Relation-Oriented Lattice Model for Resource Allocation in Heterogenous Distributed Systems

Teng Yu
Department of Computing
Imperial College London

Jan. 2016

Contents

Resource Allocation Model

State-Of-The-Art

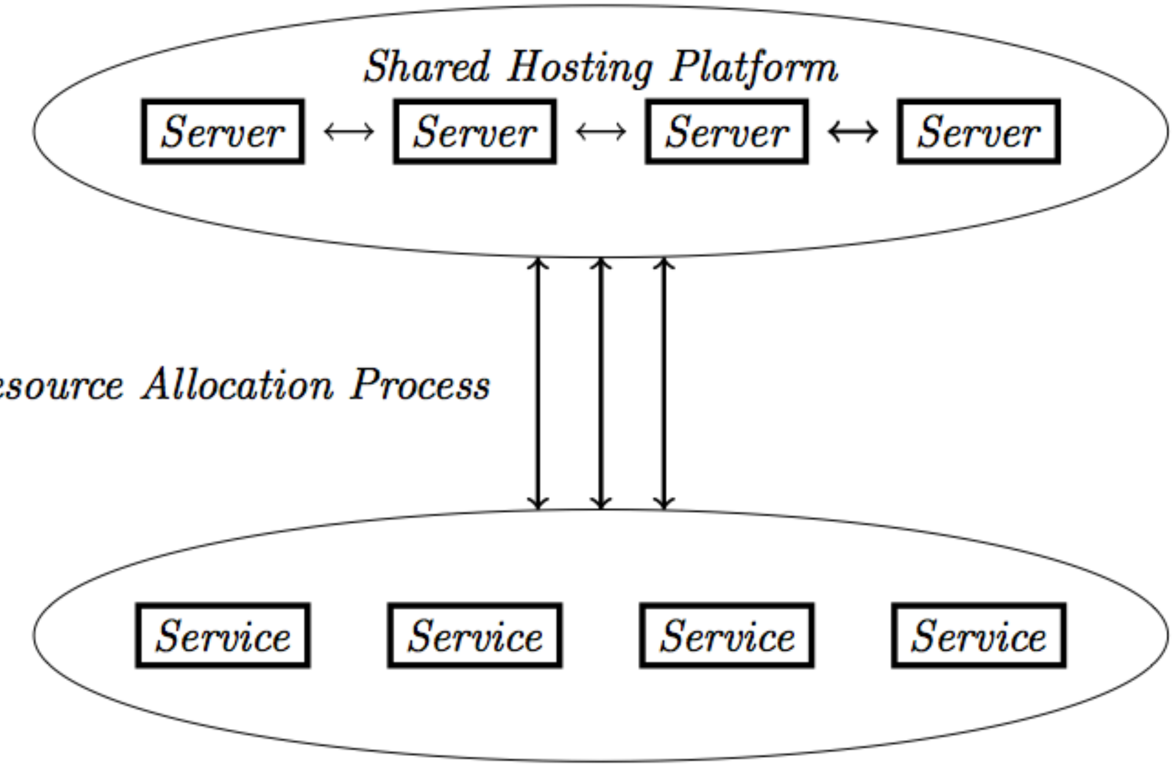
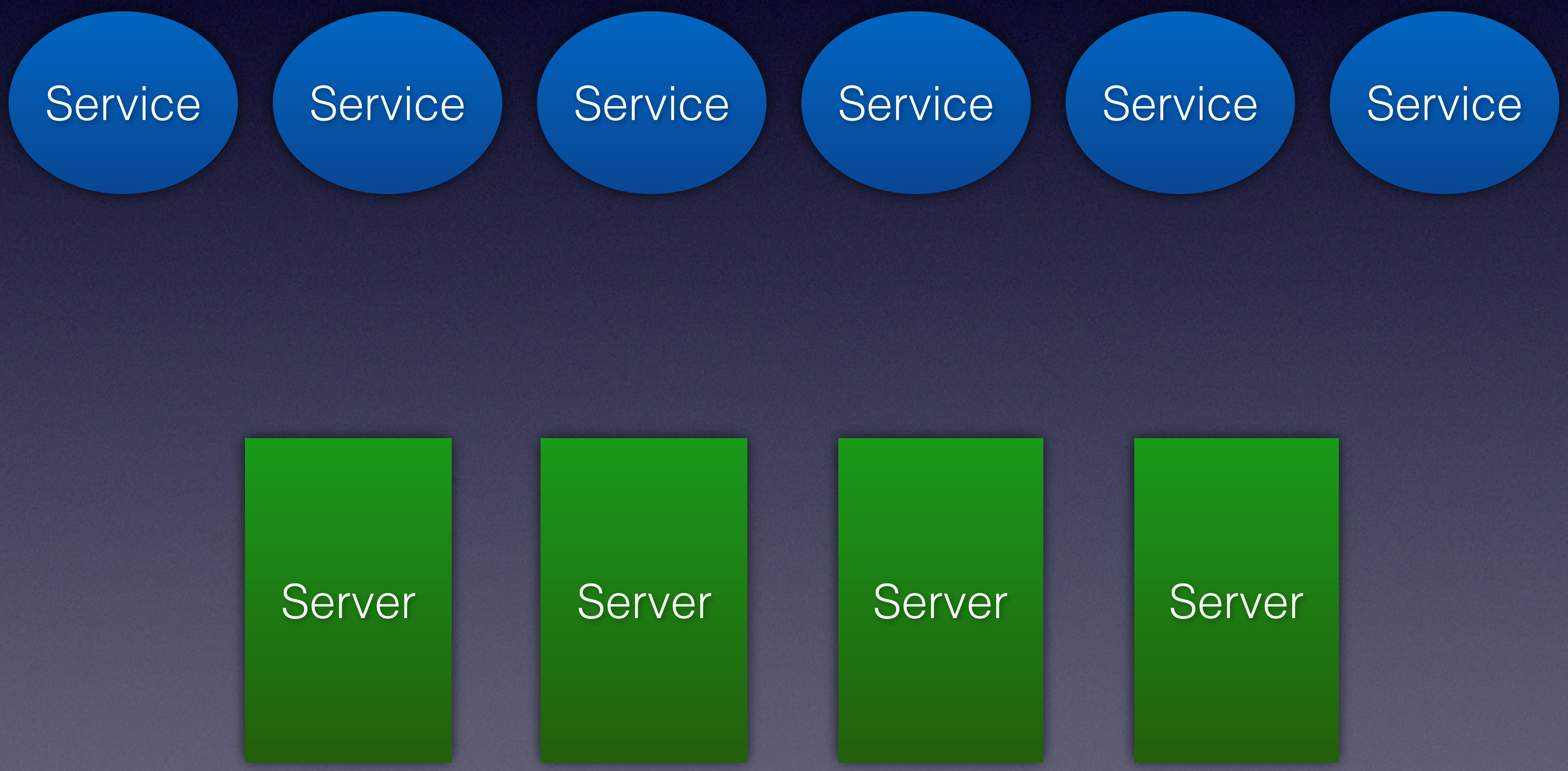


Figure 1: System Model

State Of The Art

- Bin-Packing Model



- Mixed Integer Linear Program Solver

State Of The Art

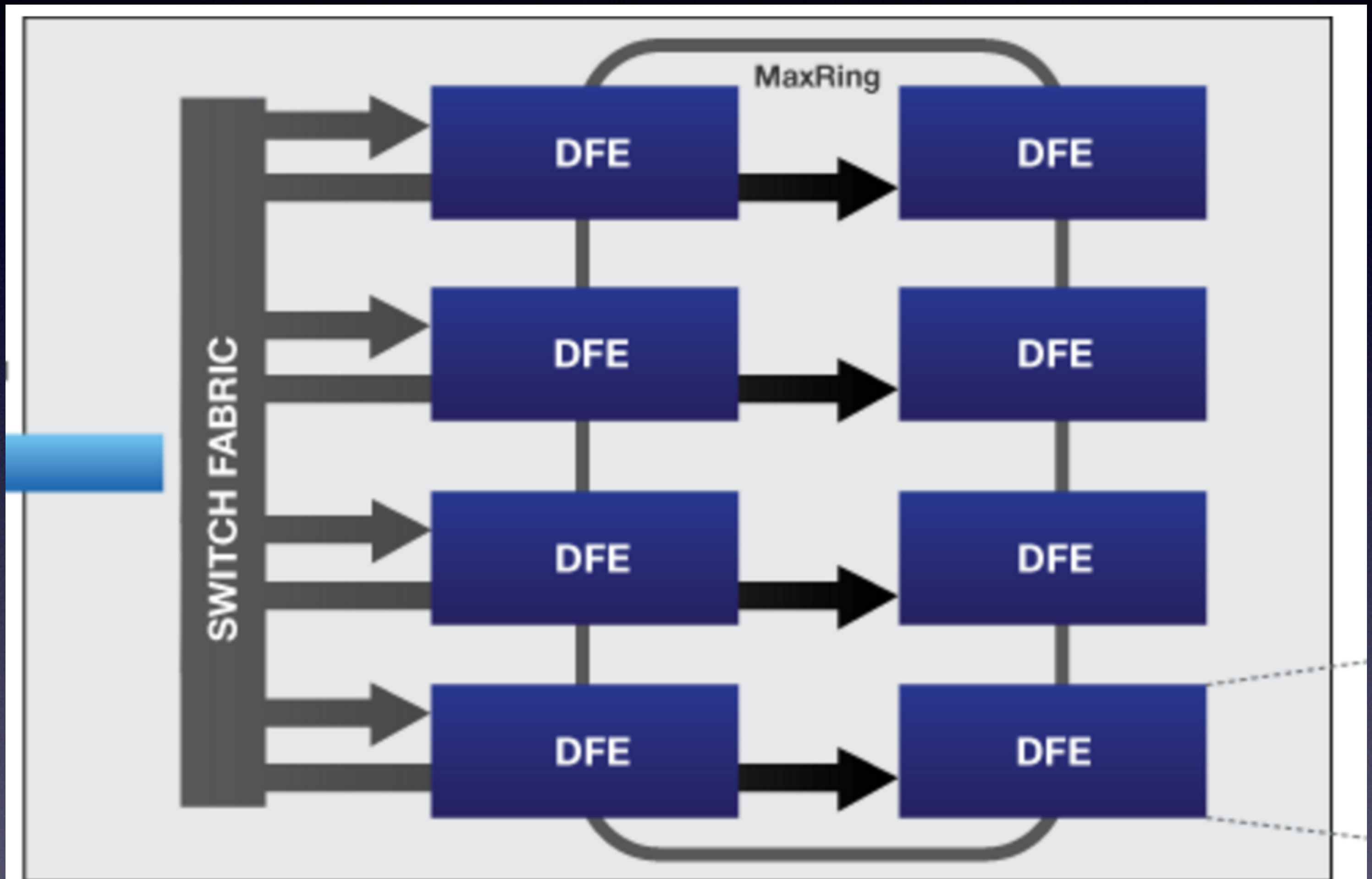
- Problems:
- Non-independent Resource Allocation Requests(RArs). e.g: shared-resource needs (Network Link), related needs (I/O, Network)
- Multi-Servers RArs

Thoroughly increase the Complexity for bin-packing model !!

Consider the problems on a concrete practical case…

MPC-X

A Concrete Example of High Performance Heterogeneous Distributed System



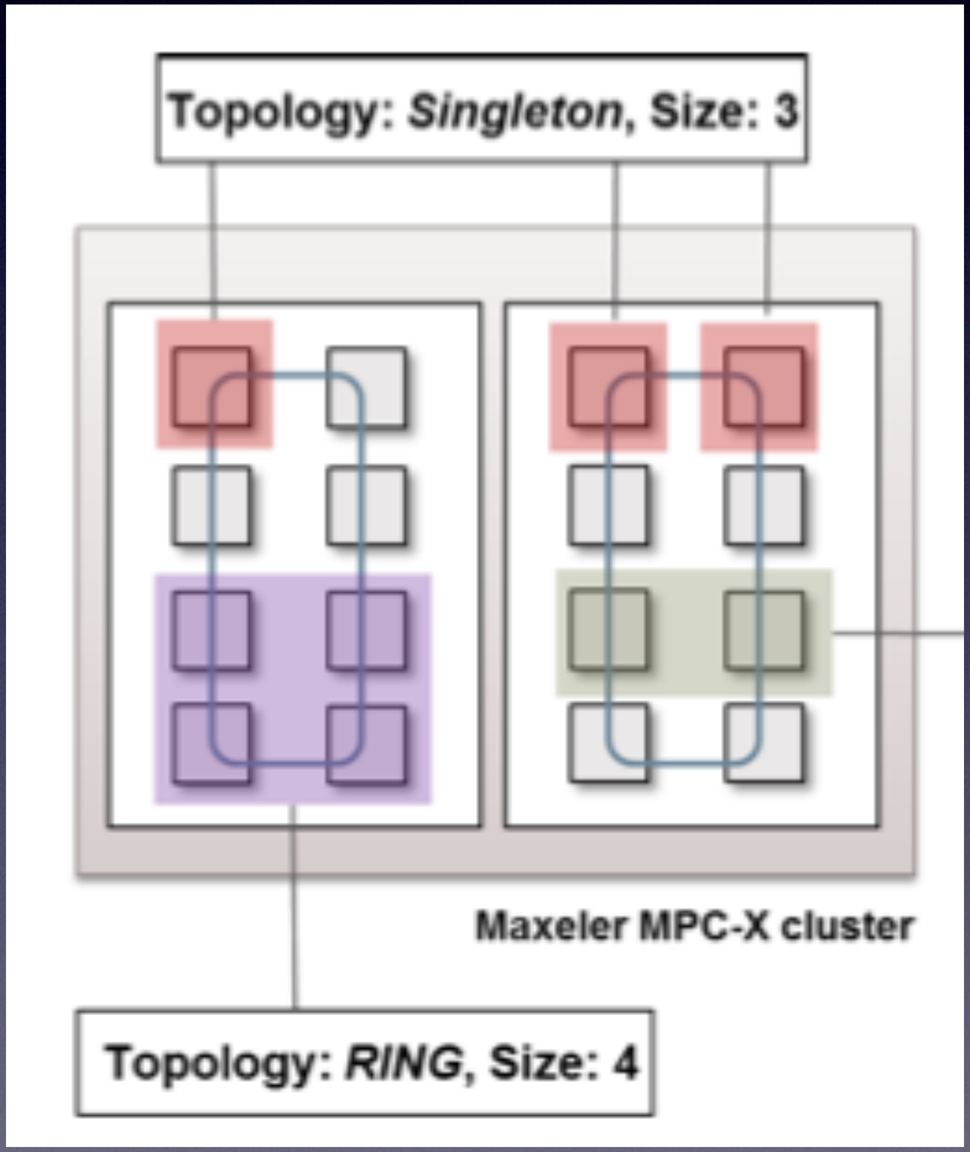
MPC-X series node architecture

- As refer to the system architecture from the Maxeler Technologies Official Website:
- MPC-X contains a cluster of Maxeler Dataflow Engines(DFE) and each of them interconnected via a ring topology.
- MPC-X provides large memory capacities, enable remote assess, redundant power supplies, lights-out management support and ***powerful computing performance***

MPC-X

A Concrete Example of High Performance Heterogeneous Distributed System

- As refer to the topology graph shown in FP7 HARNESS technical paper, there are at least two types of RArs for MPC-X device:
- Singleton RArs: Only need certain amount of DFEs to work on it, don't care about the physical positions of them in the device;
- Adjacent RArs: Need certain amount of neighboured DFEs to work on it
- We say those two types of RArs are ***non-independent*** and contain ***multi-servers*** feature



Instead of based on the traditional bin-packing model and linear programming solver, we provide a novel approach by focusing on ***relations between RArs*** and intend to generate a ***partially ordered model*** with corresponding ***ranking functions*** to direct the allocation.

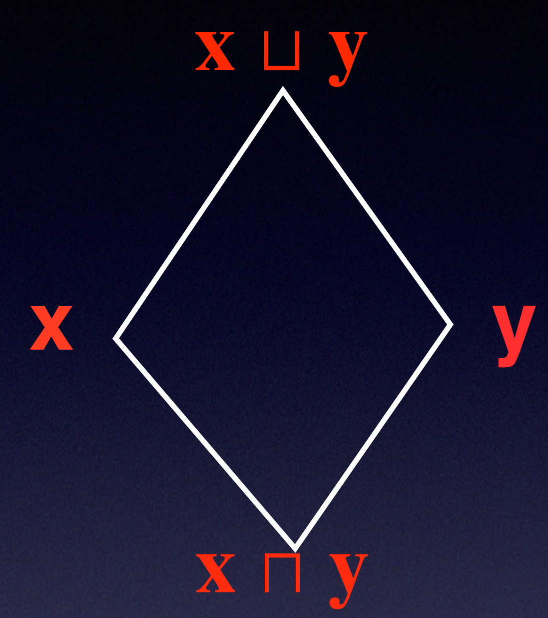
Order and Lattice theory

Definition 1: Let P be a set. An partial order on P is a binary relation ≤ on P such that,∀x,y,z∈P,
(i) $x \leq x$, (ii) $x \leq y$ and $y \leq x$ imply $x = y$, (iii) $x \leq y$ and $y \leq z$ imply $x \leq z$.

Definition 2: A set P equipped with an order relation ≤ is said to be an partially ordered set. We use the shorthand poset in this paper.

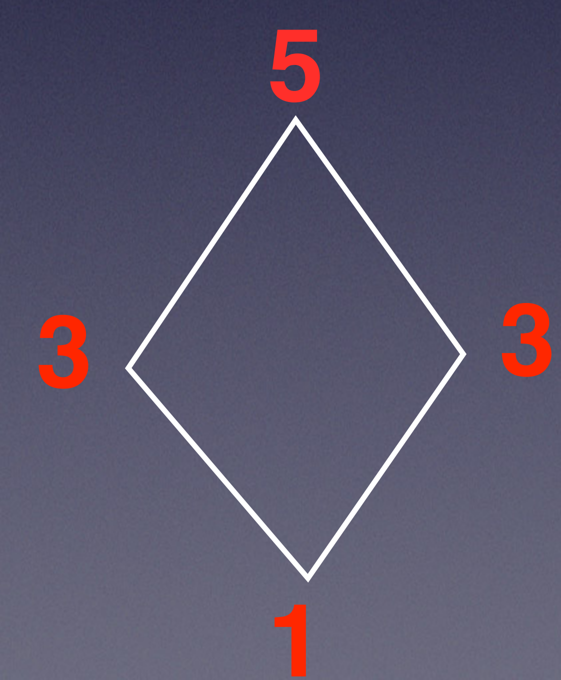
Order and Lattice theory

Definition3: Let L be a non-empty poset. If $x \sqcup y$ and $x \sqcap y$ exist for all $x, y \in L$, then L is called a ***Lattice***.



Definition4: A function f on a lattice L, $f : L \to R^{+}_{0}$ is a ***valuation*** iff

$\forall x,y \in L. f(x \sqcap y)+f(x \sqcup y)=f(x)+f(y).$



RArs Relation

Definition5: Given the relation $\leq$ between RArs:
$\forall \alpha \in \textit{cluster device}, \exists A,B \in \textit{RAr, such that } A \leq B \Leftrightarrow \alpha(B) \rightarrow \alpha(A).$

We say $\alpha(B)$ is true iff α can serve B. We say $A \preccurlyeq B$ iff $A \leq B$ and there is no interim nodes between A and B. Use $\perp$ to represent the non-resource need RAr.

Definition6: Given A^n denote a n-dimensional RAr, then:

$A^n = (\forall Ai \ i \in n) \bigwedge Ai.$

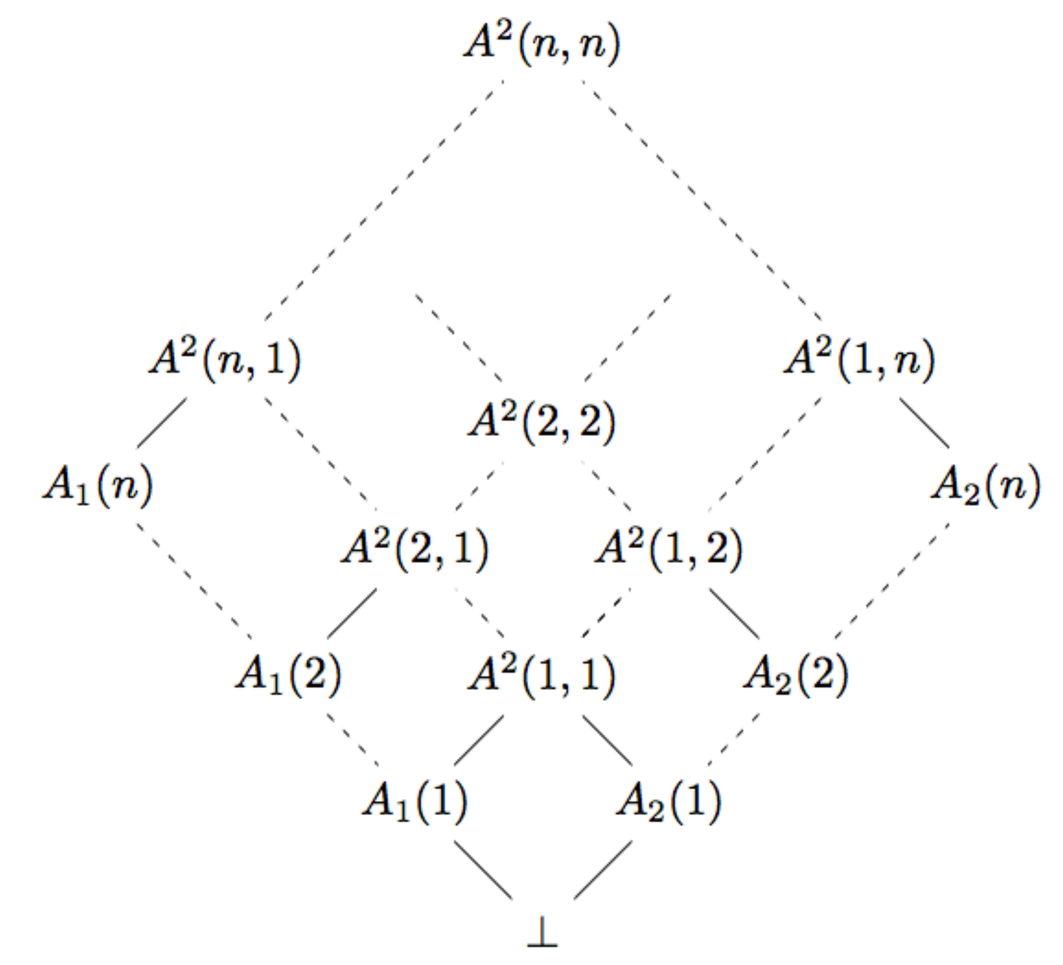


Figure 3: 2-dimensional(CPU, Memory) RArs topology

Ranking on Basic RArs Model

Define a ranking associated with each vector which is equal to the **sum of coordinates** in the vector space

Definition7:

$\forall A^2 \in$ a 2 dimensional modular structure, a ranking $R(A^2) = |A1| + |A2|$

Or considering the **height function**:

Definition8:

$\forall x,y \in$ a 2 dimensional modular structure, a height H is a function that :

(1)$H(x)=H(y)+1 \Leftrightarrow y \lessdot x$;

(2)$H(x)=0 \Leftrightarrow x=\perp$.

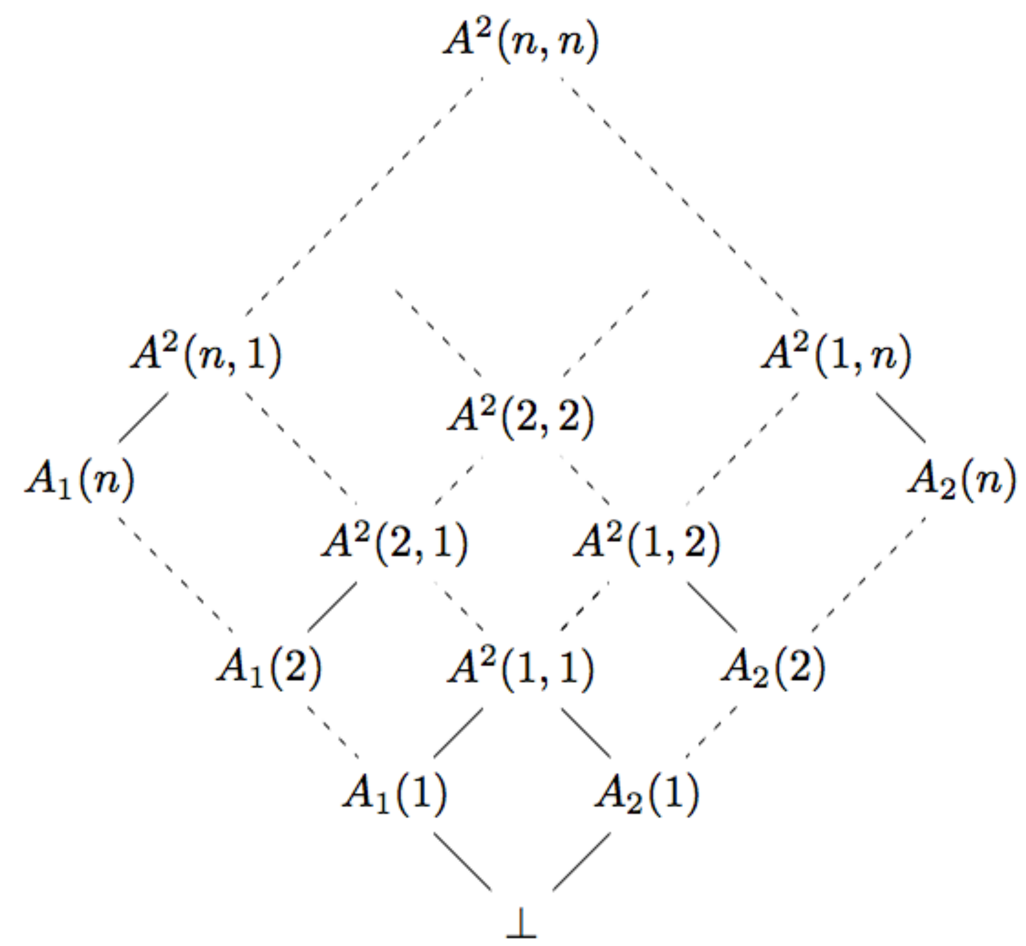
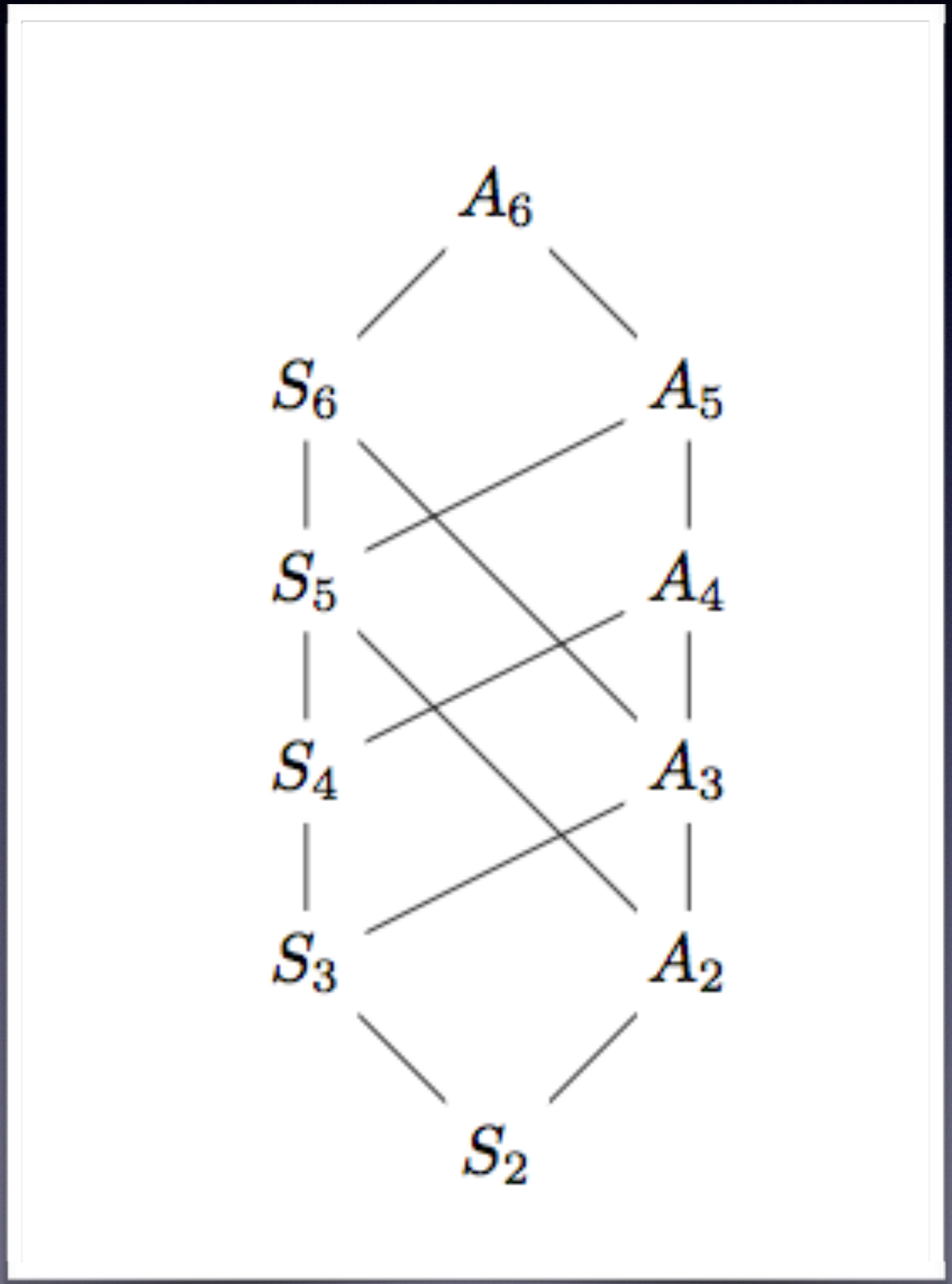


Figure 3: 2-dimensional(CPU, Memory) RArs topology

MPC-X initial RAr Model

- We use S_n (n=1…8) to represent singleton RArs; A_n (n=1…8) to represent adjacent RArs. The index represent the number of servers it requested. Then it is easy to say the foundational relations below:

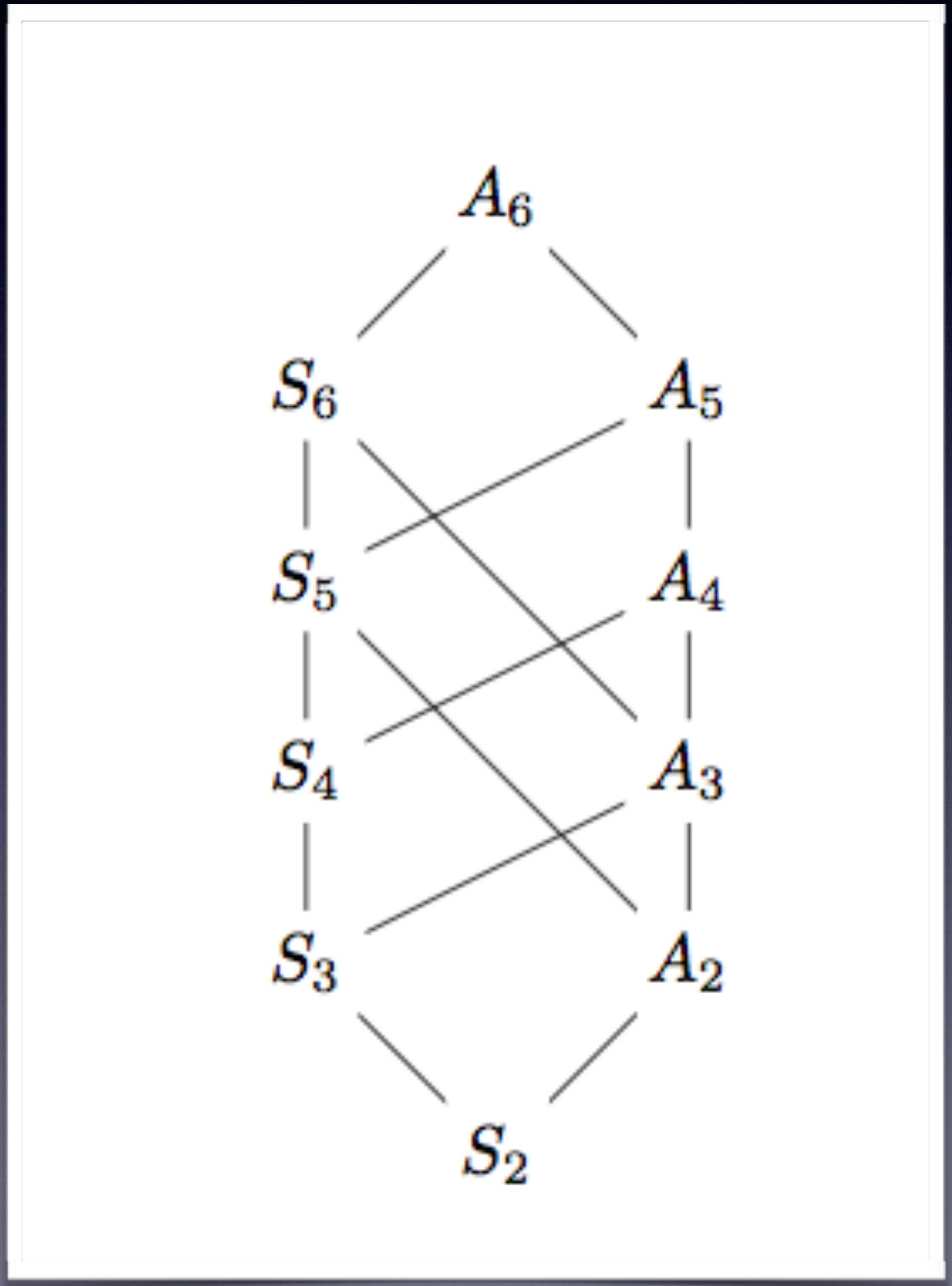
1) $\forall i \in (1,...,n-1), S_i \preccurlyeq S_{i+1}$
2) $\forall i \in (1,...,n-1), A_i \preccurlyeq A_{i+1}$
3) $\forall i \in (1,...,n), S_i \preccurlyeq A_i$
4) $A_1 = S_1$
5) $A_7 = S_7$
6) $A_8 = S_8$
7) $A_2 \preccurlyeq S_5$
8) $A_3 \preccurlyeq S_6$



Bottleneck of ranking on MPC-X initial RArs model

In a general RArs model containing some non-diamond substructure, we cannot rank it directly by neither sum of vectors coordinates or a height function.

For example, we cannot define the height value of S5 because two nodes, S4 and A2, in different level are both under it immediately.



Consider to transfer the general initial structure to a modular lattice structure by representation…

Birkhoff's Representation Theory

Definition9: Let P be a poset and let $S \subseteq P$. We say S is a ***downset*** of P iff $\forall x \in S$ and $y \leq x$, then $y \in S$.

Definition10: Given a lattice L. An element $x \in L$ is ***join-irreducible*** iff: $x=a \sqcup b$ implies $x=a$ or $x=b$ for all $a,b \in L$

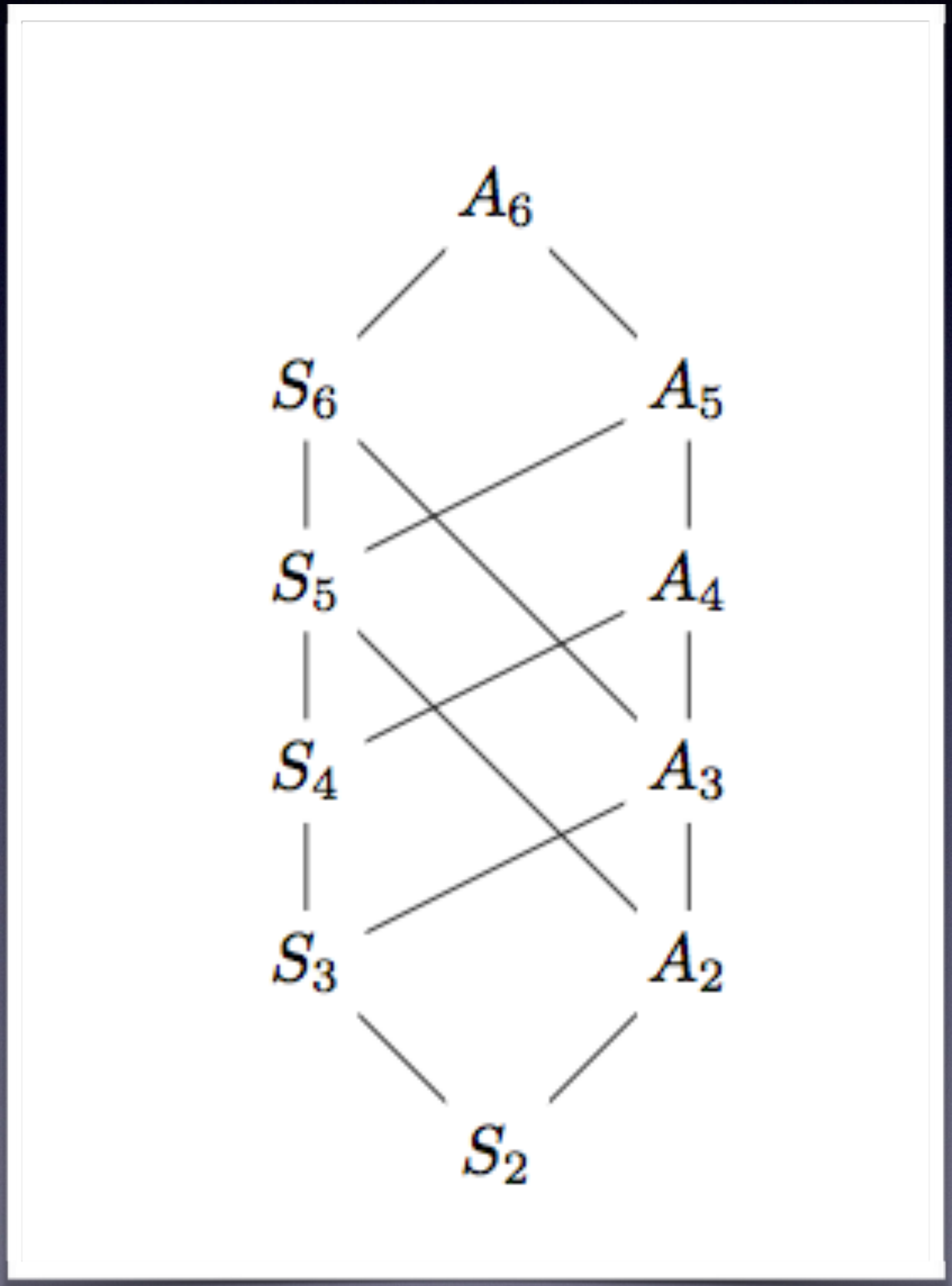
Definition11: A lattice $L = (P, \sqsubseteq)$ is ***modular*** iff

$\forall x,y,z \in L . x \sqsubseteq z \Rightarrow x \sqcup (y \sqcap z)=(x \sqcup y) \sqcap z$.

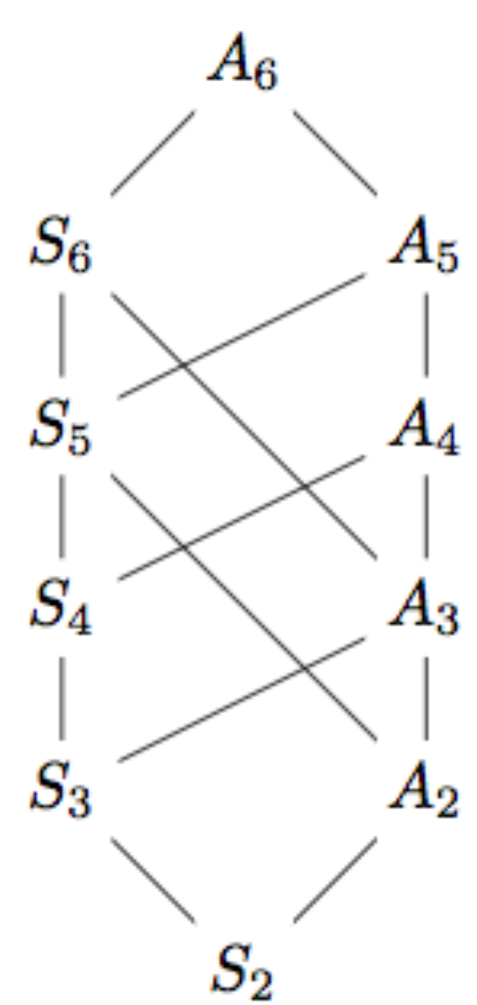
Definition12: (***Birkhoff's representation theory***) Any finite modular lattice L is isomorphic to the lattice of downsets of the partial order of the join-irreducible elements of L.

Apply Birkhoff's Representation on MPC-X initial RArs model

- I. Find out all the possible downsets of independent pairs in the initial RArs model
- II. Order those downsets by inclusion relation to construct the result modular lattice

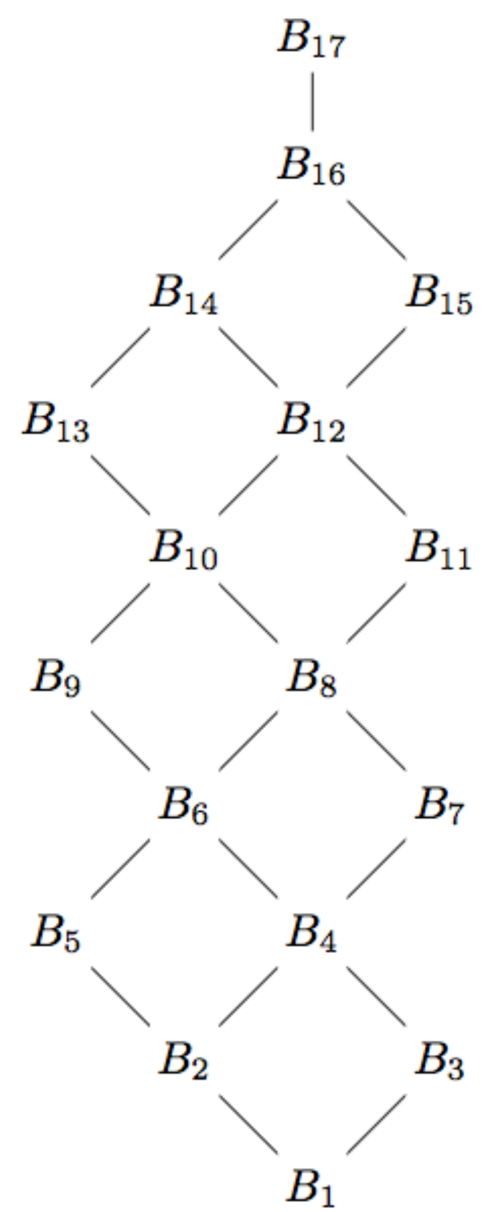


Apply Birkhoff's Representation on MPC-X initial RArs model



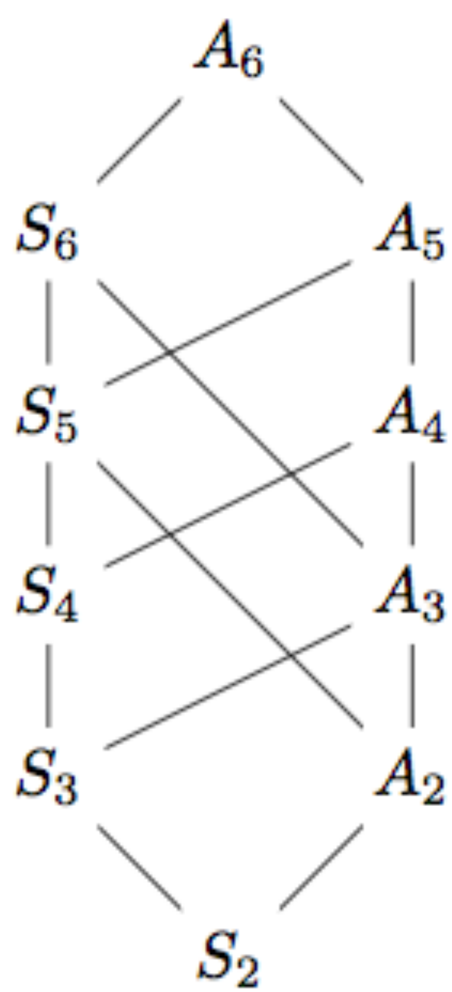
Node in lattice-topology	Corresponding Downset	Containing RArs
B_1	$S_2\downarrow$	$\{S_2\}$
B_2	$S_3\downarrow$	$\{S_3,S_2\}$
B_3	$A_2\downarrow$	$\{A_2, S_2\}$
B_4	$(S_3 \cup A_2)\downarrow$	$\{A_2, S_3, S_2\}$
B_5	$S_4\downarrow$	$\{S_4, S_3, S_2\}$
B_6	$(S_4 \cup A_2)\downarrow$	$\{A_2, S_4, S_3, S_2\}$
B_7	$A_3\downarrow$	$\{A_3, A_2, S_3, S_2\}$
B_8	$(S_4 \cup A_3)\downarrow$	$\{A_3, A_2, S_4, S_3, S_2\}$
B_9	$S_5\downarrow$	$\{A_2, S_5, S_4, S_3, S_2\}$
B_{10}	$(S_5 \cup A_3)\downarrow$	$\{A_3, A_2, S_5, S_4, S_3, S_2\}$
B_{11}	$A_4\downarrow$	$\{A_4, A_3, A_2, S_4, S_3, S_2\}$
B_{12}	$(S_5 \cup A_4)\downarrow$	$\{A_4, A_3, A_2, S_5, S_4, S_3, S_2\}$
B_{13}	$S_6\downarrow$	$\{A_3, A_2, S_6, S_5, S_4, S_3, S_2\}$
B_{14}	$(S_6 \cup A_4)\downarrow$	$\{A_4, A_3, A_2, S_6, S_5, S_4, S_3, S_2\}$
B_{15}	$A_5\downarrow$	$\{A_5, A_4, A_3, A_2, S_5, S_4, S_3, S_2\}$
B_{16}	$(S_6 \cup A_5)\downarrow$	$\{A_5, A_4, A_3, A_2, S_6, S_5, S_4, S_3, S_2\}$
B_{17}	$A_6\downarrow$	$\{A_6, A_5, A_4, A_3, A_2, S_6, S_5, S_4, S_3, S_2\}$

Apply Birkhoff's Representation on MPC-X initial RArs model



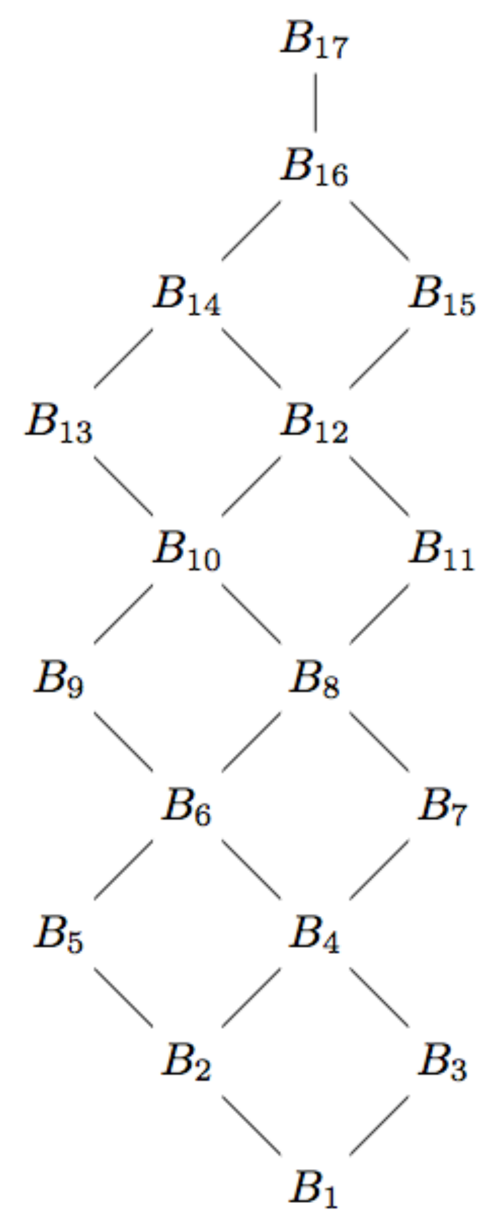
Node in lattice-topology	Corresponding Downset	Containing RArs
B_1	$S_2\downarrow$	$\{S_2\}$
B_2	$S_3\downarrow$	$\{S_3,S_2\}$
B_3	$A_2\downarrow$	$\{A_2, S_2\}$
B_4	$(S_3 \cup A_2)\downarrow$	$\{A_2, S_3, S_2\}$
B_5	$S_4\downarrow$	$\{S_4, S_3, S_2\}$
B_6	$(S_4 \cup A_2)\downarrow$	$\{A_2, S_4, S_3, S_2\}$
B_7	$A_3\downarrow$	$\{A_3, A_2, S_3, S_2\}$
B_8	$(S_4 \cup A_3)\downarrow$	$\{A_3, A_2, S_4, S_3, S_2\}$
B_9	$S_5\downarrow$	$\{A_2, S_5, S_4, S_3, S_2\}$
B_{10}	$(S_5 \cup A_3)\downarrow$	$\{A_3, A_2, S_5, S_4, S_3, S_2\}$
B_{11}	$A_4\downarrow$	$\{A_4, A_3, A_2, S_4, S_3, S_2\}$
B_{12}	$(S_5 \cup A_4)\downarrow$	$\{A_4, A_3, A_2, S_5, S_4, S_3, S_2\}$
B_{13}	$S_6\downarrow$	$\{A_3, A_2, S_6, S_5, S_4, S_3, S_2\}$
B_{14}	$(S_6 \cup A_4)\downarrow$	$\{A_4, A_3, A_2, S_6, S_5, S_4, S_3, S_2\}$
B_{15}	$A_5\downarrow$	$\{A_5, A_4, A_3, A_2, S_5, S_4, S_3, S_2\}$
B_{16}	$(S_6 \cup A_5)\downarrow$	$\{A_5, A_4, A_3, A_2, S_6, S_5, S_4, S_3, S_2\}$
B_{17}	$A_6\downarrow$	$\{A_6, A_5, A_4, A_3, A_2, S_6, S_5, S_4, S_3, S_2\}$

Apply Birkhoff's Representation on MPC-X initial RArs model

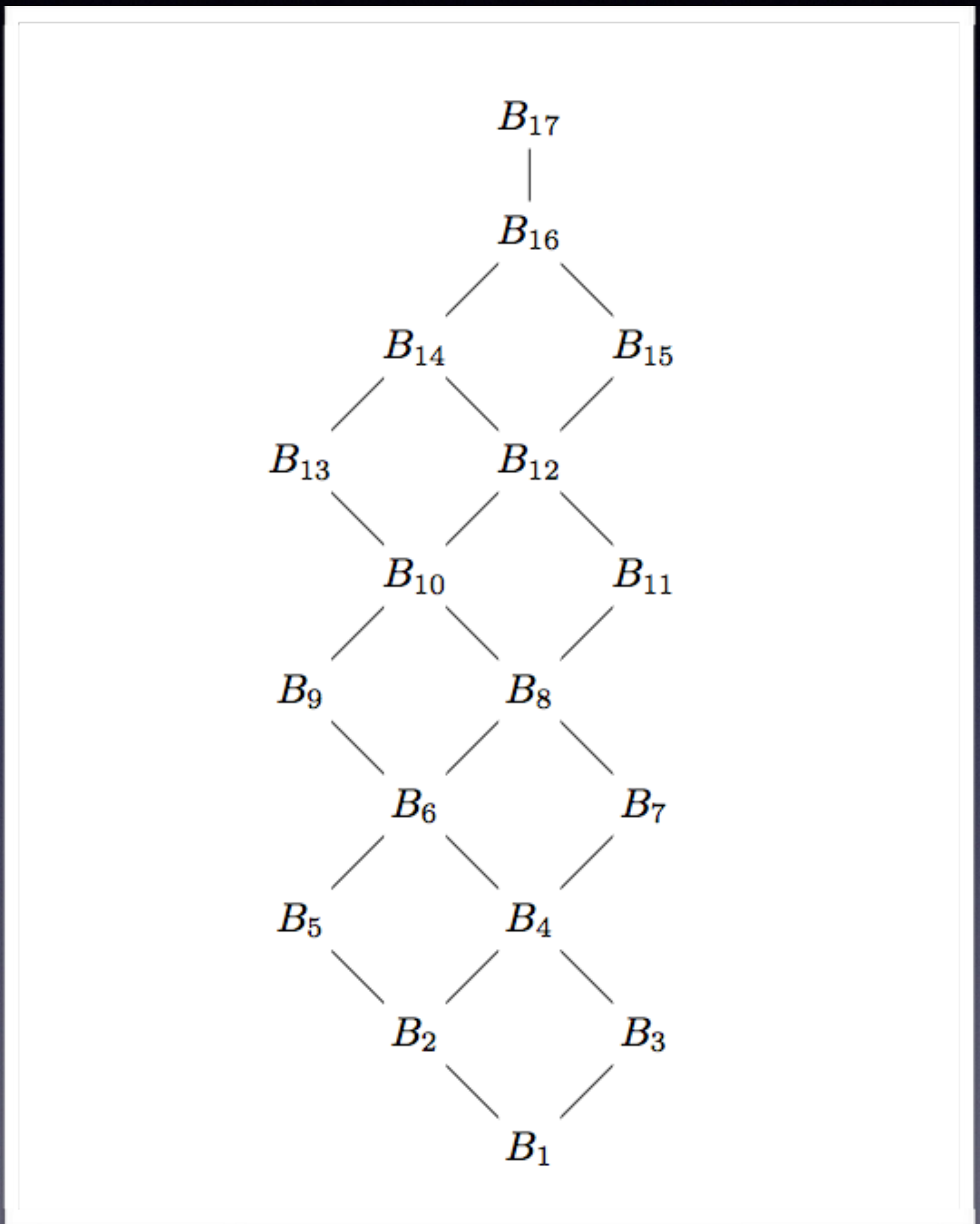


Node in lattice-topology	Corresponding Downset	Containing RArs
B_1	$S_2\downarrow$	$\{S_2\}$
B_2	$S_3\downarrow$	$\{S_3, S_2\}$
B_3	$A_2\downarrow$	$\{A_2, S_2\}$
B_4	$(S_3 \cup A_2)\downarrow$	$\{A_2, S_3, S_2\}$
B_5	$S_4\downarrow$	$\{S_4, S_3, S_2\}$
B_{10}	$(S_5 \cup A_3)\downarrow$	$\{A_3, A_2, S_5, S_4, S_3, S_2\}$
B_{11}	$A_4\downarrow$	$\{A_4, A_3, A_2, S_4, S_3, S_2\}$
B_{12}	$(S_5 \cup A_4)\downarrow$	$\{A_4, A_3, A_2, S_5, S_4, S_3, S_2\}$
B_{13}	$S_6\downarrow$	$\{A_3, A_2, S_6, S_5, S_4, S_3, S_2\}$
B_{14}	$(S_6 \cup A_4)\downarrow$	$\{A_4, A_3, A_2, S_6, S_5, S_4, S_3, S_2\}$
B_{15}	$A_5\downarrow$	$\{A_5, A_4, A_3, A_2, S_5, S_4, S_3, S_2\}$
B_{16}	$(S_6 \cup A_5)\downarrow$	$\{A_5, A_4, A_3, A_2, S_6, S_5, S_4, S_3, S_2\}$
B_{17}	$A_6\downarrow$	$\{A_6, A_5, A_4, A_3, A_2, S_6, S_5, S_4, S_3, S_2\}$

Transfer



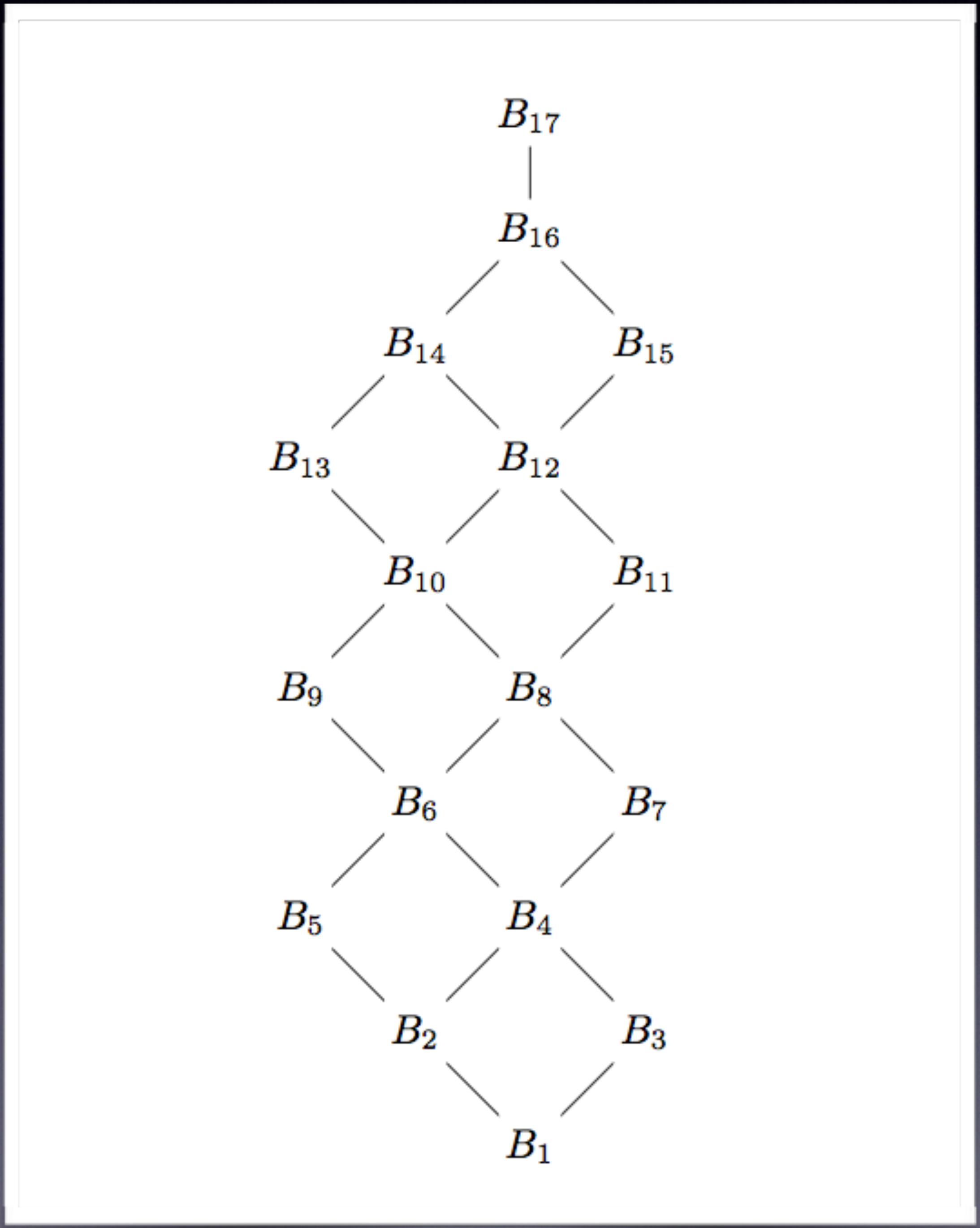
MPC-X RArs modular topology



Novel features for this model:

I. Completeness
II. Modular

Ranking on MPC-X RArs modular topology



Based on the modular feature, we can apply the same ***height function*** as mentioned in **definition7** to rank the model.

Recall **Definition7**:

$\forall$x,y $\in$ a 2 dimensional modular structure, a height H is a function that :

(1)H(x)=H(y)+1 $\Leftrightarrow$y$\leqslant$x;

(2)H(x)=0 $\Leftrightarrow$x=$\perp$.

Complexity Analysis

- For a finite poset, it is in polynomial time to implement the Birkhoff's representation to transfer to a modular lattice.
- A ranking function is easy to compute on any modular lattice in polynomial time complexity, or say within O(n) through a recursive search in which n is the amount of RArs.
- In conclusion, we can compute such a ranking on arbitrary allocation topology within polynomial time.

Evaluation

- We have implemented two prototype programs using Java-Choco at this step:
- The first one based on the traditional bin-packing model whilst the second one based on our relation-oriented model.
- Both of them compute the same tiny-input and solved by the same trivial searching algorithm as a benchmark.

Statistics Results by Trivial Model

No.	Building time	Initial propagation	Resolution	Nodes	Backtracks	Max depth	Constraints
1	0.0076s	0.013s	0.020s	4	2	3	5
2	0.0076s	0.014s	0.019s	4	2	3	5
3	0.0083s	0.010s	0.015s	4	2	3	5
4	0.0077s	0.012s	0.017s	4	2	3	5
5	0.0077s	0.011s	0.016s	4	2	3	5
6	0.0078s	0.011s	0.016s	4	2	3	5

Statistics Results by Novel Model

No.	Building time	Initial propagation	Resolution	Nodes	Backtracks	Max depth	Constraints
1	0.0053s	0.010s	0.016s	22	20	21	4
2	0.0052s	0.009s	0.016s	22	20	21	4
3	0.0051s	0.010s	0.017s	22	20	21	4
4	0.0047s	0.010s	0.017s	22	20	21	4
5	0.0065s	0.010s	0.018s	22	20	21	4
6	0.0054s	0.009s	0.016s	22	20	21	4

Future Evaluation Plan

- I. Implement our model for the MPC-X device.
- II. Implement the corresponding ranking function for the MPC-X topology
- III. Add the ranking function as a metric in different searching algorithms (such as First Fit and Best Fit) to direct the allocation and compare the performance with the traditional solver.
- Experimental dataset of RArs to MPC-X device will be used as input benchmark during the future evaluation.

Conclusion

- Achievements:
- Original work on modelling the resource allocation in heterogeneous distributed system by focusing on the RArs relations through order theory.
- Original apply Birkhoff's representation theory to model the arbitrary allocation topology to a modular structure and design ranking on it.
- Limitations:
- I. need more future evaluations;
- II. Only suited for static case at this step.

Thanks

Open Q&A

References

[1] Microsoft Assessment and Planning Toolkit (MAP). *http://www.microsoft.com/map/*.

[2] Hitesh Ballani, Paolo Costa, Thomas Karagiannis, and Ant Rowstron. Towards predictable datacenter networks. In *ACM SIGCOMM Computer Communication Review*, volume 41, pages 242–253. ACM, 2011.

[3] Nikhil Bansal, Alberto Caprara, and Maxim Sviridenko. Improved approximation algorithms for multidimensional bin packing problems. In *Foundations of Computer Science, 2006. FOCS'06. 47th Annual IEEE Symposium on*, pages 697–708. IEEE, 2006.

[4] Daniel Bleichenbacher and Jiirg Schmid. Computing the canonical representation of a finite lattice. In *Semantics of programming languages and model theory*, volume 5, page 269. CRC Press, 1993.

[5] Alberto Caprara and Paolo Toth. Lower bounds and algorithms for the 2-dimensional vector packing problem. *Discrete Applied Mathematics*, 111(3):231–262, 2001.

[6] FP7 HARNESS consortium. The harness platform: A hardware- and network-enhanced software system for cloud computing. Technical report.

[7] Brian A Davey and Hilary A Priestley. *Introduction to lattices and order*. Cambridge university press, 2002.

[8] George Grätzer. *Lattice theory: foundation*. Springer Science & Business Media, 2011.

[9] Chuanxiong Guo, Guohan Lu, Helen J Wang, Shuang Yang, Chao Kong, Peng Sun, Wenfei Wu, and Yongguang Zhang. Secondnet: a data center network virtualization architecture with bandwidth guarantees. In *Proceedings of the 6th International COnference*, page 15. ACM, 2010.

[10] William Leinberger, George Karypis, and Vipin Kumar. Multi-capacity bin packing algorithms with applications to job scheduling under multiple constraints. In *Parallel Processing, 1999. Proceedings. 1999 International Conference on*, pages 404–412. IEEE, 1999.

[11] Siva Theja Maguluri, R Srikant, and Lei Ying. Heavy traffic optimal resource allocation algorithms for cloud computing clusters. *Performance Evaluation*, 81:20–39, 2014.

[12] Heikki Mannila and Christopher Meek. Global partial orders from sequential data. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 161–168. ACM, 2000.

[13] Hien Nguyen Van, Frederic Dang Tran, and Jean-Marc Menaud. Autonomic virtual resource management for service hosting platforms. In *Proceedings of the 2009 ICSE Workshop on Software Engineering Challenges of Cloud Computing*, pages 1–8. IEEE Computer Society, 2009.

[14] GAlib: A C++ Library of Genetic Algorithm Components. *http://lancet.mit.edu/ga*, 2010.

[15] Rina Panigrahy, Kunal Talwar, Lincoln Uyeda, and Udi Wieder. Heuristics for vector bin packing. *research. microsoft. com*, 2011.

[16] Charles Prud'homme, Jean-Guillaume Fages, and Xavier Lorca. *Choco3 Documentation*. TASC, INRIA Rennes, LINA CNRS UMR 6241, COSLING S.A.S., 2014.

[17] Anshul Rai, Ranjita Bhagwan, and Saikat Guha. Generalized resource allocation for the cloud. In *Proceedings of the Third ACM Symposium on Cloud Computing*, page 15. ACM, 2012.

[18] Paul Shaw. A constraint for bin packing. In *Principles and Practice of Constraint Programming–CP 2004*, pages 648–662. Springer, 2004.

[19] Mark Stillwell, David Schanzenbach, Frédéric Vivien, and Henri Casanova. Resource allocation algorithms for virtualized service hosting platforms. *Journal of Parallel and Distributed Computing*, 70(9):962–974, 2010.

[20] Mark Stillwell, Frédéric Vivien, and Henri Casanova. Dynamic fractional resource scheduling versus batch scheduling. *IEEE Transactions on Parallel and Distributed Systems*, 23(3):521–529, 2012.

[21] Mark Stillwell, Frederic Vivien, and Henri Casanova. Virtual machine resource allocation for service hosting on heterogeneous distributed platforms. In *Parallel & Distributed Processing Symposium (IPDPS), 2012 IEEE 26th International*, pages 786–797. IEEE, 2012.

[22] Maxeler Technologies. *MPC-X, https://www.maxeler.com/products/mpc-xseries*. Maxeler Tech., 2015.

[23] Bhuvan Urgaonkar, Prashant Shenoy, Abhishek Chandra, Pawan Goyal, and Timothy Wood. Agile dynamic provisioning of multi-tier internet applications. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, 3(1):1, 2008.

[24] Shahin Vakilinia, Mustafa Mehmet Ali, and Dongyu Qiu. Modeling of the resource allocation in cloud computing centers. *Computer Networks*, 91:453–470, 2015.

[25] Hien Nguyen Van, Frederic Dang Tran, and Jean-Marc Menaud. Sla-aware virtual resource management for cloud infrastructures. In *Computer and Information Technology, 2009. CIT'09. Ninth IEEE International Conference on*, volume 1, pages 357–362. IEEE, 2009.