PAM & PAL

Policy-Aware Virtual Machine Migration and Placement in Dynamic Cloud Data Centers

Hugo Flores Vincent Tran Bin Tang

Department of Computer Science
California State University Dominguez Hills,
Carson, CA 90747, USA
This work was supported by NSF Grant CNS-1911191.

IEEE INFOCOM, July 2020

Outline

Goals of Project

- Focused on policy-aware data centers **(PADCs)**, wherein virtual machine**(VM)** traffic traverses in
  1. **Ordered:** Sequence of middle boxes.
  2. **Unordered:** Unsequence of middle boxes.
- Proposed two new VM placement and migration problems.
  1. **PAL:** Policy-aware virtual machine placement.
  2. **PAM:** Policy-aware virtual machine migration.
- Satisfied the constraint: resources capacity of PADCs.
- Addressed dynamic communicating traffic in cloud data center such as cloud chatbots[1]

[1] Slack, the collaboration software that moves work forward. http://slack.com., and Amazon lex. https://aws.amazon.com/lex/.

Policy Aware Data Centers (PADCs)

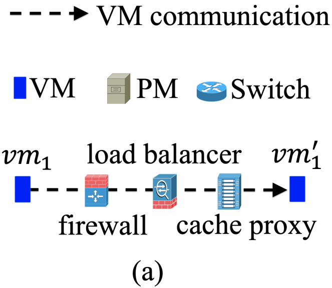


Figure 1: A data center policy with 1 virtual machine pair example

Policy Aware Data Centers (PADCs)

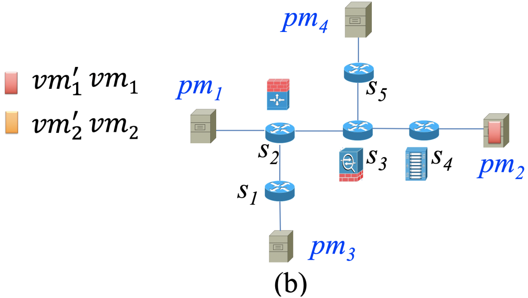


Figure 2: A PADC example with 2 virtual machine pairs

Policy Aware Data Centers (PADCs)

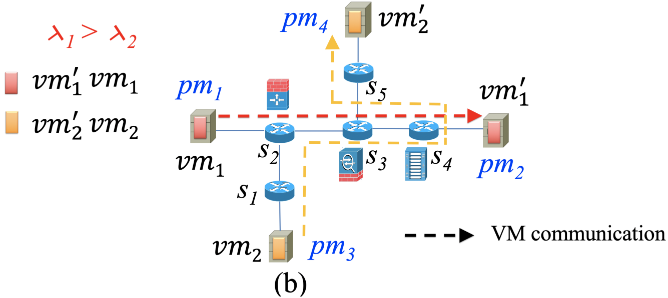


Figure 2: A PADC example with 2 virtual machine pairs

Policy Aware Data Centers (PADCs)

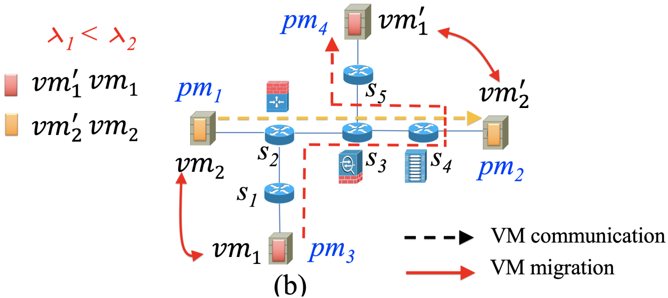


Figure 2: A PADC example with 2 virtual machine pairs

Fat Tree

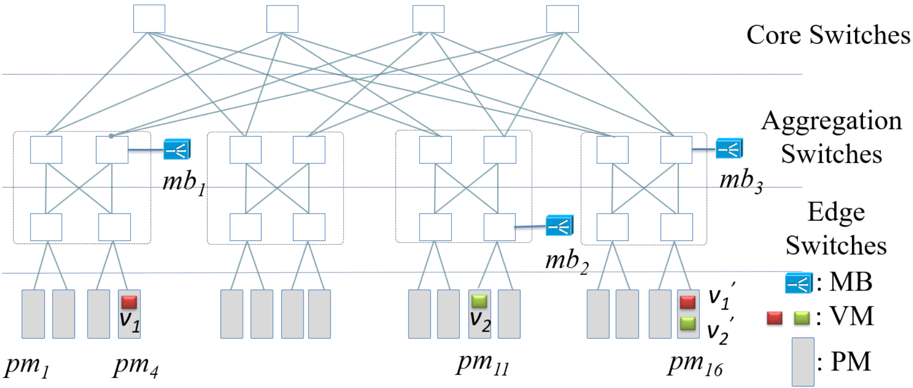


Figure 3: A PADC (k=4) with 16 PMs, 3 MBs, and two VM pairs

Ordered Policy: $vm_1 \rightarrow vm_1'$

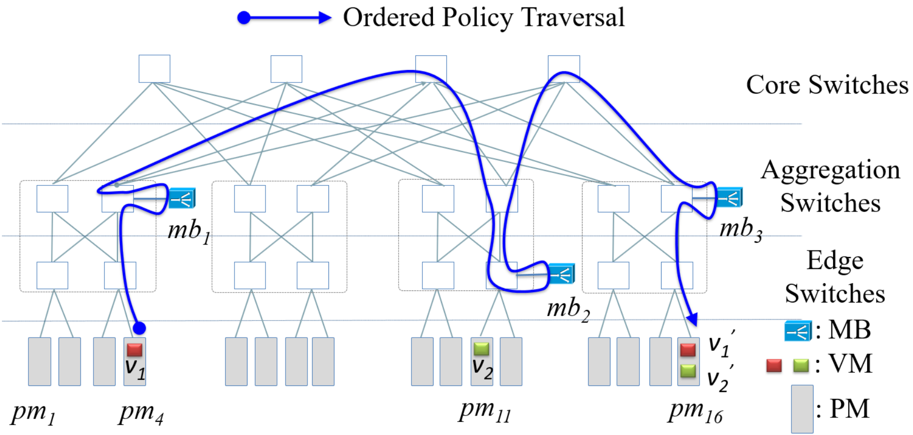


Figure 4: A PADC (k=4) with 16 PMs, 3 MBs, and two VM pairs

Unordered Policy: $vm_2 \rightarrow vm_2'$

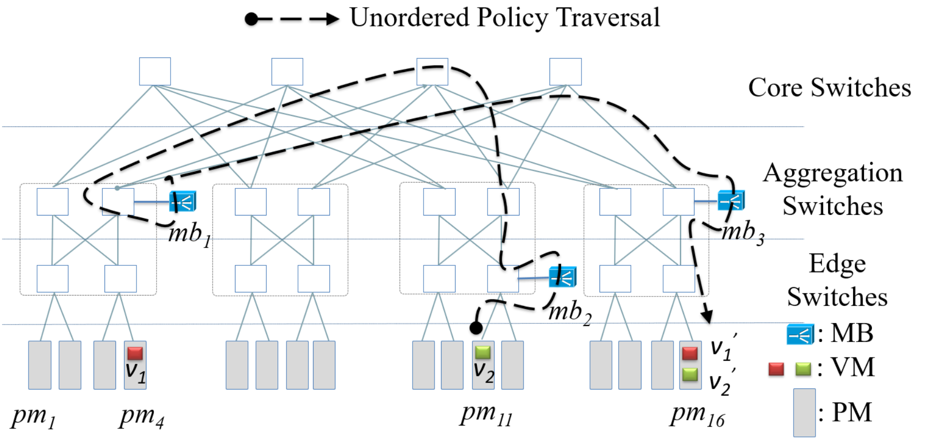


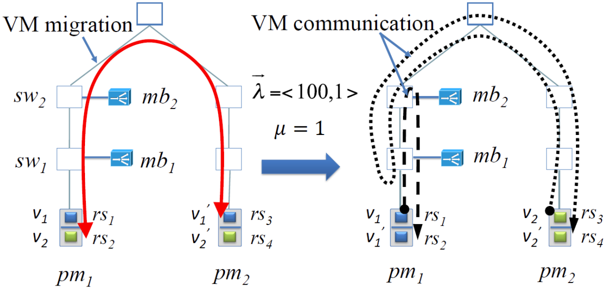
Figure 5: A PADC (k=4) with 16 PMs, 3 MBs, and two VM pairs

Network Model

- An undirected general graph $G(V, E)$.
  - $V = V_p \cup V_s$
    - $V_p = \{pm_1, pm_2, \cdots, pm_{|V_p|}\}$.
    - V_s is a set of switches.
  - E is a set of edges, each connecting either one switch to another switch or a switch to a physical machine.
- A set of n middle boxes, $M = \{mb_1, mb_2, \cdots, mb_n\}$ with mb_j installed at switch $sw_j \in V_s$.
- ℓ pairs of communicating VMs $P = \{(v_1, v_1'), (v_2, v_2'), \cdots, (v_\ell, v_\ell'), \}$ in which v_i is source and v_i' is destination.
- $\overrightarrow{\lambda} =< \lambda_1, \lambda_2, \cdots, \lambda_\ell >$.
  - λ_i is the traffic rate or transmission rate between (v_i, v_i').
  - Change over time.

A $k = 2$ data Center with $n = 2, \ell = 2$

Figure 5: Computation the total cost in ordered policy



(a) Before migration (b) After migration

Before Migration	After Migration
$C_t(m) = \underbrace{6 \cdot 100}_{(v_1, v_1')} + \underbrace{6 \cdot 1}_{(v_2, v_2')} = 606$	$C_t(m) = \underbrace{4 \cdot 100}_{(v_1, v_1')} + \underbrace{10 \cdot 1}_{(v_2, v_2')} + \underbrace{(6+6) \cdot 1}_{\text{migration cost}} = 422$

Problem formulation of PAL for Ordered Policy

The total communication cost of all the l VM pairs under a VM placement function p as $C_c(p)$

$$C_c(p) = \underbrace{\sum_{i=1}^{l} \lambda_i \cdot \sum_{j=1}^{n-1} c(sw_j, sw_{j+1})}_{MBs} + \underbrace{\sum_{i=1}^{l} \lambda_i \cdot (c(p(v_i), sw_1) + c(sw_n, p(v_i')))}_{c_{in}(rs), c_e(rs)}$$

The objective of PAL is to find a VM placement p to minimize $C_c(p)$ while satisfying resource constraint of PMs: $|\{v \in \mathbb{V} | p(v) = i\}| \leq rc_i, \forall i \in \mathbb{V}_p$.

PAL Algorithm for Ordered Policy

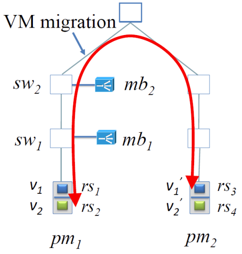
Algorithm 1

- **Input:** A PADC with ordered policy $(mb_1, mb_2, ..., mb_n)$, VM pairs $\mathbb{P}$, $V_p = \{pm_i\}$, resource capcacity rc_i.
- **Output:** A placement p and the total communication cost $C_c(p)$
- **Algorithm(summary)**
  1. Sort resource slots in ascending order of their ingress (and egress) costs
  2. Find optimal resource slots for (v_k, v_k')
  3. Descending frequencies communication of virtual machine pair
  4. Place VM pairs and calculate cost
- **Return:** A placement p and total communication cost.

Theorem

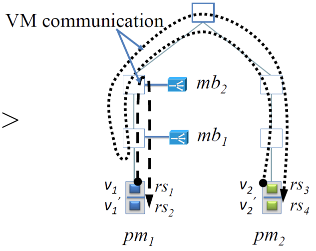
PAL Algorithm for Ordered Policy finds the VM placement that minimizes total communication cost for the ℓ VM pairs.

PAL Algorithm for Ordered Policy



(a) Before migration

$$\vec{\lambda} = < 100, 1 >$$



(b) After migration

Initial

$v_1' : rs_2 \rightarrow rs_1,\ v_2 : rs_1 \rightarrow rs_2$

$\mathcal{I}$	rs_1	rs_2	rs_3	rs_4
$\mathcal{E}$	rs_1	rs_2	rs_3	rs_4

PAL Ordered Alg.

$\mathcal{I}^{opt}$	rs_1	rs_3
$\mathcal{E}^{opt}$	rs_2	rs_4

Total Communication cost: $100 \cdot 4 + 1 \cdot 10 = 410$

Problem Formulation of PAL for Unordered Policy

Total Communication Cost: $C_c(p, \vec{\pi}) = \sum_{i=1}^{\ell} c_i^{p,\pi^i}$, where

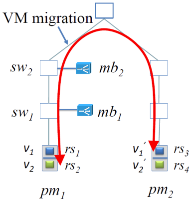
1. Given p and π^i, denote (v_i, v_i')'s communication cost as

$$c_i^{p,\pi^i} = \lambda_i \cdot c\left(p(v_i, sw(\pi^i(1))\right) + \lambda_i \cdot \sum_{j=1}^{n-1} c\left(sw(\pi^i(j)), sw(\pi^i(j+1))\right) + \lambda_i \cdot \left(sw(\pi^i(n)), p(v_i')\right)$$

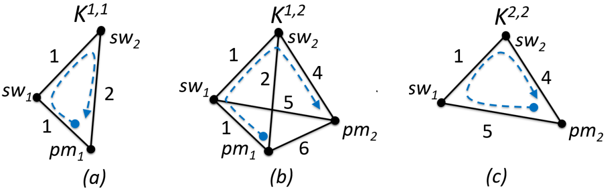
2. MB traversal function π^i: a permutation function indicating the j^{th} MB that (v_i, v_i') visits is $mb_{\pi^i j}$

PAL Algorithm for Unordered Policy

The algorithm achieves 2-approximation when $\ell = 1$



(a) Before migration



$X = (1, 1, 4)$ $X = (1, 2, 6)$ $X = (2, 2, 10)$

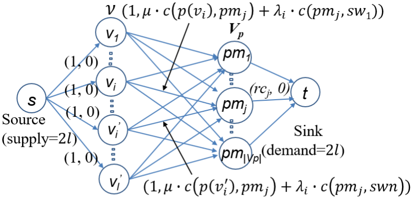
- (v_1, v_1') is placed at pm_1 in (a), and (v_2, v_2') at pm_2 in (c).
- Total cost: 410

Problem formulation of PAM for Ordered Policy

Let $C_t(m)$ be the total migration and communication cost of all pairs after m. Then total cost is calculated as follow:

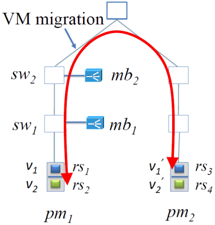
$$
\begin{aligned}
C_t(m) = \quad & \overbrace{\sum_{i=1}^{\ell} \lambda_i \cdot \sum_{j=1}^{n-1} c(sw_j, sw_{j+1})}^{\textit{communication cost between middle boxes}} \\
& + \underbrace{\sum_{i=1}^{\ell} \left(\mu \cdot c(p(v_i), m(v_i)) + \lambda_i \cdot c(m(v_i), sw_1)\right)}_{\textit{migration and communication } v_i \rightarrow \textit{first mbs}} \\
& + \underbrace{\sum_{i=1}^{\ell} \left(\mu \cdot c(p(v_i'), m(v_i')) + \lambda_i \cdot c(sw_n, m(v_i'))\right)}_{\textit{migration and communication } v_i' \rightarrow \textit{last mbs}}
\end{aligned}
$$

Transforming PADC to FLow Network for k = 2

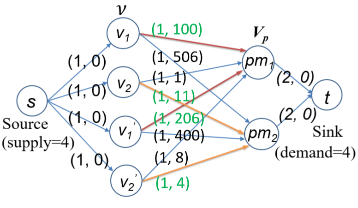


(a)

$$C_t = 100 + 11 + 206 + 4 + \underbrace{101}_{sw_1 \to sw_2} = 422$$



(a) Before migration



(b)

PAM is equivalent to minimum cost flow problem.

Graph transformation and MCF results for PADC

Problem formulation of PAM for Unordered Policy

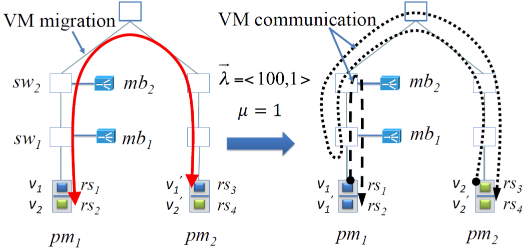
Let $\vec{\pi} =< \pi^1, \pi^2, ..., \pi^\ell >$ and $C_t(m, \vec{\pi})$ denote the total cost of all the VM pairs with m and $\vec{\pi}$. Then,

$$C_t(m, \vec{\pi}) = \overbrace{\sum_{i=1}^{\ell} (\mu \cdot c(p(v_i), m(v_i)) + \mu \cdot c(p(v_i'), m(v_i')))}^{\textit{total migration cost}}$$

$$\underbrace{+ \sum_{i=1}^{\ell} \lambda_i \cdot \left(\sum_{j=1}^{n-1} c\left(sw(\pi^i(j)), sw(\pi^i(j+1))\right) + c\left(m(v_i, sw(\pi^i(1))\right) + c\left(sw(\pi^i(n)), m(v')\right) \right)}_{\textit{total communication cost}}$$

PAM Algorithm for Unordered Policy

1. Compute the path route between two physical machine.
2. Find PM pair for VM pair, capacity of each physical machine is checked.
3. Update migration scheme and total cost.



(a) Before migration (b) After migration

Migrate $v_1' \rightarrow pm_1 : 406,\ v_2 \rightarrow pm_2 : 16,\ C_t(m) = 422$

Simulation Results

- Fat-tree PADCs, $k = 8$ with 128 PMs and $k = 16$ with 1024 PMs.
- The Traffic rates of VM pairs are in range of $[0, 1000]$.
- Traffic rates[2]: 25% in $[0, 300)$, 70% in $[300, 700]$, and 5% in $(700, 1000]$.
- VM Pairs[3]: 80% are placed into the PMs under the same edge switches, 20% under different edge switches.
- Each data point in an average of 20 runs with 95% confidence interval.
- In each run a new set of VM pairs are placed or migrated in the PADC.

Table 1: COMPARING PAM AND PAL ALGORITHMS

	Ordered Policy	Unordered Policy	Existing Work
PAL	Optimal (Algo. 1)	Approx-PAL (Algo. 3)	TrafficAware
PAM	MCF	Approx-PAM (Algo. 4)	PLAN

[2] Facebook Data Centers(Section 5.1, A. Roy, H. Zeng, J. Bagga, G. Porter, and A. C. Snoeren. Inside the social network's (datacenter) network. In Proc. of ACM SIGCOMM 2015.)

[3] Cisco Design Guide, Cisco virtualized multi-tenant data center, version 2.0 compact pod design guide. http://hyperurl.co/hpj2xt.

PLAN & Traffic Aware

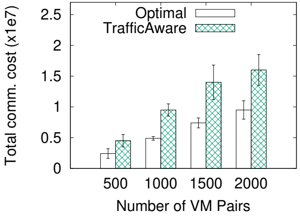
1. **PLAN[4] - Greedy algorithm**
   - Works in rounds.
   - Each round, computes that which VM is migrated to which PM with available resources, such that the utility of this migration is the maximum among all the VMs that have not been migrated.
   - Continues until all the VMs are migrated, or no more VM migration gives any positive utility.
2. **Traffic Aware[5]**
   - Only assigns VMs to the same PMs or PMs in close proximity, and does not consider the proximity of the PMs to the MBs.
   - In ordered policy, it places VM pairs (in non-ascending order of their traffic rates) to the PMs that are closest to the ingress switch.
   - In unordered policy, TrafficAware always places VM pairs in the same PM if possible.

---

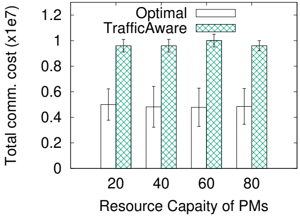
[4] L. Cui, F. P. Tso, D. P. Pezaros, W. Jia, and W. Zhao. Plan: Joint policy- and network-aware vm management for cloud data centers. IEEE Transactions on Parallel and Distributed Systems, 28(4):1163–1175, 2017

[5] X. Meng, V. Pappas, and L. Zhang. Improving the scalability of data center networks with traffic-aware virtual machine placement. In Proc. of IEEE INFOCOM 2010.

Comparing with TrafficAware



(a) Varying ℓ, $n = 3$, $rc = 40$

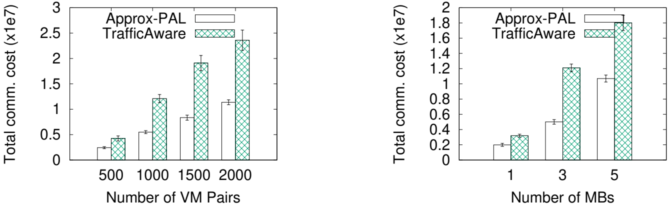


(b) Varying rc, $\ell = 1000$, $n = 3$

Figure 6: Comparing with TrafficAware in ordered policy, $k = 16$

- In (a), varies the number of VM pairs ℓ and shows that Optimal yields 46-49% less costs than TrafficAware.
- In (b), varies resource capacities of PMs rc and shows that Optimal outperforms TrafficAware by around 48%.

Comparing with TrafficAware (Cont.)

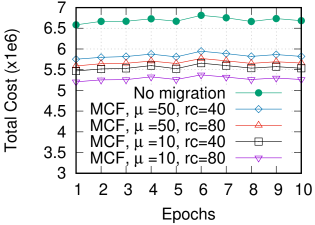


(a) Varying ℓ, $n = 3$, $rc = 40$ (b) Varying n, $\ell = 1000$, $rc = 40$

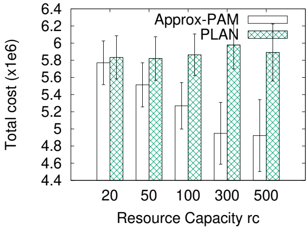
Figure 7: Comparing with TrafficAware in unordered policy, $k = 16$

- Varies ℓ as well as number of MBs n and shows that Approx-PAL outperforms TrafficAware by 37-58% in all scenarios.

Effect of VM Migrations



(a) Effects of VM migration

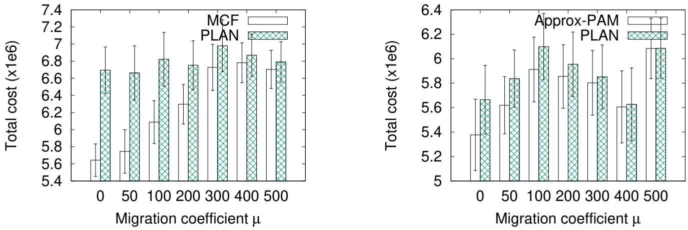


(b) Comparing with PLAN

Figure 8: $k = 8, \ell = 1000, m = 3$

- VM migration reduces the total costs by up to 25% ($\mu = 10$) and $rc = 80$ compared to without migration.

Comparing with PLAN



(a) Ordered Policy (b) Unordered Policy

Figure 9: Comparing with PLAN, $k = 8, \ell = 1000, n = 3, rc = 40$

- In (a), ordered policy, the MCF outperforms are the PLAN by around 20% when μ small. With the increase of μ, PLAN and MCF start to perform close due to high migration cost.
- In (b), unordered policy, Approx-PAM outperforms PLAN slightly for the entire range of μ.

Comparing with PLAN (cont.)

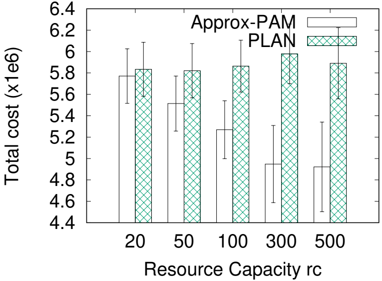


Figure 10: Comparing with PLAN, $k = 8, \ell = 1000, n = 3, rc = 40$

- Compares Approx-PAM and PLAN by varying *rc* while fixing μ as 20, and shows that Approx-PAM outperforms PLAN for the entire range of *rc*.
- With the increase of the *rc*, the performance difference between Approx-PAM and Greedy gets larger, around 30%.

Conclusion

- VM migration in an effective technique to alleviate dynamic VM traffic in PADCs.
- Working together, PAL and PAM place and then migrate VMs in the event of dynamic traffic fluctuation, achieving optimal and near-optimal network resource management for a PADC's lifetime.
- Uncovered a suite of new policy-aware problems and designed optimal, approximation, and heuristic algorithms.

Future Work

- Study if the optimality and approximability of our algorithms still hold when VMs have different resource demands.
- Study how VNF migration mitigates diverse and dynamic traffic and design a holistic VNF+VM migration scheme to achieve ultimate resource optimization in PADCs.

Thank you

github: https://github.com/vtran42/pal-pam-datacenter