct</u> mean direct operation across different line or different railway companies (operators) line

1. Railway congestion in TMA

Railway congestion has been a serious problem since 1960s.

Average in-vehicle congestion rate (CR) decreased to 170% in 2005 in comparison with the 250% in 1960s.



However commuters *still* suffer from physical and mental burdens caused by railway congestion during peak period.

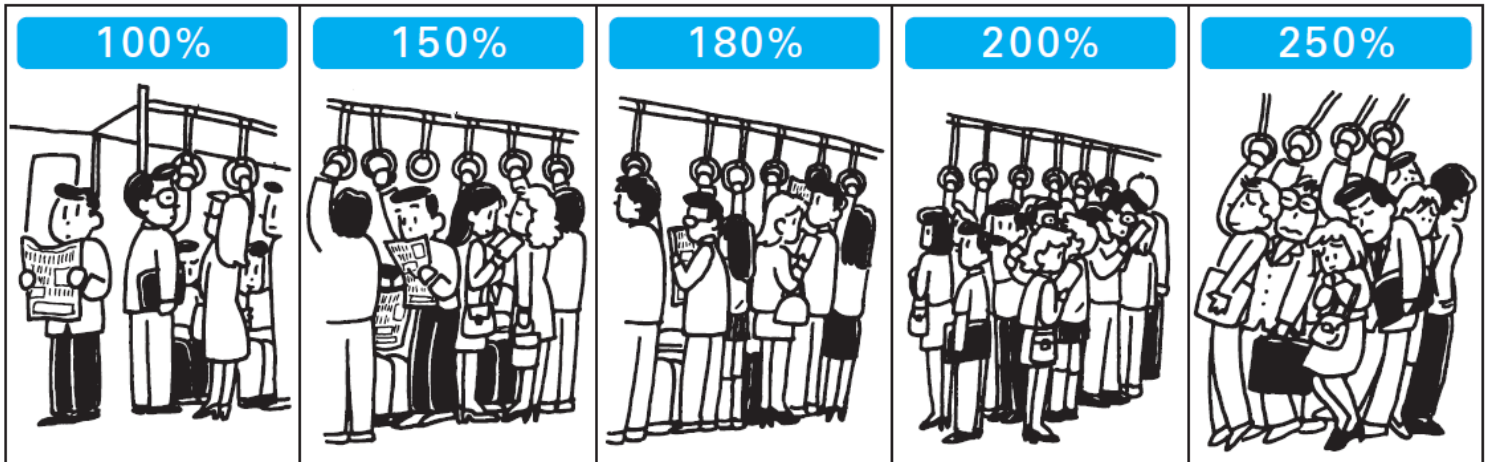
1. Railway congestion in TMA

Railway congestion has been a serious problem since 1960s.

Average in-vehicle congestion rate (CR) decreased to 170% in 2005 in comparison with the 250% in 1960s.

"Congestion Rate (CR)": index of congestion widely used in Japan.

CR(%) := (passenger link volume) / (link capacity)*100



Source: Ministry of Land, Infrastructure, Transport and Tourism (MLIT)

1. Masterplan for urban railway in TMA

- **The 18th masterplan for urban railway network in TMA**

  (Planed in 2000 by the *National Council for Transport Policy*)

  This masterplan proposed the following target toward 2015:

  - Average congestion rate: **less than 150%**
  - Maximum congestion rate for each line: **less than 180%**
  - Reduce transfer time between stations and reduce congestion in station

➡ ***However***, the current conditions in 2010 are still:

- ✓ Average congestion rate: About **170%**
- ✓ Maximum congestion rate: More than **180%** in **18** lines
- ✓ Transfer time: increase **11%** in 120 stations of TMA

Therefore, these targets have not yet been well achieved probably due to

- Budget constraints for constructing new lines
- Unexpected induced congestion

1. Types of congestions and their countermeasures

- **Typical 4 types of railway congestions** (Morichi, 2000)
  i. In-vehicle congestion
  ii. Train delay (Railroad congestion)
  iii. Terminal congestion
  iv. Congestion caused by the closure of railroad crossings

  ⇨ **Network congestion**

- **Some measures were already conducted for reducing congestions:**
  - Constructions of
    connecting lines / four-tracks line / Mutually direct operation
  - Increase in service frequency
  - Renovation of stations (e.g. Transfer route, entrance gate, platform door)

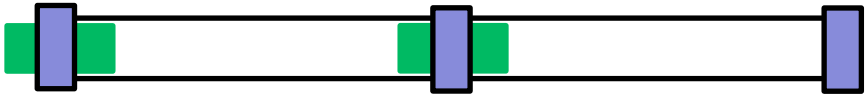
Indeed, these measures helped to reduce serious congestion to some extent by increasing capacity and reducing travel time.

1. Negative side-effects of urban rail policies

However, some railway policies, such as high frequent operation and mutually direct operation, have caused "new network congestion problems".

Ex. Increase in service frequency

- In the case of moderate frequency level, delay do not occur so often. Therefore, in-vehicle and waiting times do not vary and provide regular services.



- In the case of *high* frequency level, however, delay occurs so often. Therefore, in-vehicle and waiting times tend to increase by "knock-on train delay ".

Previous demand forecasting methods were not able to explicitly consider these network congestion problems.

2. Literature Review

- Railway network services have some specific characteristics:
  - Access/egress to stations
  - Service frequency (waiting time)
  - Line transfer
  - Capacity of train vehicle and railroad

- Railway demand forecasting model in TMA:
  - Kato *et al.* (2010): Standard deterministic/stochastic user equilibrium model.
  - Ieda *et al.* (2002): Disaggregate route-departure time model .
  - Yai *et al.* (1997): Disaggregate multinomial probit based route choice model.

➡ These models are analogous to car drivers behavior in road networks, and do not consider some typical aspects of public transportation:
The effects of "Service Frequency (or Waiting Time at stops)"

2. Literature Review

- <u>Public transit assignment models:</u>
  - Spiess and Florian (1989), Nguyen and Pallotino (1988): frequency-based transit assignment model (FBTA) based on concept of "Hyperpath".
  - De Cea and Fernandez (1993): considered congestion with effective frequency.
  - Kurauchi *et al.* (2003): capacity constrained transit assignment model.
  - Cepeda *et al.* (2006): application to the real urban network (Stockholm in Sweden, Winnipeg in Canada)
  - Schmöcker *et al.* (2008): Dynamic FBTA in London Metro Network.
  - Yaginuma *et al.* (2010): considered two congestion effects and application to real small size network.

➡

- FBTA model explicitly consider the some public transportation specific properties, frequency especially. However, there **are no practical applications to large-scale network** like in TMA.
- Solution algorithms of FBTA models are **highly computational** exhaustive compared to standard equilibrium assignment models.

2. Research purpose

To conduct quantitative evaluation of railway congestion reducing polices, this study develops a railway demand forecasting system considering congestion effects:

1. Develop the hyperpath-based railway route assignment model considering in-vehicle and railroad congestions, station to station transfer behavior.
2. Accelerate hyperpath search by the parallelization for the practical use of the proposed system.
3. Apply the system to the Tokyo metropolitan area network and check the validity and the effectiveness of the proposed modeling system.

3. Definition of hyperpath

- Behavioral assumptions:
  - Transit: operated based on **frequency** (not on schedule)
  - Passenger: can only observe the next line to be served

  ⇒Passengers make en-route choice with **common lines problems**

Common lines problems (Chiriqui & Ronolland 1975)

*Passenger who takes the first vehicle to come within his/her "**attractive set**" lines can get to his/her destination earliest.*

Attractive set (choice set) := "Hyperpath"

Strategy of riding trains to the destination (choice set of links)

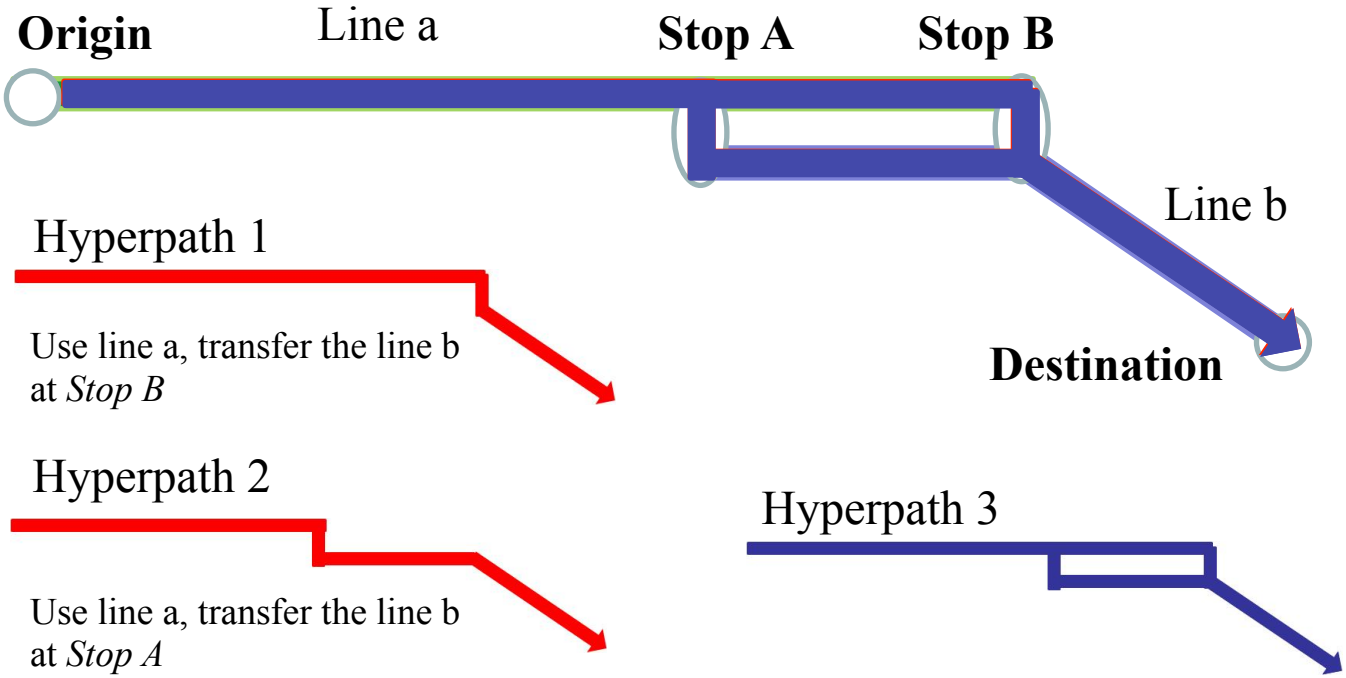
Passenger's hyperpath

= An attractive set with the *minimum* expected cost.

3. Illustration of hyperpath

Example: 1 OD 2 lines network

If at *Stop A* an user sees a vehicle to *Destination*, transfer the line. Otherwise continue to ride to *Stop B*.



Single-route strategies

Multiple-route strategies

A hyperpath is composed of *a set of single and/or multiple routes*.

3. Validity of applying FBTA models to TMA

Trains are usually operated based on timetable (schedule) in Japan.

But in TMA, there are

- ✓ Highly frequent operations (e.g. 2 minutes headway during peak).
- ✓ Chronic train delay due to severe congestion.

⇒ Passenger might not understand timetable perfectly but might recognize service frequency.

- During peak period, FBTA model might be reasonable. So, I apply FBTA in TMA rail network.
- FBTA enables to conduct an analysis with a static framework leading to the computational efficiency.

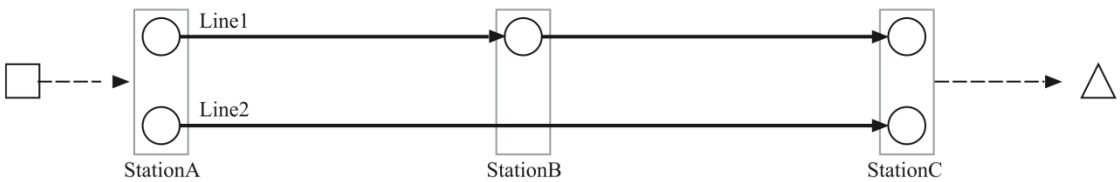
3. Network representation of hyperpath

Hypergraph Representation:

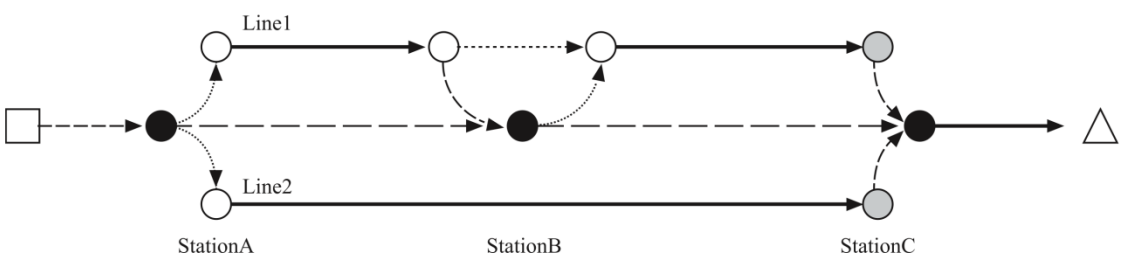
Describe passenger behavior in detail between OD pairs:

(Origin-Destination)

Example Transit Network



Hyper-graph Network



Nodes

- □ : Origin
- ○ : Boarding
- △ : Destination
- ● : Stop(station)
- ● : Alighting

Links

Travel Time

→ : Line	$t_a \geq 0,$	$f_{(a)_l} = \infty$
→ : Boarding	$t_a = 0,$	$f_{a(l)} \geq 0$
→ : Alighting	$t_a \geq 0,$	$f_{a(l)} = \infty$
→ : Stopping	$t_a \geq 0,$	$f_{a(l)} = \infty$
→ : Access	$t_a \geq 0,$	$f_{a(l)} = \infty$
→ : Egress	$t_a \geq 0,$	$f_{a(l)} = \infty$
→ : Transfer	$t_a \geq 0,$	$f_{a(l)} = \infty$

Service Frequency

3. Passenger's behavioral principle

- Passenger chooses the hyperpath p that minimizes his/her expected generalized cost g:

$$g_p = \boxed{\phi \sum_{a \in L_p \cap D_p} \alpha_{ap} t_a} + \boxed{\omega \sum_{a \in T_p} \alpha_{ap} t_a} + \varphi \sum_{i \in S_p} \frac{\beta_{ip}}{F_{ip}} + \xi \sum_{a \in L_p} \alpha_{ap} CDU_a,$$

$\phi, \varphi, \xi, \omega, \kappa, \psi$: Parameter,
α , β: through probability,
x_a: link volume.

Sum-up of frequency in link a include out-going node i :

$$F_{ip} = \sum_{a \in OUT_{p(i)}} f_{l(a)},$$

Effective frequency:

$$f'_{l(a)} = \frac{1}{w_{l(a)}} = \left[\frac{1}{F_{ip}} + \rho \left(\frac{x_a}{Cap_{l(a)}} \right)^{\kappa} \right]^{-1},$$

Congestion disutility:

$$CDU_a = t_a \left(\frac{x_a}{Cap_{l(a)}} \right)^{\psi}.$$

1. Expected in-vehicle time in line links
   - unchanged cost by congestion.
2. Station to station transfer time in transfer links
   - unchanged cost by congestion.

3. Passenger's behavioral principle

- Passenger chooses the hyperpath p that minimizes his/her expected generalized cost g:

$$g_p = \boxed{\phi \sum_{a \in L_p \cap D_p} \alpha_{ap} t_a} + \boxed{\omega \sum_{a \in T_p} \alpha_{ap} t_a} + \boxed{\varphi \sum_{i \in S_p} \frac{\beta_{ip}}{F_{ip}}} + \xi \sum_{a \in L_p} \alpha_{ap} CDU_a,$$

$\phi, \varphi, \xi, \omega, \kappa, \psi$: Parameter,
α , β: through probability,
x_a: link volume.

Sum-up of frequency in link a include out-going node i :

$$F_{ip} = \sum_{a \in OUT_{p(i)}} f_{l(a)},$$

Effective frequency:

$$f'_{l(a)} = \frac{1}{w_{l(a)}} = \left[\frac{1}{F_{ip}} + \rho \left(\frac{x_a}{Cap_{l(a)}} \right)^{\kappa} \right]^{-1},$$

Congestion disutility:

$$CDU_a = t_a \left(\frac{x_a}{Cap_{l(a)}} \right)^{\psi}.$$

3. Expected waiting time in nodes

- Effective frequency (EF) is defined at the inverse of frequency and changes with the delay (railroad congestion) level.

- EF describes the frequency change via "*congestion index*".
  If link flow increases, EF decreases.
  If capacity increases, EF increases.

3. Passenger's behavioral principle

- Passenger chooses the hyperpath p that minimizes his/her expected generalized cost g:

$$g_p = \phi \sum_{a \in L_p \cap D_p} \alpha_{ap} t_a + \omega \sum_{a \in T_p} \alpha_{ap} t_a + \varphi \sum_{i \in S_p} \frac{\beta_{ip}}{F_{ip}} + \xi \sum_{a \in L_p} \alpha_{ap} CDU_a,$$

$\phi, \varphi, \xi, \omega, \kappa, \psi$: Parameter,
α , β: through probability,
x_a: link volume.

Sum-up of frequency in link a include out-going node i :

$$F_{ip} = \sum_{a \in OUT_{p(i)}} f_{l(a)},$$

Effective frequency:

$$f'_{l(a)} = \frac{1}{w_{l(a)}} = \left[\frac{1}{F_{ip}} + \rho \left(\frac{x_a}{Cap_{l(a)}} \right)^{\kappa} \right]^{-1},$$

Congestion disutility:

$$CDU_a = t_a \left(\frac{x_a}{Cap_{l(a)}} \right)^{\psi}.$$

Feedback loop exists if congestion are considered (Update link volume: x_a)

4. Congestion disutility in line links

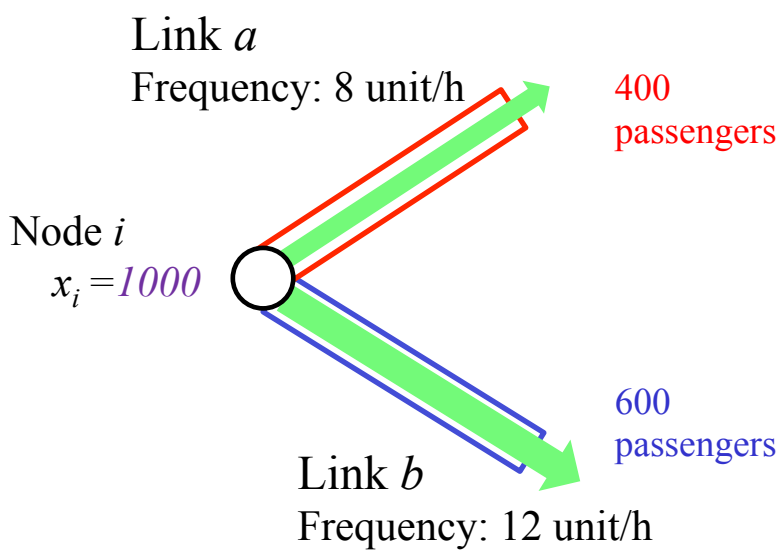
- passengers perceived disutility change with in-vehicle congestion, and is defined as the product of in-vehicle time and congestion index.

3. Loading passenger demand to the network

Loading passenger OD demand on the minimum-cost hyperpath:

- Passengers are loaded to each link proportionally to the service frequency of the links.

Example: two outgoing links



When *1000* passengers come to node *i*, they are loaded to links *a* and *b* as follows:

link a = *400* ($= 1000 \times \frac{8}{8+12}$)

link b = *600* ($= 1000 \times \frac{12}{8+12}$) .

3. Solution algorithm (Method of Successive Average: MSA)

STEP 0. Initialization
SET: Iteration n=0,
Link flow: x_a=0, CDU=0

STEP 1-1. Minimum cost hyperpath search
For a given destination.

STEP 1-2. Flow assignment
Using frequency for given STEP 1-1 results.
And repeat flow assignment for all origin

Repeat for all destinations

STEP 2. Update link flow and variables
Link flow=average x_n+1 and x_n
Update :EF, CDU. n+1

Repeat this until n reached to the threshold or it satisfies convergence criteria.

MSA is an iterative method to search the equilibrium condition.

3. Mathematical property of the proposed model

I check the equilibrium property of the proposed model using VIP (variational inequality problem) by referring Wu et al. (1994).

Assume that passenger flows on the network satisfy Wardrop's user equilibrium condition.

[VIP] Find equilibrated link flows x^* and hyperpath flows h^* which satisfy

$$g_p(x^*)^T(x - x^*) + \boldsymbol{W}^T(h - h^*) \geq 0$$

$$\forall\, x \in \boldsymbol{X} \text{ and } \forall\, h \in \boldsymbol{\Omega}$$

where,

$\boldsymbol{W}$: vector of waiting times for all hyperpath

$\boldsymbol{\Omega}$: set feasible hyperpath flow

$\boldsymbol{X}$: set of feasible link flow

[Theorem 1]

Assume that $g_p(x)$ is continuous and Ω is a compact and convex set. Then there is at ***least one solution*** *of VIP*

[Theorem 2]

Consider the VIP. Then, x^* and $W^T h^*$ are ***unique*** *if $g_p(x)$ is a strictly monotone function.*

- ✓ The generalized cost functions of the proposed model depends only on the vector of link flows.
- ✓ Functions for effective frequency and congestion disutility are both convex.

Through rough check, I anticipate that the proposed model would satisfy the theorems by Wu *et al*.

However, there is a need for rigorous proof.

4. Strategy of accelerating algorithm

1. Program execution environment

verification by focusing on the types of **compiler** and **optimization** options.

- Compiler

  GCC (GNU Compiler Collection) and Intel

- Optimization options

Compile Option	Effect of Option
Defaultx (None option)	None optimization
O2	Increasing the execution speed by vectorization
O3	O2 + Optimize of memory data access
fast (use only Intel)	O3 + Optimize of function used in compiler

4. Strategy of accelerating algorithm

2. Faster data access

To accelerate the large network data access, I apply the **linked list** data structure.

3. Faster hyperpath search algorithm

Apply the **Hyperpath-Dijkstra algorithm** (Cominetti & Correa 2001).

⇨ Theoretical computation time reduces from $O(m^2)$ to $O(m \log m)$

(m: number of links)

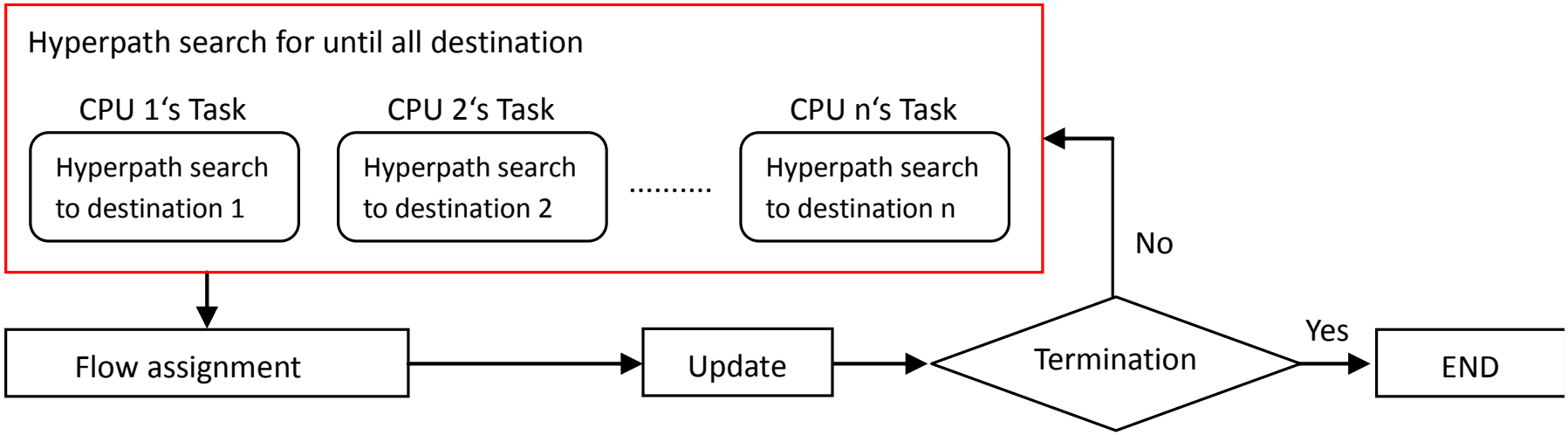
4. Acceleration of iterative process

Apply the **parallelization** of hyperpath search

4. Parallelization of hyperpath search

- In solution algorithm, the "hyperpath search" account for the 91% of the total computation time (Yaginuma et al. 2010).
- It is possible to conduct hyperpath search for each destination node separately.

➡ Parallelization of hyperpath search for each destination node is implemented with multi-thread computing.



4. Testing with an artificial small network

Specification

- 10×10 square grid network
- Create the 100 test networks and transform them into hypergraph representation

  ⇒ ・Average # of nodes: 897

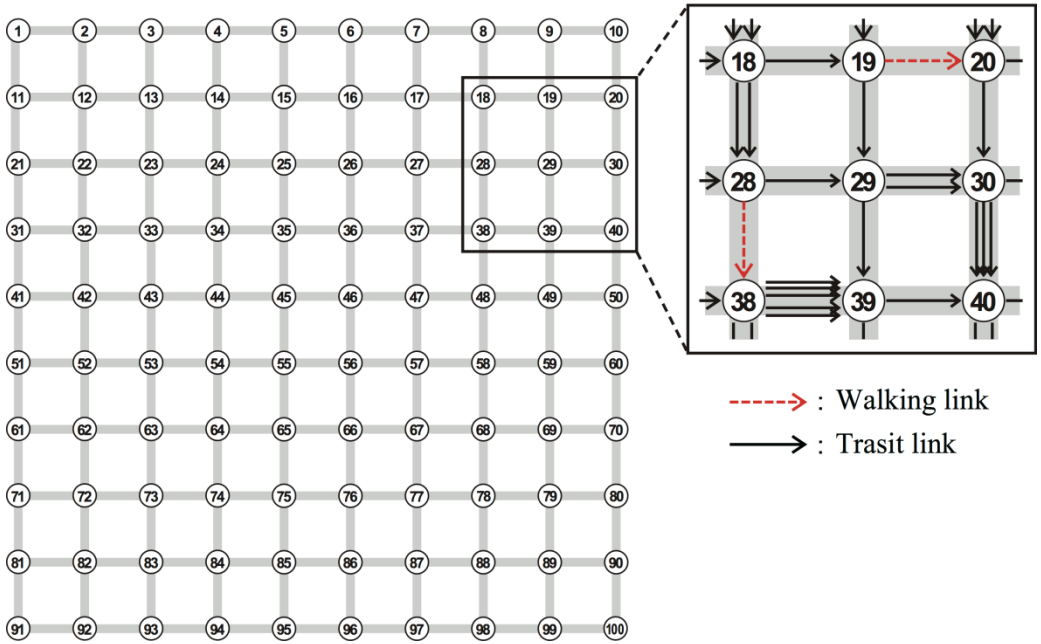
  ・Average # of link : 2687

- Implementation environment:

  C++ with Open MP library

- Machine spec:

  Intel Core i7 with 8G memory



Variable	Generating procedure	Setting range
Number of mamimam parallel lines	uniform random numbers	1～10
Link travel time	uniform random numbers	8～12
Link frequency	uniform random numbers	5～20
OD pair	uniform random numbers	50 pair
Walking link	Generate a random rate of 10% in the case of number of parallel line at 1	

4. Convergence check of MSA

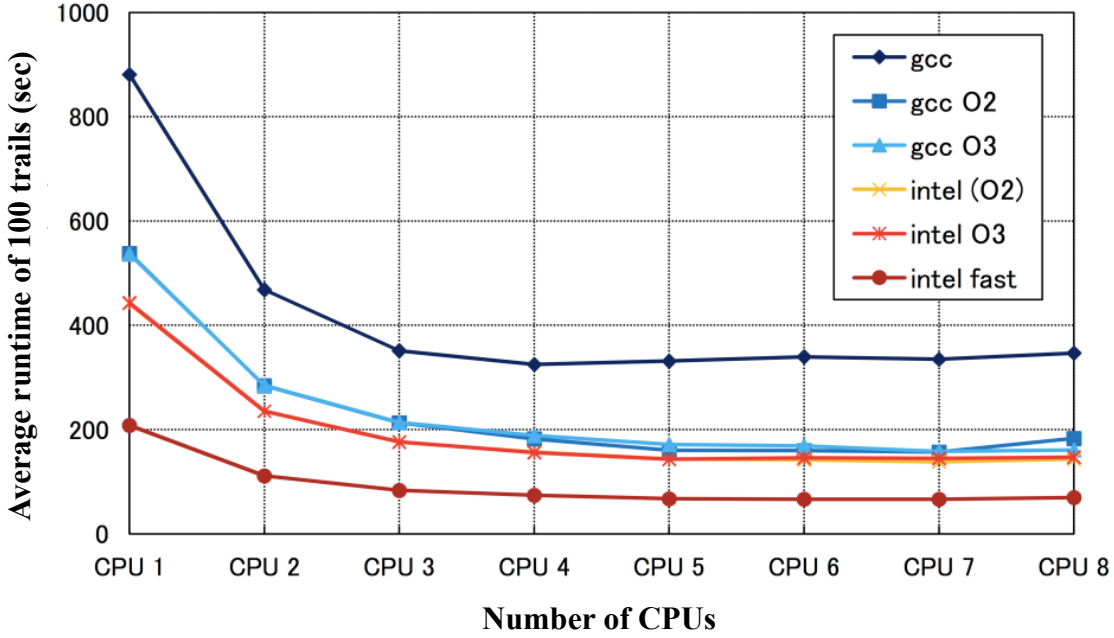
Example of convergence of MSA algorithm:

- The algorithm seems to be well converged.

  ⇒ The model has reached convergence in equilibrium.

4. Effects of parallelization

Effect of parallelization:



- The parallelization significantly decreased runtime.
  - ✓ Difference in compiler brought 1.99 times and difference in compiler option brought 2.13 times difference in computational speeds.
  - ✓ However, "Depletion Effect (Amdahl, 1967)" became distinct with the increase in the number of CPUs.

5. Application to Tokyo's railway network

- Network data

1. Nodes & links: 2005 Tokyo metropolitan transportation census*
2. Link flow & OD: 2005 Tokyo metropolitan transportation census

   ***Tokyo metropolitan census** is an authorized survey for every 5 year. It corrects railway users' daily travel behavior (e g. departure time, origin-destination & route)

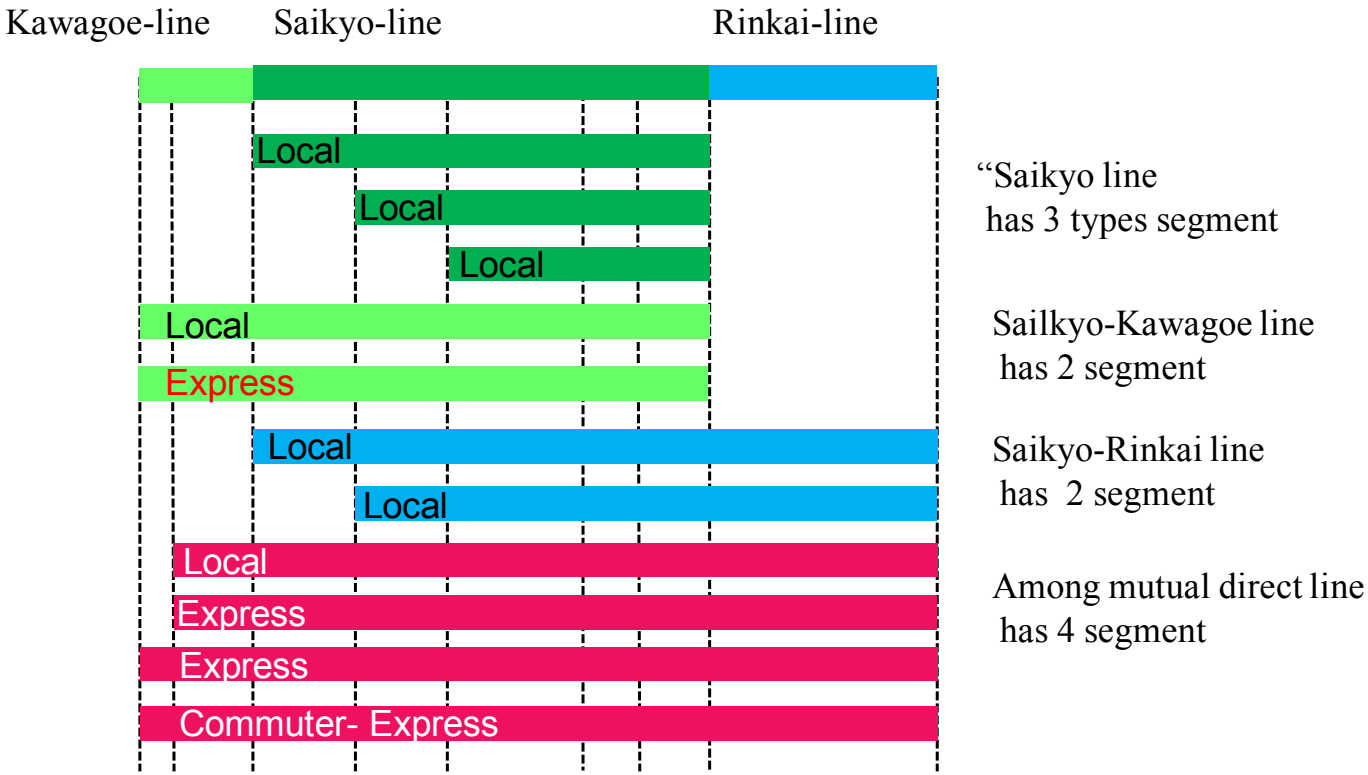
3. Link travel time: Railway timetable in TMA
4. Frequency(each line segment): Railway timetable in TMA
5. Train vehicle capacity: 120(parson/vehicle) * unit (number of vehicle)

➡ Convert these data into the hypergraph representation

5. Line segmentation

Many lines operate the same lines but with different origin and destination stations and have several types of operations (e.g. Local train, Express train, Mutual direct...)

Ex.) Saikyo line (one of the longest north-south corridor line with mutual operation)



Input network data is increase by mutual direct and different OD stations

5. Tokyo's railway network

- Specification:

<u>Physical network</u>:

Lines : 124

Line Segments:837

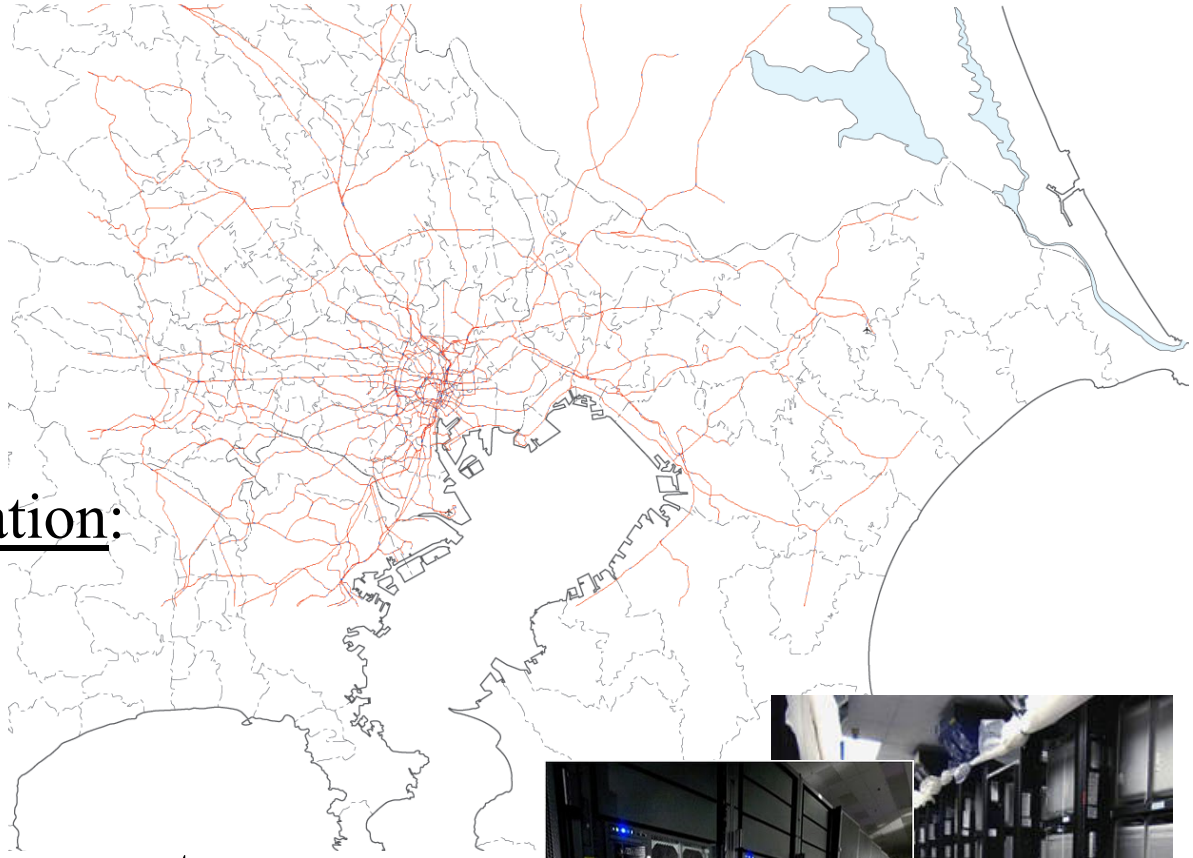
Stations : 1416

Links : 40701

<u>Hypergraph representation</u>:

Nodes : 29902

Links:55972



- Calculation environment:

✓ C++ with Open MP library

✓ "TSUBAME2.0" (64CPUs available)

TSUBAME 2.0

5. Calibration of model parameters

- Numbers of parameters ω and ψ are set based the calibration which tries to fit the data with its prediction through trial and error repetitions.
- Numbers of other parameters are the same with the previous literature on transit assignment in the TMA (Kato et al. 2007; Iwakura et al. 2000).

Generalized cost function:

$$g_p = \phi \sum_{a \in L_p \cap D_p} \alpha_{ap} t_a + \omega \sum_{a \in T_p} \alpha_{ap} t_a + \varphi \sum_{i \in S_p} \frac{\beta_{ip}}{F_{ip}} + \xi \sum_{a \in L_p} \alpha_{ap} CDU_a,$$

Effective frequency:

$$f'_{l(a)} = \frac{1}{w_{l(a)}} = \left[\frac{1}{F_{ip}} + \rho \left(\frac{x_a}{Cap_{l(a)}} \right)^{\kappa} \right]^{-1},$$

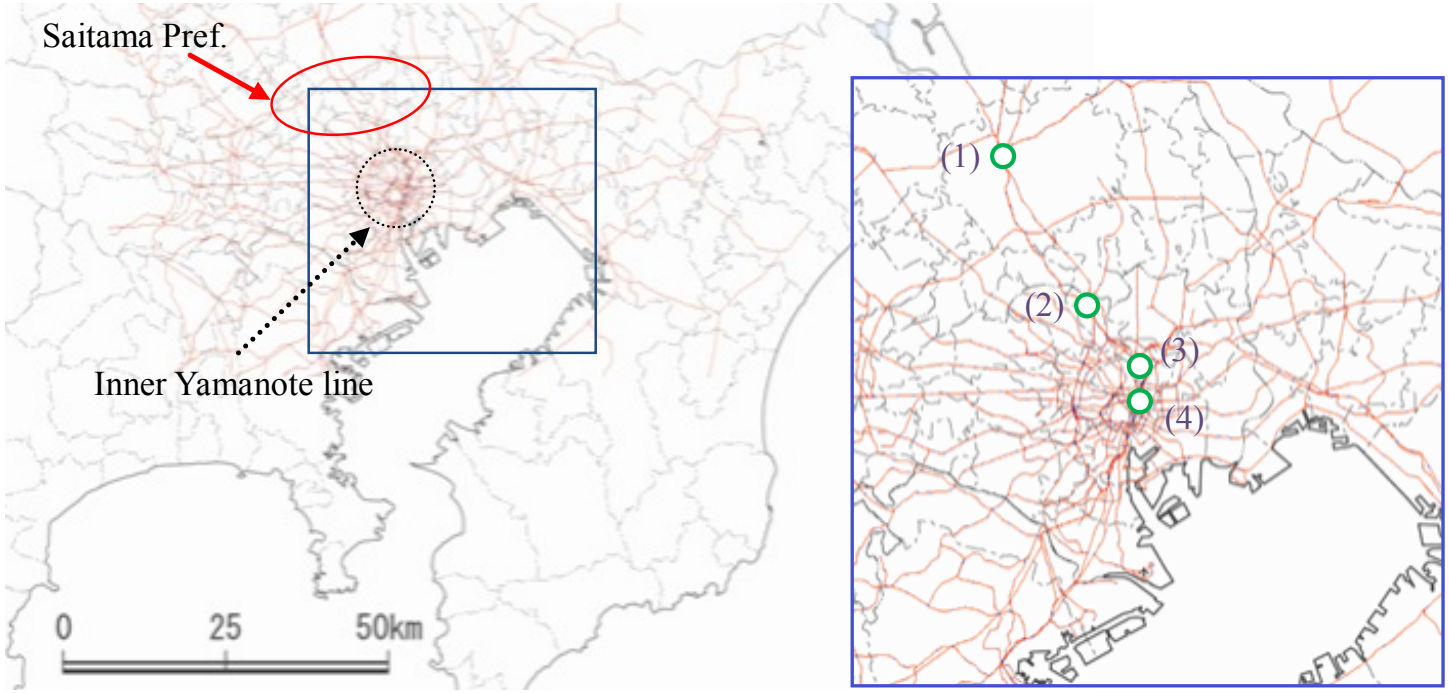
Congestion disutility:

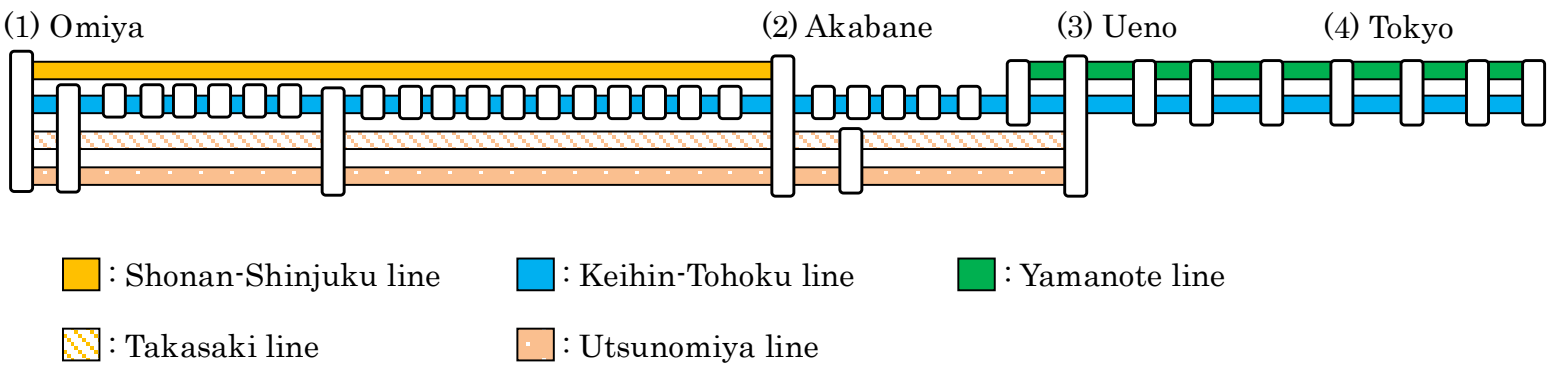
$$CDU_a = t_a \left(\frac{x_a}{Cap_{l(a)}} \right)^{\psi}.$$

Φ	φ	ζ	ω	ρ	κ	Ψ
1.00 (Normalised)	*1.45*※	*2.07*※	*1.20*	*1.06*※	*2.21*※	*2.50*

* indicates parameters taken from Kato *et al*. (2007) and Iwakura *et al*. (2000)

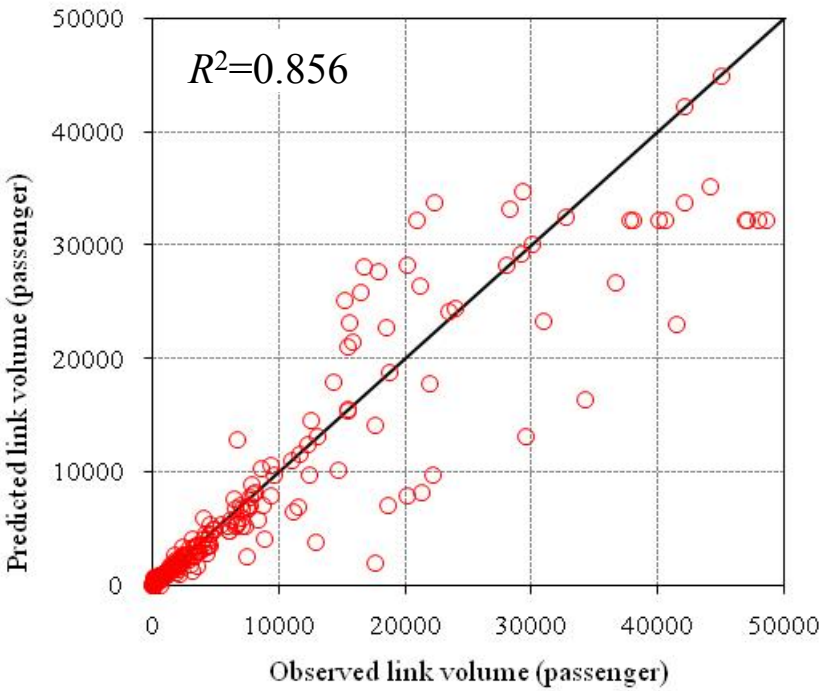
5. Results of overlapping section (example)



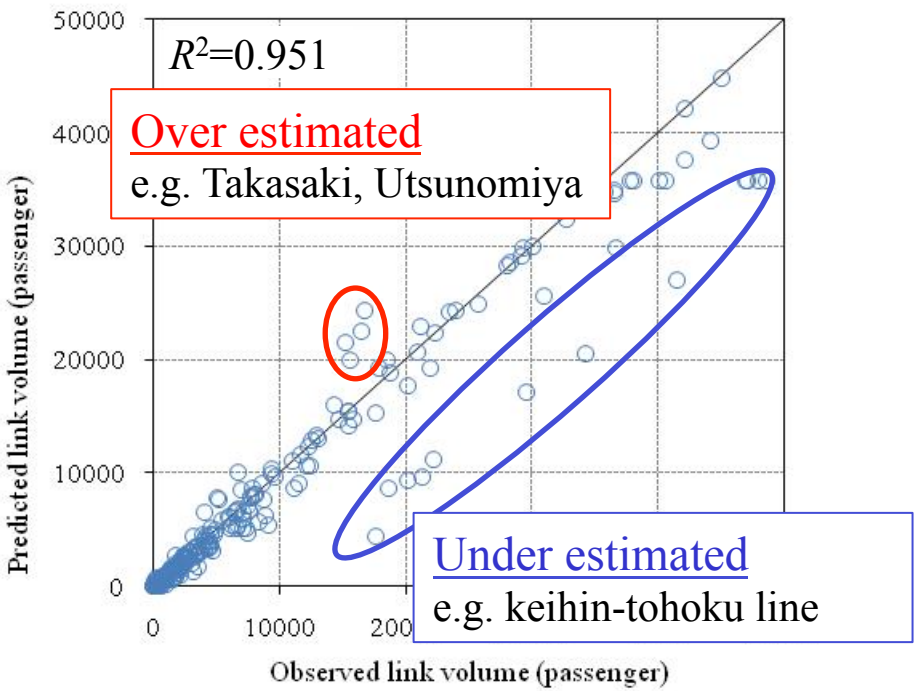


4. Goodness-of-fit of the Models (with 412 observed link flows)

Without congestion effects:



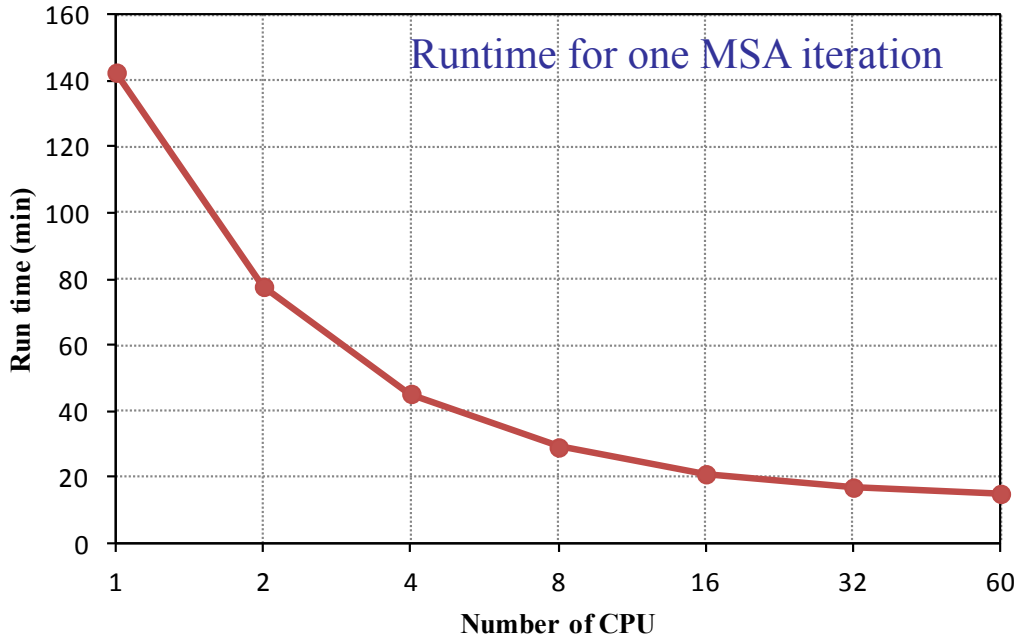
With congestion effects:



- Congestion effects would be crucial for Tokyo network.
- However, in some links even with congestion, the traffic volume is underestimated or overestimated.

4. Effects of parallelization

- Again, the parallelization decreased runtime significantly.



✓ Parallelization of hyperpath search enabled to reduce the run time significantly for large-scale applications.

✓ However, "Depletion effect" becomes distinct with the increase in the number of processers.

6. Conclusions & Future Works

Conclusions:

- We proposed a frequency-based railway route assignment model with two different congestion effects (in-vehicle and on-railroad) and applied this to the TMA rail network.
- The model improved the goodness of fit compared with the one without congestion effects.
- Parallelization of hyperpath search enabled to reduce the run time significantly for large-scale applications.

Future works:

- Consider the transfer behavior between platforms.
- Introduce the railway fare to generalized cost function.
- Further *speed-up* with the improvement of parallelization code and parameter fitting compatible.

Thank you for your kind attention.