Modelling the Bandwidth Allocation Problem in Mobile Service-Oriented Networks

Bo Gao, Ligang He and Chao Chen
Department of Computer Science, University of Warwick, Coventry, CV4 7AL, UK
liganghe@dcs.warwick.ac.uk

ABSTRACT

When the services requested by mobile application workflows are distributed over a network of mobile smart devices, the question arises as to which service should be allocated with how much bandwidth and when in order to satisfy service demands? Furthermore, the mobility of smart mobile devices brings forward the challenge to determine how changes in mobile network conditions affect the bandwidth requirements of interacting services. In this paper, we construct a Network I-O model to describe the bandwidth dependencies in mobile service-oriented networks incorporating and extend on the principles of the Leontief Input-Output model in economics. Various factors such as bandwidth and service demand are accounted for in the model. The network I-O model lays the foundation for future objective developments in ubiquitous mobile computing scenarios. Results from simulation studies are presented to demonstrate the effectiveness of the proposed methods.

Keywords

Mobile service-oriented networks; bandwidth allocation; mobile application workflows

1. INTRODUCTION

Technologies that were once thought to be futuristic like self-driving vehicles and wearable digital assistants are fast becoming a reality. The proliferation of mobile smart devices is radically changing the way applications are delivered. One key characteristic of mobile applications is their unique ability to sense and capture information about the immediate environment its user is in. However, the downside of being in such a dynamic environment is the limited and unreliable wireless network capacity. Therefore allocating the correct amount of bandwidth to the right mobile service on the right device at the right time becomes an important issue.

Consider as an example a wireless network of personal mobile smart devices as shown in Fig. 1. Each service module serves a different purpose to the user and is able to run independently. However, when connected via a network, these services can also be dynamically combined to serve more complex mobile application workflows. For instance, as illustrated in Fig. 1 by the coloured lines, the user can use the tablet to stream a remote video file from the storage service by transcoding it to a format that is readable by the smart TV. At the same time, if the smartphone is also available (has adequate bandwidth) to the network, the tablet can transcode the video file to a voice stream that can be transcribed to a text stream on the smartphone. This text information can then be translated on the tablet to a language chosen by the user and streamed to the smart TV as subtitles.

Observe that the process of service composition is dynamic and non-deterministic. The composition decision may be influenced by many factors such as 1) Network Conditions: the smartphone might not have enough bandwidth therefore the "Voice-to-Text" service on the smartphone cannot receive the voice stream from the "Transcode" service on the tablet; 2) Dynamic Application Information: the film may be in a language which the user can understand, and therefore the "Translation" service is not included in the workflow.

In this paper, we refer to this type of mobile networks as *mobile service-oriented networks* (MSON)s on which a universe of services is distributed, and mobile applications are executed as dynamic compositions of these services. This service-oriented network structure commonly exist among vehicular ad hoc networks (VANETs) [14, 7], urban sensing networks [6, 15], biomedical application workflows [16, 12] and other smart environment applications.

One of the key challenges that comes with the research in MSON is the constrained and unpredictable wireless network connection capacity (e.g., bandwidth). In contrast to desktop based SOA networks [12], focus of the bandwidth allocation problem has shifted from the centre of the network (considered fast in an MSON) to the access points at the edge of the network, which raises the question: *How*

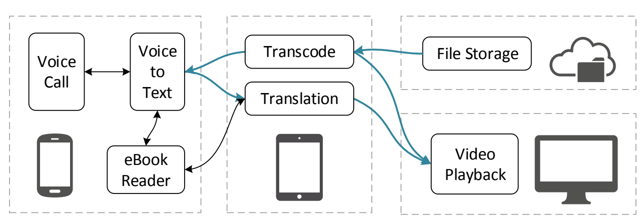

Figure 1: A simple mobile service-oriented network example.

MSWiM'15, November 2–6, 2015, Cancun, Mexico.
 ISBN 978-1-4503-3762-5/15/11 ...$15.00.
DOI: http://dx.doi.org/10.1145/2811587.2811628.

much bandwidth should each service require and when in order to satisfy the desired QoS? Moreover, since the services interact with each other dynamically during workflow executions, it is a non-trivial task to investigate the impact of the changes in service demand (e.g., the arrival rate of workflows) on bandwidth allocations for individual services in order to maintain the QoS. The problem is further complicated by the mobility of the mobile devices. As mobile devices move around, their bandwidth or latency may change, which brings the challenges to adjust the bandwidth allocations for the interacting services.

In this paper, we first discuss the applications of an MSON and related work in the next section. We then extend the Leontief I-O model in economy and formulate the bandwidth dependencies of an MSON with a Network I-O model in section 3. We conduct a series of parametric simulations in section 4 to verify and demonstrate the applicabilities of our network I-O model. We conclude the paper and discuss future directions of this work in section 5.

2. MSON AND RELATED WORK

An MSON infrastructure can be observed from many research areas. In vehicular wireless systems, many applications are built on top of cooperative networks of mobile smart devices installed on smart vehicles. Exemplar applications includes BitTorrent-styled location significant content downloading [14] and vehicle-to-vehicle environment and safety sensing [7]. In biomedical applications, a mobile application workflow is used in [16] to describe a sensor-based biomedical application which includes mobile devices used as both sensors and data processing units. The motivating scenario addressed in [12] describes the benefit of using service composition in a hospital resource scheduling application. In other mobile application areas, a mobile P2P file sharing framework is presented in [10]. A framework for mobile P2P social content sharing is presented in [4]. These studies all share the same underlying MSON infrastructure.

The idea of mobile devices as service hosts is also an active research topic. In [13], an SOA-based approach is presented to support interactions between business applications running on J2ME. In [9], opportunistic composition of sequentially-connected service over a decentralised mobile ad hoc network is proposed. Experiments conducted in [11] demonstrates that this opportunistic communication is viable at scale. A middleware is created in [8] to reduce user perceived latency while accessing remote services on mobiles by pre-fetching and caching data according to a sequence prediction algorithm. The same technique has been shown to reduce battery cost. All these studies are about developing the service-oriented architecture on mobile devices. None of the work discusses the bandwidth requirement of mobile services.

3. NETWORK I-O MODEL

3.1 Input-Output Analysis in Economics

Suppose a nation's economy is divided into n sectors that produce goods or services. Let x_i be the value of goods or services produced by sector i, we then have a *production vector* $\boldsymbol{x} \in \mathbb{R}^n$ to list the output from all sectors of the economy. In order to avoid waste and deficiency, production is planned in accordance to the demand of goods and services which originates from two channels: ***External demand*** represents consumer demands, exports, planned surplus, etc. from the economy. Let d_i be the external demand of sector i, then $\boldsymbol{d} \in \mathbb{R}^n$, namely the *external demand vector*, lists the external demand (output) of all sectors of the economy. ***Intermediate demand***, represents intra-sector demand of good and services. For instance, assume a small town with two primary industries: a steel plant and a railway. Then in order to produce goods, the steel plant requires services from the railway. To represent the intermediate demand, a square matrix $A \in \mathbb{R}^{n\times n}$, namely the *consumption matrix*, is assumed, in which a_{ij} denotes the production (input) needed from sector i per unit of production (output) by sector j. With this definition, we have that in order to produce x_j units of good or services, sector j will demand $x_j a_{ij}$ units from sector i, that is the intermediate demand by sector j from sector i. When the economy's production balances the total demand for that production exactly, we have:

$$\underbrace{\boldsymbol{x}}_{\text{production}} = \underbrace{A\boldsymbol{x}}_{\text{intermediate demand}} + \underbrace{\boldsymbol{d}}_{\text{external demand}} \tag{1}$$

which is the cornerstone of the Leontief Input-Output model of economics. This model helps economists understand how changes in one sector affect others, and predict the production level required to balance the demand exactly.

3.2 The Economy of Mobile Service-Oriented Networks

We consider each service as a sector of the network economy. Entailed by the SOA paradigm, services are combined and possibly recombined to create complex applications (workflows) that serve the demand of the end users. This composition of services is a dynamic run-time decision process which adapts to: the fluctuating network conditions (which is especially true for a mobile network), various application-dependent QoS level requirements [17, 2, 1] and dynamic application information (e.g., whether the user requires translation for a video). This means that the exact execution sequence of services is not predefined and therefore the communication demands between services are non-deterministic. This behaviour is similar to that of the common economies analysed by the Leontief I-O model. For instance, consider manufacturing and raw material as two sectors of an economy. Each product of the manufacturing sector has its own bill of materials and may require different amount of input from the raw material sector. Furthermore, a repair service may avoid input from the raw material sector completely if it does not require any replacement parts.

Data as Commodity: Services (sectors) of an MSON economy produce and exchange data to serve the demands of its end users. This data as a commodity may carry information requested by the user (e.g., query services), which may be the product in accordance to user input (e.g., image processing service), or simply be the confirmation from the service that the user's request has been recorded (e.g., flight check-in service).

Bandwidth as Currency: Exchange of data is facilitated by the network. One unit of network bandwidth facilitates the exchange of one unit of data in one unit of time. Similar to the common currency (e.g., one US dollar) used in an economy to measure goods and services of different sectors, one unit of bandwidth is the common currency of an MSON economy.

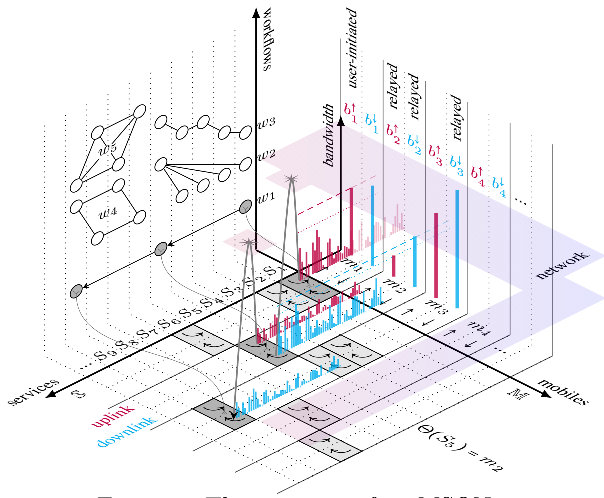

Figure 2: The economy of an MSON

Exchange of Data: Let $\mathbb{S}$ denote the universe of services distributed over the network containing a set $\mathbb{M}$ of mobile devices, according to a mapping scheme $\Theta : \mathbb{S} \to \mathbb{M}$. (In Fig. 2, we have $\Theta(s_1) = m_1, \Theta(s_2) = m_1, \Theta(s_3) = m_3$, and so on.) For each service $s_i \in \mathbb{S}$, assuming that historical data (e.g., collected by filtering logging data) are available [3, 5], and let β_i, measured in units of bandwidth, denote the (average) size of data produced by each run of s_i as an intermediate step of a service composition. The effect of β_i is three-fold:

First, as an intermediate product, β_i needs to be communicated to the next service(s) $s_j \in \mathbb{S}$ as instructed by the service composition (application workflow). If s_j is not located on the same device as s_i, then β_i needs to be sent from its host $\Theta(s_i) \in \mathbb{M}$ over the MSON. We define a co-location indicator

$$\omega_{ij} = \begin{cases} 0 & \text{if } \Theta(s_i) = \Theta(s_j), \\ 1 & \text{otherwise.} \end{cases} \tag{2}$$

so that $\omega_{ij} = 1$ indicates that, if destined to s_j, the task of sending β_i would consume the uplink bandwidth of $\Theta(s_i)$. As illustrated in Fig. 2, when a service workflow w_1 is initiated on m_1, because the next service (s_5) is located on a different device (m_2), m_1 has to first upload the data to the network. We refer to this type of bandwidth cost as *self-initiated cost* later on in this subsection. We assume that a square matrix $P = [p_{ij}]_{|\mathbb{S}|\times|\mathbb{S}|}$, in which p_{ij} denotes the probability that a run of s_i is to be succeed by a call to s_j (β_i is to be sent to s_j), is known through profiling [3, 5]. Together with our co-location indicator ω_{ij}, we define

$$\rho_i = \sum_j p_{ij}\omega_{ij} \tag{3}$$

which gives the probability that each unit product (data) of s_i is to be uploaded to the MSON by $\Theta(s_i)$. Note that $\sum_j p_{ij}$ is not necessarily one, because each run of s_i is not necessarily succeeded by a call to another service.

Second, for a service s_j to receive β_i, the downlink bandwidth of $\Theta(s_j)$ is consumed. In the example illustrated in Fig. 2, this receive action is taken by m_2 which hosts s_5. This creates a dependency between the consumption of the uplink bandwidth of the sender device and the downlink bandwidth of the receiver device in the MSON economy. We define

$$\eta_{ij} = \frac{p_{ij}\omega_{ij}}{\sum_k p_{ik}\omega_{ik}}, \quad s_k \in \mathbb{S} \tag{4}$$

which gives the probability that a unit of data sent by s_i to the MSON is to be received by s_j. We refer to the cost occurred in this type of process as *relayed cost* later on in this section.

Third, depending on the specification of the application workflow, the service which received β_i may be requested to further communicate with other services. Take for instance the example workflow illustrated in Fig. 2, s_5 is to continue the workflow and communicate with s_9. This action consumes the uplink bandwidth of m_2 and the downlink bandwidth of m_3. As such, following a service workflow, a sequence of services in the service composition would be requested to perform communication tasks. This chain effect exists in every application and is triggered by the execution of the head services of the workflow.

Self-initiated Cost vs. Relayed Cost: Following previous discussion, we discover that the production of data, and thus the cost of bandwidth, in a MSON can also be classified into two classes: ***Self-initiated*** production of data refers to data generated by the head services executed at the start of every application workflow and exhibits the same characteristics as the external demands in Leontief's model. Devices that hosts these head services bear the cost of sending data to subsequent services. These costs are in the form of uplink bandwidth of the sender device, and are initiated solely by the service itself (e.g., triggered by user action). Let λ_i denote the arrival rate of s_i, then the self-initiated cost to the uplink of $\Theta(s_i)$ which we denote $c_i^\uparrow$ is given by

$$c_i^\uparrow = \lambda_i\beta_i\rho_i \tag{5}$$

The other class of data production is in contrast caused by services that executed prior in the application workflow and thus no consequent cost is self-inflicted. We refer to this as ***relayed*** production of data. Bandwidth cost from this class of data production can be in forms of both uplink and downlink bandwidth. We derive the cost function of this class in the proof of Theorem 1.

3.3 Network I-O Model

Given a service $s_i \in \mathbb{S}$, let $x_i = x_i^\uparrow + x_i^\downarrow$ denote its total, uplink and downlink bandwidth costs respectively. We now construct a model that derives these values with the limited information we have about the MSON, i.e., Θ, P, $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$.

Definition 1. For each pair of services $\{s_i, s_j\} \in \mathbb{S}^2$, the elements of the *uplink consumption coefficient matrix* of $\mathbb{S}$, denoted $A^\uparrow = [a_{ij}^\uparrow]_{|\mathbb{S}|\times|\mathbb{S}|}$ is given by

$$a_{ij}^\uparrow = \frac{1}{\beta_j\rho_j} p_{ji}\beta_i\rho_i \tag{6}$$

THEOREM 1. *Let $\boldsymbol{x}^\uparrow = [x_i^\uparrow]_{|\mathbb{S}|\times 1}$ denote the uplink bandwidth demand vector of $\mathbb{S}$, and $\boldsymbol{c}^\uparrow = [c_i^\uparrow]_{|\mathbb{S}|\times 1}$ denote the self-initiated demand vector of $\mathbb{S}$, then when the network is in equilibrium (meaning that each service is given the amount of bandwidth it requires to run without delay) the following equation holds*

$$\underbrace{\boldsymbol{x}^\uparrow}_{\text{uplink cost}} = \underbrace{A^\uparrow\boldsymbol{x}^\uparrow}_{\text{relayed uplink demand}} + \underbrace{\boldsymbol{c}^\uparrow}_{\text{self-initiated demand}} \tag{7}$$

PROOF. From our earlier discussion in 3.2, we know that the send (uplink) action of a service s_i is triggered by two sources, namely self-initiated and relayed. With $c_i^{\uparrow}$ defined in (5), let $h_{ji}^{\uparrow}$ denote the uplink demand that is relayed from s_j to s_i, i.e., when s_j immediately precedes s_i in an application workflow. Therefore

$$x_i^{\uparrow} = \sum_j h_{ji}^{\uparrow} + c_i^{\uparrow} \tag{8}$$

With (3) we derive that each run of s_j and s_i is to generate data of size $\beta_j\rho_j$ and $\beta_i\rho_i$ respectively. If service s_j were to be allocated an uplink bandwidth of $x_j^{\uparrow}$, as an equilibrium entails, s_j would execute $x_j^{\uparrow}/\beta_j\rho_j$ times. From the communication probability matrix P, we know that for every one run of s_j there is a probability p_{ji} a subsequent run of s_i is triggered. Therefore we have

$$h_{ji}^{\uparrow} = \frac{x_j^{\uparrow}}{\beta_j\rho_j} p_{ji}\beta_i\rho_i \overset{(8)}{\Rightarrow} x_i^{\uparrow} = \sum_j \frac{x_j^{\uparrow}}{\beta_j\rho_j} p_{ji}\beta_i\rho_i + c_i^{\uparrow} \tag{9}$$

Consider $i \in \{1, 2, \cdots, |\mathbb{S}|\}$, (9) derives the same set of equations as given by taking (6) into (7). □

Definition 2. For each pair of services $\{s_i, s_j\} \in \mathbb{S}^2$, the elements of the *downlink consumption coefficient matrix* of $\mathbb{S}$, denoted $A^{\downarrow} = [a_{ij}^{\downarrow}]_{|\mathbb{S}|\times|\mathbb{S}|}$ is given by

$$a_{ij}^{\downarrow} = \eta_{ji} = \frac{p_{ji}\omega_{ji}}{\sum_k p_{jk}\omega_{jk}}, \quad s_k \in \mathbb{S} \tag{10}$$

THEOREM 2. *Let $\boldsymbol{x}^{\downarrow} = [x_i^{\downarrow}]_{|\mathbb{S}|\times 1}$ denote the downlink bandwidth demand vector of $\mathbb{S}$, then when the network is in equilibrium (meaning that each service is given the amount of bandwidth it requires to run without delay) the following equation holds*

$$\underbrace{\boldsymbol{x}^{\downarrow}}_{\textit{downlink cost}} = \underbrace{A^{\downarrow}\boldsymbol{x}^{\uparrow}}_{\textit{relayed downlink demand}} \tag{11}$$

PROOF. It is easy to understand that within the MSON, the downlink cost is totally dependent on the uplink cost in the sense that no receive action is required if no data was sent, and that all data sent by a service in context of the MSON must be received by another service of the MSON. On this basis, let $h_{ji}^{\downarrow}$ denote the downlink cost relayed from data sent from s_j to s_i, i.e., the amount of data sent from s_j to s_i, and we have

$$x_i^{\downarrow} = \sum_j h_{ji}^{\downarrow} \tag{12}$$

Recall from (4) that the probability that a unit of data sent by s_i to s_j is given by η_{ij}, we derive

$$h_{ji}^{\downarrow} = x_j^{\uparrow}\eta_{ji} \overset{(12)}{\Rightarrow} x_i^{\downarrow} = \sum_j x_j^{\uparrow} \frac{p_{ji}\omega_{ji}}{\sum_k p_{jk}\omega_{jk}}, \quad s_k \in \mathbb{S} \tag{13}$$

Similarly to the proof of theorem 1, by enumerating (13) with $i \in \{1, 2, \cdots, |\mathbb{S}|\}$, we get the same set of equations as given by taking (10) into (11). □

To conclude the network I-O model, we gather the per-service cost from both markets and derive the total bandwidth cost for a host device $m \in \mathbb{M}$ as

$$b_m = b_m^{\uparrow} + b_m^{\downarrow} = \sum_i x_i^{\uparrow} + \sum_i x_i^{\downarrow} = \sum_i x_i, \quad \Theta(s_i) = m \tag{14}$$

with b_m, $b_m^{\uparrow}$ and $b_m^{\downarrow}$ denote the total, uplink and downlink bandwidth cost of m.

4. SIMULATIONS

In this section, we conduct a series of simulation studies based on two types of service topologies: centralised and chain (illustrated by w_2 and w_3 in Fig. 2) to demonstrate the basic dynamics of the network I-O model. We assume a service-to-mobile allocation scheme given as $\Theta(s_1) = \Theta(s_2) = M_1$, $\Theta(s_3) = \Theta(s_4) = M_2$ and $\Theta(s_5) = M_3$ in both sets of experiments.

Effect of Service Arrival Rate: The service arrival rate is a key QoS metric for mobile application workflows. In this set of simulations, we demonstrate the dynamics of the network I-O model by examining the effect of increase in $\boldsymbol{\lambda}$ on the bandwidth costs of all services in $\mathbb{S}$. Furthermore, we map each service to a mobile device and examine the effect of the same action on each device's total bandwidth requirement.

In a centralised topology (w_2), we identify s_5 to be the core service and gradually increase λ_5 from 20 to 40. Results as illustrated in the first row of Fig. 3 show that the increased traffic is evenly relayed to the downlink bandwidth cost of the other services (due to the service topology), and because the traffic relayed back from the leaf services are less significant (due to the communication pattern), M_3 which hosts s_5 does not require great increase in downlink capacity.

In a chain topology (w_3), we identify the head service s_1 to be the core service and increase λ_1 to double its initial value. As shown in the second row of Fig. 3, s_1 itself does not demand much extra bandwidth since its succeeding service is located on the same device (M_1). This co-location factor also explains why only $x_2^{\uparrow}$ and $x_4^{\uparrow}$ is showing an increase in the first plot and $x_3^{\downarrow}$ and $x_5^{\downarrow}$ in the second plot. When these values are summarised per device, b_2 shows the greatest increase because it has to accommodate both the increase in $x_3^{\downarrow}$ and $x_4^{\uparrow}$.

Effect of Per Service Data Size: In this set of simulations, we examine the effect of increase in per request data size (i.e., $\boldsymbol{\beta}$). In practice, this can be observed when the per frame resolution of a video stream from one user to the other is changed. As shown in the third row of Fig. 3, as we increase β_5, both $x_5^{\uparrow}$ and $b_5^{\uparrow}$ increase as they do in the first row of Fig. 3. However, $x_5^{\downarrow}$ remains unchanged. Furthermore, the uplink bandwidth demand of all other services and their hosts remain unchanged. This is because the increase in β_5 does not affect the relayed uplink bandwidth of the service which is called by s_5, therefore the effect of increase in $\boldsymbol{\beta}$ is more confined within the MSON than that in $\boldsymbol{\lambda}$. The same can be observed from the fourth row of Fig. 3 which illustrates the result from a chain topology (increase in β_1).

Alternative Allocation Scheme: One common bandwidth allocation scheme, as an alternative scheme to our network I-O model, evenly distributes the available bandwidth to the services it hosts. As a result, the service rate of an MSON is prematurely capped by the service which requires the most amount of bandwidth as shown by λ' of Fig. 3 (zoom). It can be seen that the scheme as given by the I-O model, capped at λ'', realise greater potential from the MSON.

5. CONCLUSION

This work extends the existing I-O model in economics to model the service bandwidth allocation problem in mobile service-oriented networks (MSON)s. A network I-O model is

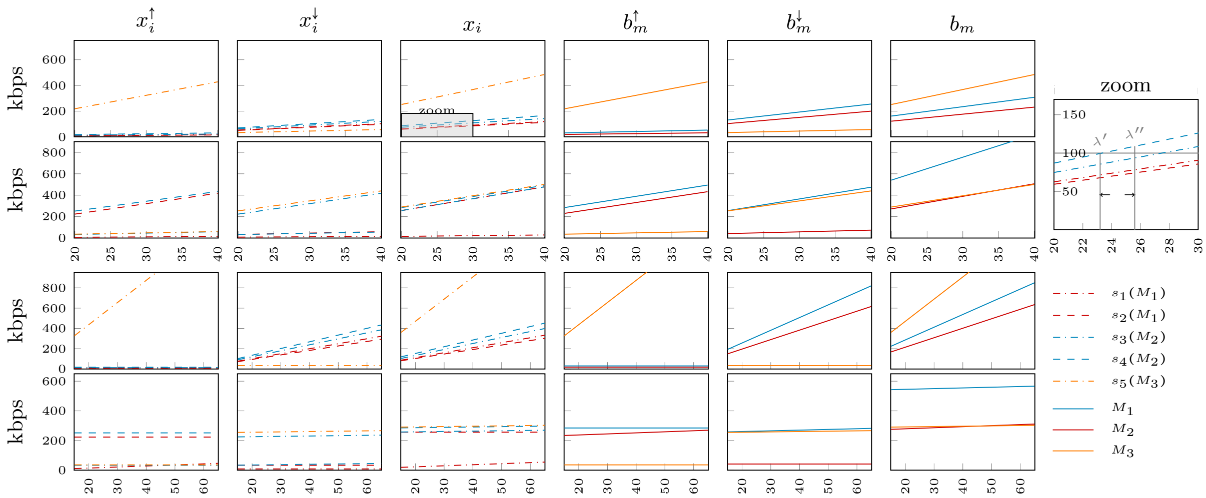

Figure 3: The increases in λ_5 of w_2, λ_1 of w_3, β_5 of w_2 and β_1 of w_3 are projected onto the x-axis of each of the four rows of plots respectively. The effects of these increases including the uplink ($x_i^{\uparrow}$), downlink ($x_i^{\downarrow}$) and total (x_i) bandwidth demands of each service (s_1 to s_5 as in plot legends), and the uplink ($b_m^{\uparrow}$), downlink ($b_m^{\downarrow}$) and total (b_m) bandwidth demands of each device (M_1 to M_3 as in plot legends) are presented in each of the six columns of plots respectively.

constructed to describes the bandwidth dependency and allocation problem in mobile networks. Elements of an MSON are considered as economic entities with their interdependencies (in terms of bandwidth demand and service QoS) as the underlying structure of the network economy. The network I-O model lays the foundation for future objective developments in ubiquitous mobile computing scenarios. For future work, we would like to extend the application of our network I-O model to include scenarios in which dynamic service allocation schemes are implemented in the network and eliminate the assumption of given service-to-device mappings. The network I-O model presented in this paper would be an essential instrument to optimise the mapping strategies in such scenarios.

Acknowledgement

This work is sponsored by the Research Project Grant of the Leverhulme Trust (Grant No. RPG-101).

6. REFERENCES

[1] Alrifai, M., and Risse, T. Combining global optimization with local selection for efficient QoS-aware service composition. In *Proceedings of the 18th International Conference on World Wide Web, WWW* (2009).

[2] Ardagna, D., and Pernici, B. Adaptive service composition in flexible processes. *IEEE Transactions on Software Engineering 33*, 6 (2007), 369–384.

[3] Bodík, P., Menache, I., Chowdhury, M., Mani, P., Maltz, D. A., and Stoica, I. Surviving failures in bandwidth-constrained datacenters. In *Proceedings of ACM SIGCOMM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication, SIGCOMM* (2012).

[4] Chang, C., Srirama, S. N., and Ling, S. An adaptive mediation framework for mobile p2p social content sharing. In *Proceedings of the 10th International Conference on Service-Oriented Computing, ICSOC* (2012).

[5] Chen, Y., Jain, S., Adhikari, V. K., Zhang, Z.-L., and Xu, K. A first look at inter-data center traffic characteristics via Yahoo! datasets. In *Proceedings of IEEE INFOCOM* (2011).

[6] Cuff, D., Hansen, M., and Kang, J. Urban sensing: Out of the Woods. *Communications of the ACM 51*, 3 (2008).

[7] Ferreira, M., Fernandes, R., Conceição, H., Viriyasitavat, W., and Tonguz, O. K. Self-organized traffic control. In *Proceedings of the seventh ACM international workshop on VehiculAr InterNETworking, VANET* (2010).

[8] Gob, A., Schreiber, D., Hamdi, L., Aitenbichler, E., and Muhlhauser, M. Reducing user perceived latency with a middleware for mobile SOA access. In *Proceedings of IEEE International Conference on Web Services,* (2009).

[9] Groba, C., and Clarke, S. Opportunistic composition of sequentially-connected services in mobile computing environments. In *Proceedings of IEEE International Conference on Web Services, ICWS* (2011).

[10] Huang, C.-M., Hsu, T.-H., and Hsu, M.-F. Network-aware P2P file sharing over the wireless mobile networks. *IEEE Journal on Selected Areas in Communications 25*, 1 (2007).

[11] Liu, S., and Striegel, A. D. Exploring the potential in practice for opportunistic networks amongst smart mobile devices. In *Proceedings of 19th International Conference on Mobile Computing & Networking, MobiCom* (2013).

[12] Mabrouk, N. B., Beauche, S., Kuznetsova, E., and Georgantas, N. QoS-aware service composition in dynamic service oriented environments. In *Proceedings of ACM/IFIP/USENIX 10th International middleware Conference, Middelware* (2009).

[13] Natchetoi, Y., Wu, H., and Zheng, Y. Service-Oriented mobile applications for ad-hoc networks. In *Proceedings of IEEE Conference on Services Computing, SCC* (2008).

[14] Park, J.-S., Lee, U., Oh, S. Y., Gerla, M., and Lun, D. S. Emergency related video streaming in VANET using network coding. In *Proceedings of the 3rd International Workshop on Vehicular Ad Hoc Networks* (2006).

[15] Shi, J., Zhang, R., Liu, Y., and Zhang, Y. PriSense: Privacy-Preserving Data Aggregation in People-Centric Urban Sensing Systems. In *Proceedings of IEEE INFOCOM* (2010).

[16] Viswanathan, H., Lee, E. K., and Pompili, D. Enabling real-time in-situ processing of ubiquitous mobile-application workflows. In *Proceedings of IEEE International Conference on Mobile Ad-Hoc and Sensor Systems, MASS* (2013).

[17] Yu, T., Zhang, Y., and Lin, K.-J. Efficient algorithms for web services selection with end-to-end QoS constraints. *ACM Transactions on the Web 1*, 1 (2007).