



Lagopus switchの評価

2014年10月14日

Lagopus switch* : 特徴

- DPDKを組み込んだ仮想スイッチ
- 受信、送信、パケット処理を行うコアを予め指定
- コア数により性能が変化

* : <http://lagopus.github.io/>

Lagopus : インストール手順

- lagopus/QUICKSTART.mdを参照
 - 下記のOSでインストール

OS	Ubuntu14.04 (minimal)
Network	1GBASE-T (4ports)

Lagopus : 起動

- lagopusスイッチ

```
test@dppk% sudo lagopus -d -l hoge hoge -- -c1f -n1 -- -p3 [ /etc/lagopus ]
```

- CPU core : (0,1,2,3,4)
- Memory channel : 1
- Port : (0,1)

lagopus.confが存在する
ディレクトリで実行

- Ryu コントローラ*

- 検証に応じてあらかじめフローを追加する
- packet in 処理については評価しない

* : <http://osrg.github.io/ryu/>

Lagopus : 機能

- OpenFlow機能 (確認済)

Match
in_port
dl_src/dst
ethertype
nw_src/dst
vlan_vid

Actions
output
push_vlan, pop_vlan
push_mpls
set_field(vlan_vid, mpls_label, mpls_ttl)
group (all, indirection, watch_port)
meter (rate, burst)
decrement_nw_ttl
set_field (eth_src/dst)

- 未確認

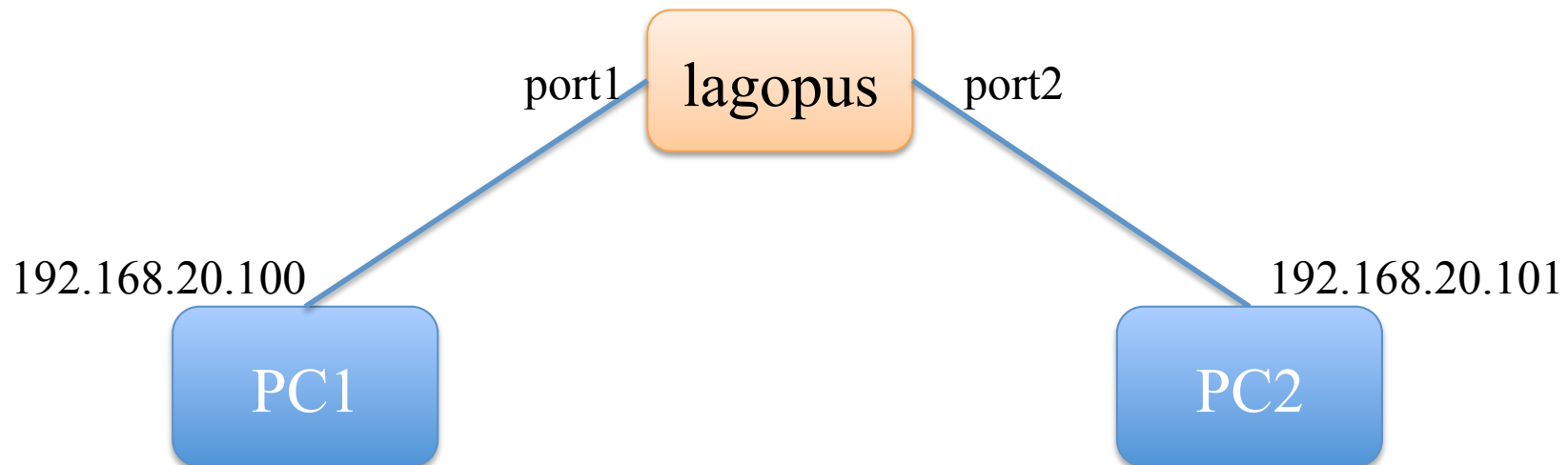
Match
tcp/udp_src/dst

Actions
pop_mpls

Lagopus : 性能評価(ping)

- 評価環境

OS	Ubuntu14.04
CPU	Intel(R) Atom(TM) 2.40GHz 8core
Memory	8Gbytes
Network	1GBASE-T (4ports)
Chipset	FGBGA-1283 (Intel)



Lagopus : 性能評価(ping)

- PC1からPC2へpingを送信

- memory channel : 1
- port mask : 0x3

match	action
in_port=1	output:2
in_port=2	output:1

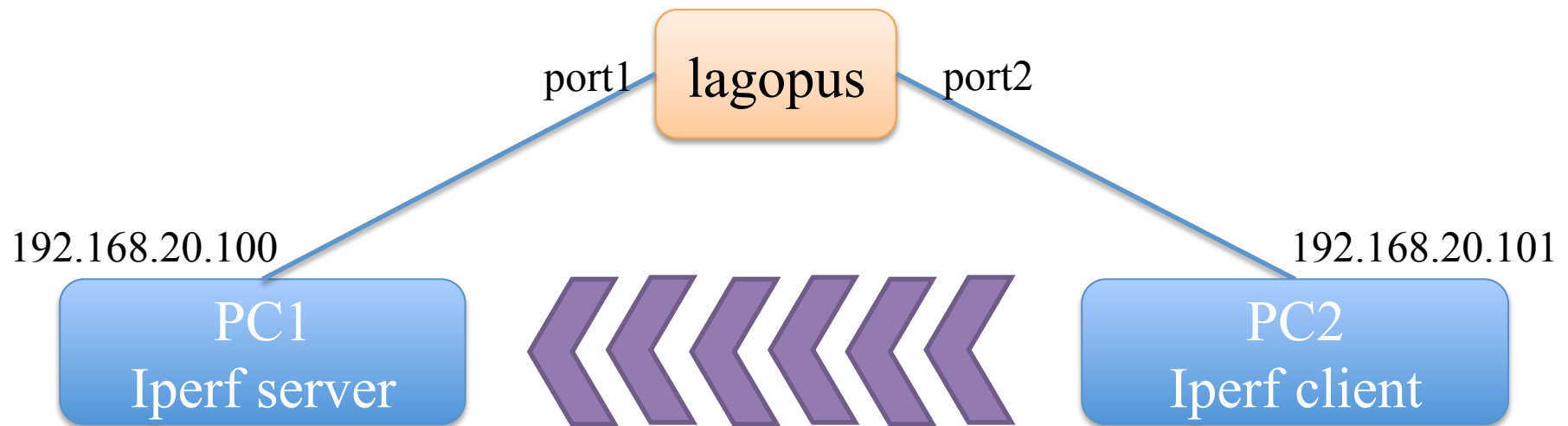
core mask (number of core)	latency
0x3 (1)	20ms
0x7 (2)	11ms
0xf (3)	13~15ms
0x1f(4)	5ms~7ms
0x3f(5)	6ms~8ms(~11ms)
0x7f(6)	7ms~9ms
0xff (7)	6ms~8ms

- IP forwarding, OpenvSwitch: 0.280ms

Lagopus : 性能評価(Iperf)

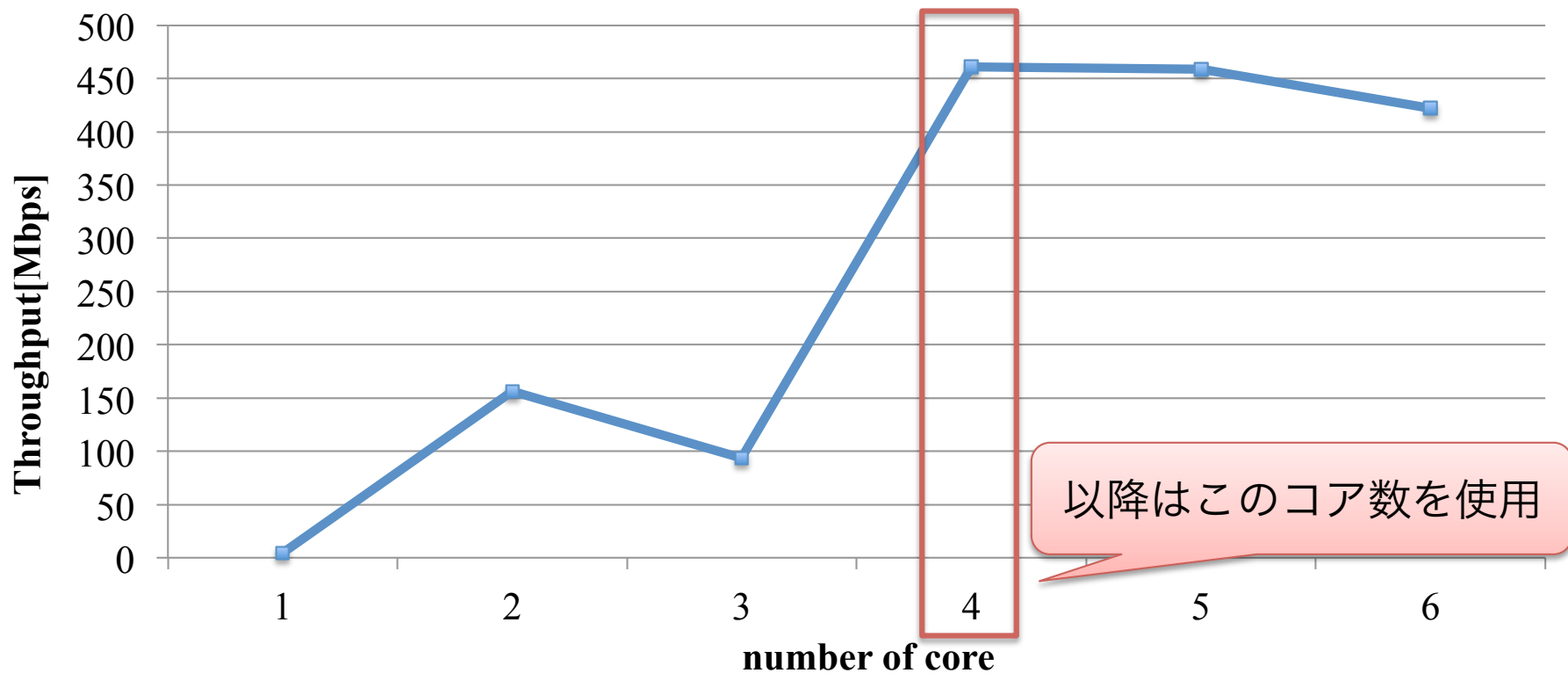
- 評価環境

OS	Ubuntu14.04
CPU	Intel(R) Atom(TM) 2.40GHz 8core
Memory	8Gbytes
Network	1GBASE-T (4ports)
Chipset	FGBGA-1283 (Intel)



Lagopus : 性能評価(Iperf)

- 評価対象 : 計測時間60sの平均スループット
- port mask : 0x03(0,1)
- フローエントリ : pingと同様



Lagopus : 性能評価 (パケット処理速度)

- 100bytes (パケットヘッダを含む)

```
[ 3] local 192.168.20.100 port 39549 connected with 192.168.20.101 port 5001
[ ID] Interval      Transfer      Bandwidth
[ 3]  0.0-10.0 sec  39.1 MBytes  32.8 Mbits/sec
[ 3] 10.0-20.0 sec  39.1 MBytes  32.8 Mbits/sec
[ 3] 20.0-30.0 sec  39.1 MBytes  32.8 Mbits/sec
[ 3] 30.0-40.0 sec  39.2 MBytes  32.8 Mbits/sec
[ 3] 40.0-50.0 sec  39.1 MBytes  32.8 Mbits/sec
[ 3]  0.0-60.0 sec  235 MBytes  32.8 Mbits/sec
```

68.33Kpps

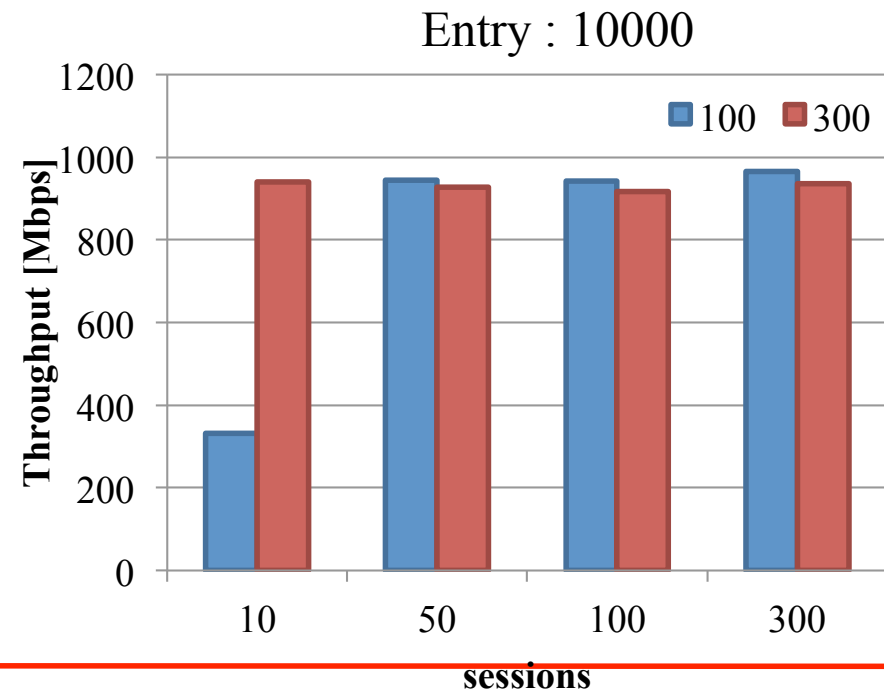
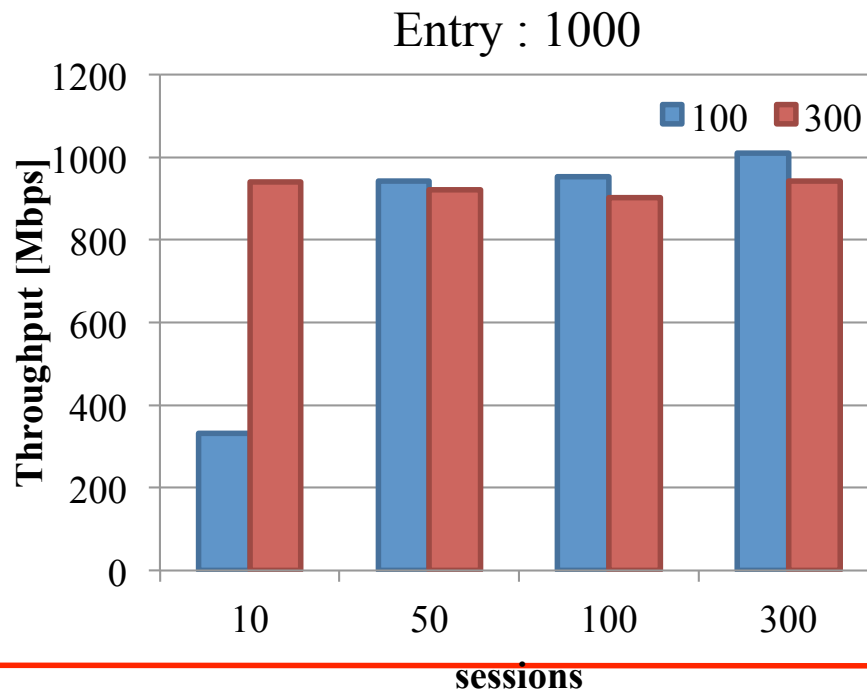
- 300bytes (パケットヘッダを含む)

```
[ 3] local 192.168.20.100 port 39548 connected with 192.168.20.101 port 5001
[ ID] Interval      Transfer      Bandwidth
[ 3]  0.0-10.0 sec  170 MBytes  142 Mbits/sec
[ 3] 10.0-20.0 sec  170 MBytes  142 Mbits/sec
[ 3] 20.0-30.0 sec  170 MBytes  142 Mbits/sec
[ 3] 30.0-40.0 sec  170 MBytes  142 Mbits/sec
[ 3] 40.0-50.0 sec  170 MBytes  142 Mbits/sec
[ 3] 50.0-60.0 sec  170 MBytes  142 Mbits/sec
[ 3]  0.0-60.0 sec 1019 MBytes  142 Mbits/sec
```

68.26Kpps

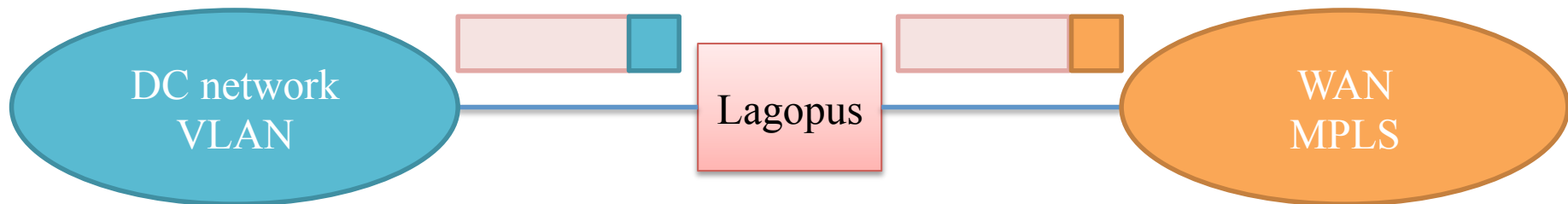
Iperf : 性能評価(Entry/Session)

- エントリ数 : 1000,10000
 - match : (dl_dst/src, nw_dst/src/ dl_type)
 - actions : (output)
- パケットサイズ : 100, 300bytes
- 評価対象 : 80秒間の平均スループット



Iperf : 性能評価(actions)

- Actions処理の概要
 - DC→WAN : VLAN POP, MPLS PUSH

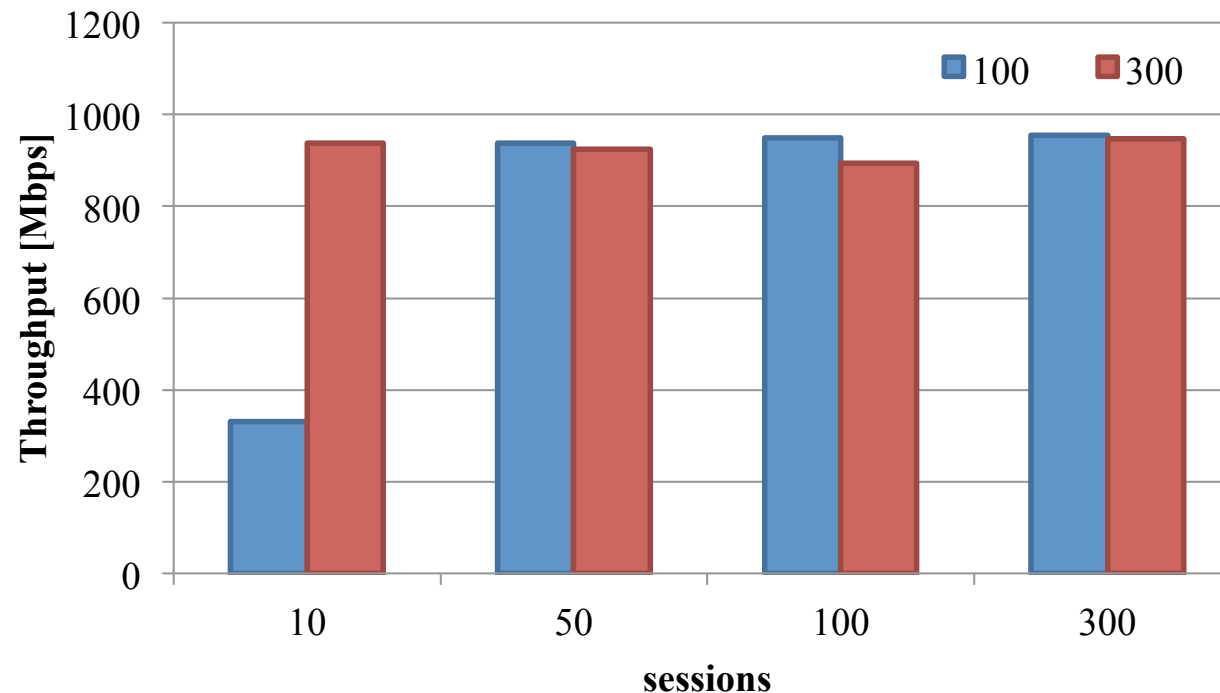


- Flow Table
 - PUSH MPLSをVLAN PUSHで代用

```
dpdk> show flow
Bridge: br0
Table id: 0
  priority=1000,idle_timeout=0,hard_timeout=0,flags=0,cookie=0,packet_count=57691532,byte_count=87477030548,in_port=1 actions=pop_vlan,push_vlan:0x8100,set_field:4116->vlan_vid,set_field:aa:aa:aa:aa:aa:aa->eth_src,set_field:74:d4:35:6b:d7:ea->eth_dst,dec_nw_ttl,output:2
  priority=1000,idle_timeout=0,hard_timeout=0,flags=0,cookie=0,packet_count=45380385,byte_count=3296356462,in_port=2 actions=output:1
```

Iperf: 性能評価(actions)

- 評価結果
 - 評価内容はentry/actionsと同様



meter

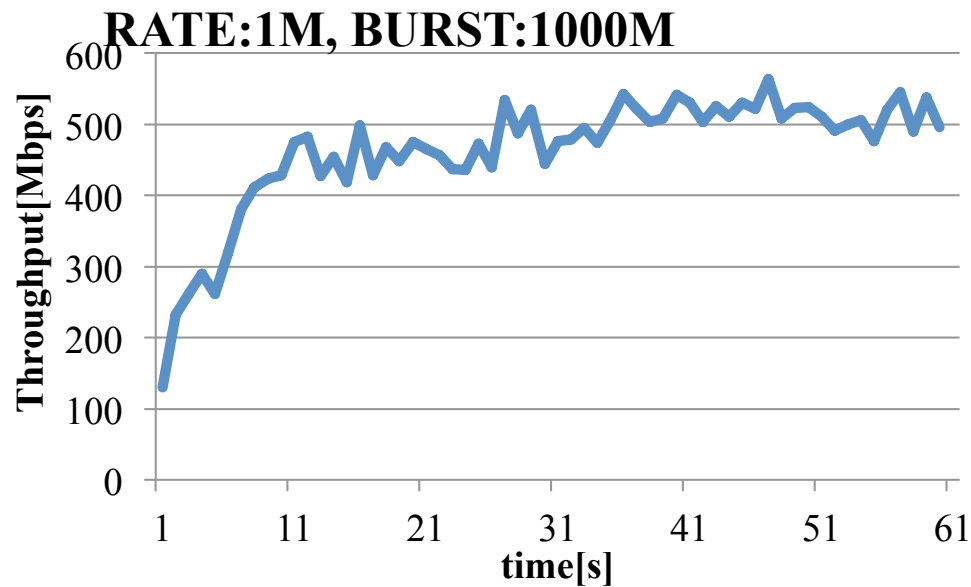
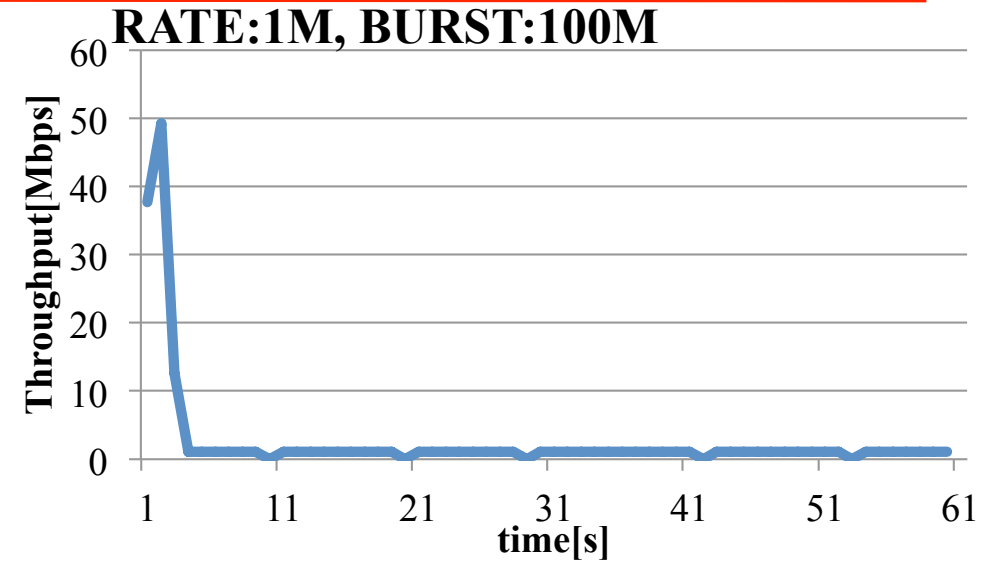
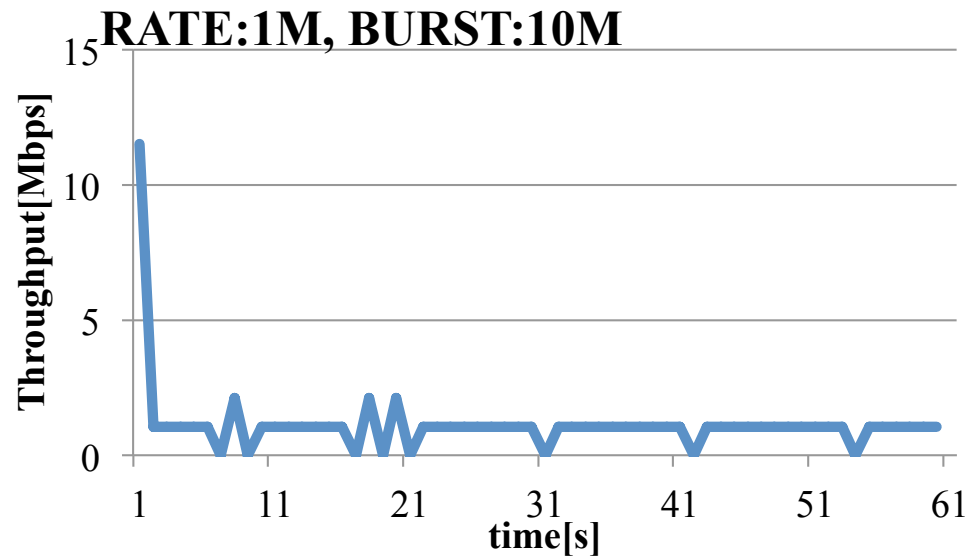
- `./meterGenerator.py -i 1 -R 1000 -B X`

match	action
in_port=1	meter_id=1,output=2
in_port=2	output=1

meter_id	meter
1	rate=1000, burst=X

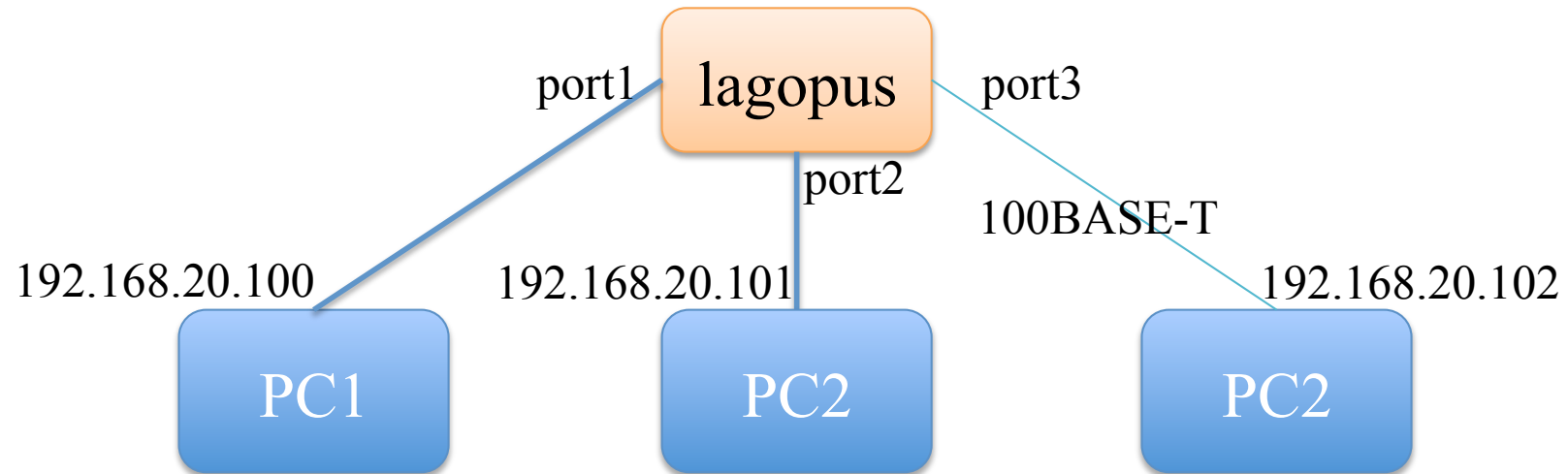
- Iperfを使用

meter



Lagopus : group

- 評価環境



group: all

match	action	group_id	type	watch_port	bucket1	bucket2
in_port=1	output:3	1	all	any	output:1	output:2
in_port=2	output:3					
in_port=3	group:1					

- PC3からPC1,2へpingを送信

```
ray:~ shin$ cat all1.txt
PING 192.168.20.100 (192.168.20.100): 56 data bytes
64 bytes from 192.168.20.100: icmp_seq=0 ttl=64 time=45.620 ms
64 bytes from 192.168.20.100: icmp_seq=1 ttl=64 time=21.388 ms
64 bytes from 192.168.20.100: icmp_seq=2 ttl=64 time=23.643 ms
ray:~ shin$ cat all2.txt
PING 192.168.20.101 (192.168.20.101): 56 data bytes
64 bytes from 192.168.20.101: icmp_seq=0 ttl=64 time=44.571 ms
64 bytes from 192.168.20.101: icmp_seq=1 ttl=64 time=19.274 ms
64 bytes from 192.168.20.101: icmp_seq=2 ttl=64 time=23.139 ms
ray:~ shin$
```

group: indirection

match	action	group_id	type	watch_port	actions
in_port=1	group:1	1	indirection	any	group:10
in_port=2	output:1	10	all	any	output:2

```
PING 192.168.20.101 (192.168.20.101) 56(84) bytes of data.  
64 bytes from 192.168.20.101: icmp_seq=1 ttl=64 time=21.0 ms  
64 bytes from 192.168.20.101: icmp_seq=2 ttl=64 time=21.3 ms  
64 bytes from 192.168.20.101: icmp_seq=3 ttl=64 time=20.2 ms  
64 bytes from 192.168.20.101: icmp_seq=4 ttl=64 time=20.9 ms  
64 bytes from 192.168.20.101: icmp_seq=5 ttl=64 time=21.7 ms  
64 bytes from 192.168.20.101: icmp_seq=6 ttl=64 time=22.3 ms  
64 bytes from 192.168.20.101: icmp_seq=7 ttl=64 time=20.7 ms  
64 bytes from 192.168.20.101: icmp_seq=8 ttl=64 time=21.6 ms  
64 bytes from 192.168.20.101: icmp_seq=9 ttl=64 time=19.4 ms  
64 bytes from 192.168.20.101: icmp_seq=10 ttl=64 time=20.4 ms  
  
--- 192.168.20.101 ping statistics ---  
10 packets transmitted, 10 received, 0% packet loss, time 9014ms  
rtt min/avg/max/mdev = 19.498/21.009/22.362/0.794 ms
```

group: watch_port

match	action	group_id	type	watch_port	actions
in_port=1	group:1	1	all	3	output:2
in_port=2	output:1				

- 10s後にport3を抜去

```
muramatsu@ubuntu:~$ cat watch_port.txt
PING 192.168.20.101 (192.168.20.101) 56(84) bytes of data.
64 bytes from 192.168.20.101: icmp_seq=1 ttl=64 time=21.2 ms
64 bytes from 192.168.20.101: icmp_seq=2 ttl=64 time=21.7 ms
64 bytes from 192.168.20.101: icmp_seq=3 ttl=64 time=20.1 ms
64 bytes from 192.168.20.101: icmp_seq=4 ttl=64 time=17.5 ms
64 bytes from 192.168.20.101: icmp_seq=5 ttl=64 time=22.3 ms
64 bytes from 192.168.20.101: icmp_seq=6 ttl=64 time=19.4 ms
64 bytes from 192.168.20.101: icmp_seq=7 ttl=64 time=17.9 ms
64 bytes from 192.168.20.101: icmp_seq=8 ttl=64 time=23.3 ms
64 bytes from 192.168.20.101: icmp_seq=9 ttl=64 time=21.5 ms
64 bytes from 192.168.20.101: icmp_seq=24 ttl=64 time=43.7 ms
64 bytes from 192.168.20.101: icmp_seq=25 ttl=64 time=21.1 ms
64 bytes from 192.168.20.101: icmp_seq=26 ttl=64 time=21.5 ms
64 bytes from 192.168.20.101: icmp_seq=27 ttl=64 time=20.6 ms
64 bytes from 192.168.20.101: icmp_seq=28 ttl=64 time=23.0 ms
64 bytes from 192.168.20.101: icmp_seq=29 ttl=64 time=22.7 ms
64 bytes from 192.168.20.101: icmp_seq=30 ttl=64 time=19.0 ms

--- 192.168.20.101 ping statistics ---
30 packets transmitted, 16 received, 46% packet loss, time 29135ms
rtt min/avg/max/mdev = 17.547/22.350/43.746/5.770 ms
muramatsu@ubuntu:~$
```

Lagopus : まとめ

- QUICKSTARTに従えばすぐに利用可能
- OpenFlow 1.3の機能は使用可能
 - Meter機能の動作は不安定
- 問題点
 - パケット処理の有無に関わらずCPU負荷が常に100%
 - -c 0x1fの場合(2,3,4,5), 0xffの場合(2,3,4,5,6,7,8)が100%
 - シングルストリームの場合OVSに劣る
 - マルチストリームの場合同等のスループット
- 今後の期待
 - 処理性能の向上
 - パケットイン機能の高速処理