An availability model of exchange point and the JPNAP case study

Katsuyasu Toyama
JPNAP
by Internet Multifeed Co.

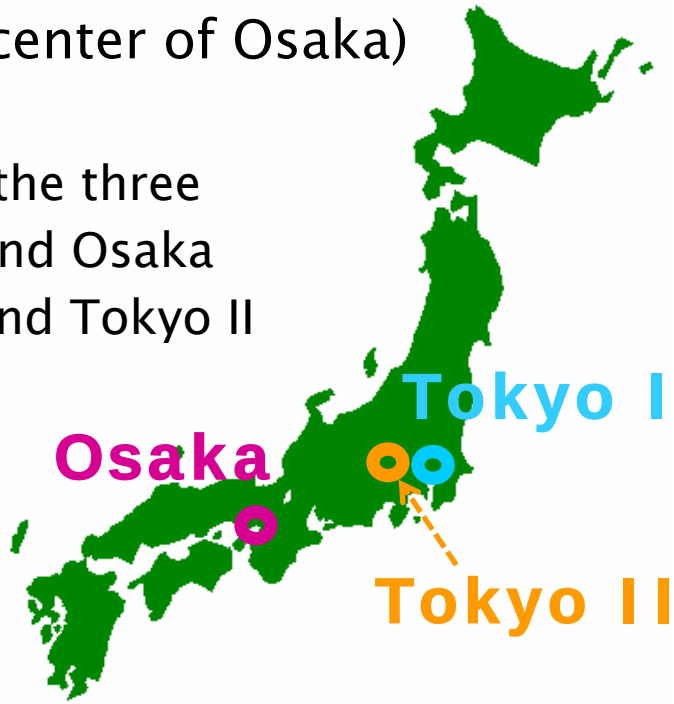
Outline

- Brief introduction of JPNAP
- An availability model of exchange point
- JPNAP solution for improving availability
- JPNAP case study

JPNAP:
the largest exchange point in Japan

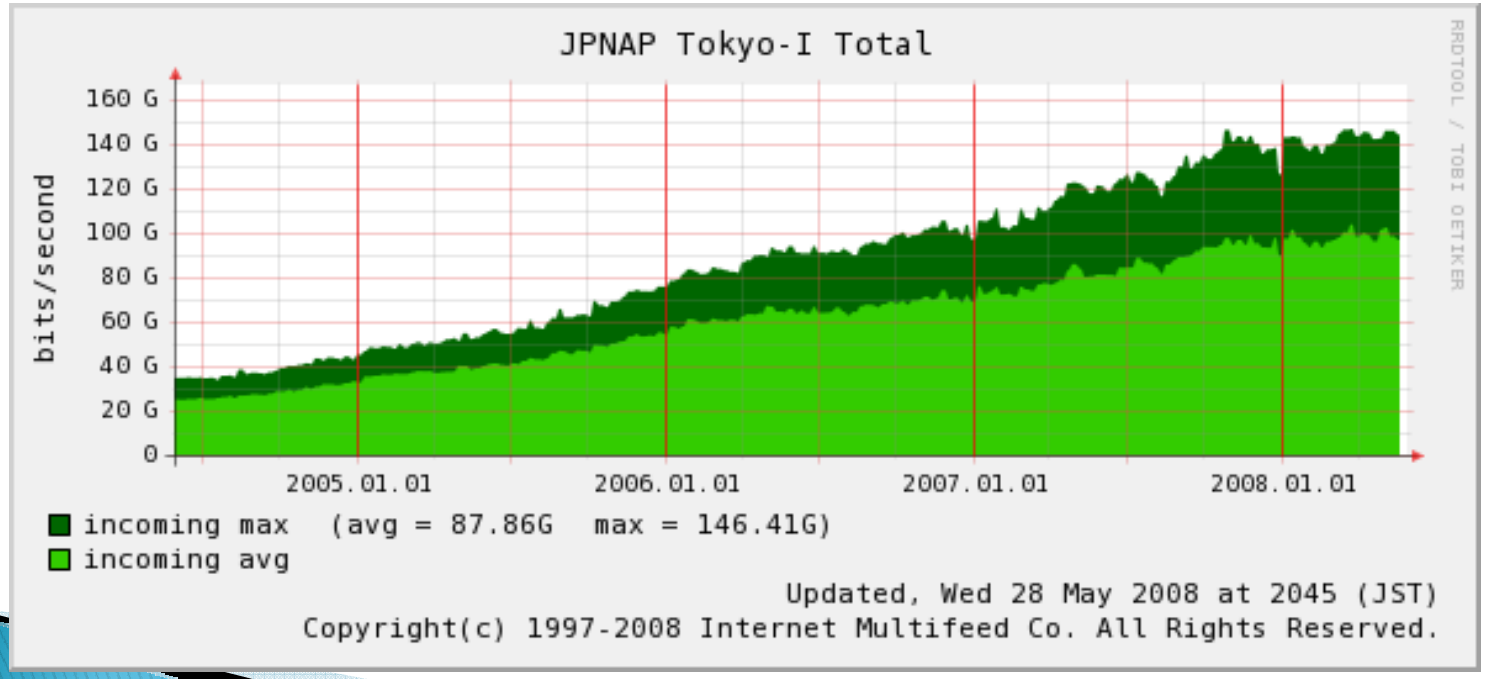
- Provided by INTERNET MULTIFEED CO. since 2001
- Currently three exchange points:
  - JPNAP Tokyo I (Otemachi, the center of Tokyo)
  - JPNAP Tokyo II (Ikebukuro, the hill side of Tokyo)
  - JPNAP Osaka (Dojima, the center of Osaka)

    - No interconnection among the three
    - 400 miles between Tokyo and Osaka
    - 10 miles between Tokyo I and Tokyo II



JPNAP:
the largest exchange point in Japan

- The largest internet traffic are exchanged at JPNAP
  - 146Gbpsat peak, at JPNAP Tokyo 1

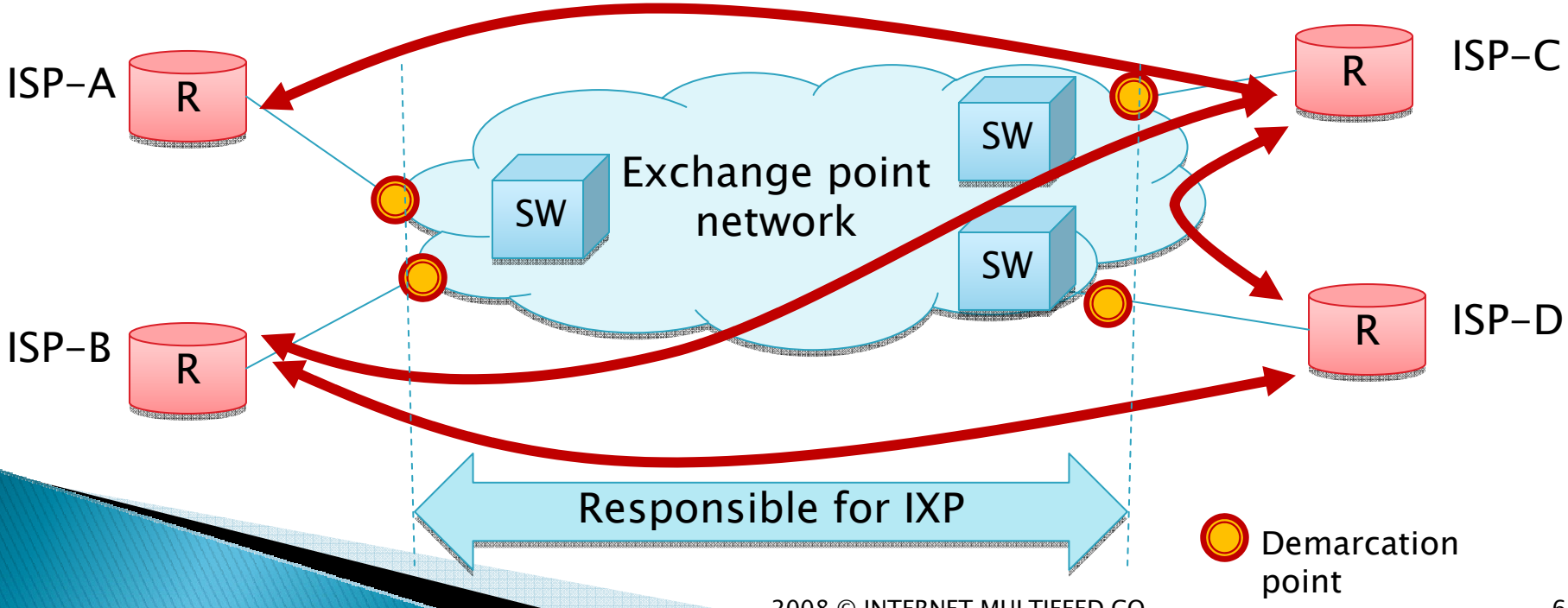


Outline

- Brief introduction of JPNAP
- An availability model of exchange point
- JPNAP solution for improving availability
- JPNAP case study

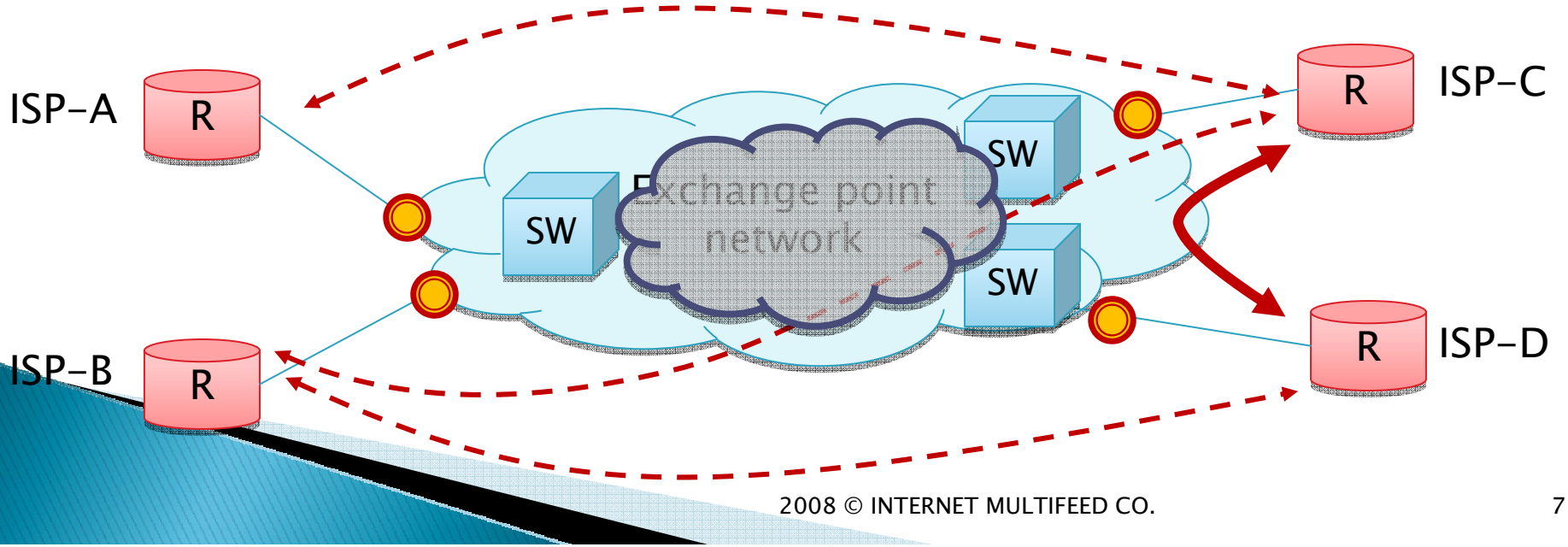
An availability model of exchange point

- Definition: exchange point is "available," if:
  1. No packet loss and no link failure in switch cloud and
  2. All ports for customers are up and running



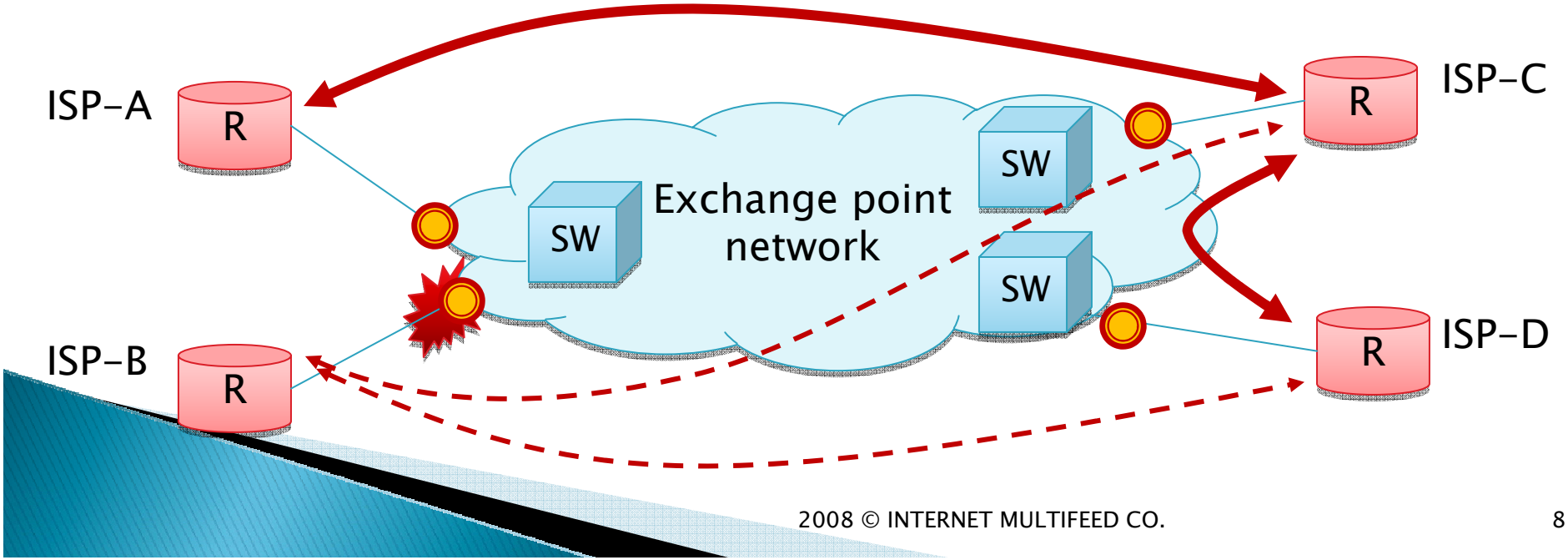
“Not available” cases (1)

- Packet loss or link down in switching cloud
  - Exchange point should be regarded as big pipe for all the customers, with enough capacity/bandwidth.
    - *It is not the case that a customer is going to push or pull more traffic into/from the switching cloud than the contracted bandwidth.*



“Not available” cases (2)

- One port for ISP is down
  - Due to device failure or maintenance of IXP
    - *Other customers who do not peer with ISP-B are not affected. However, an exchange point provider should maintain the environment where all the customers can peer with each other anytime.*

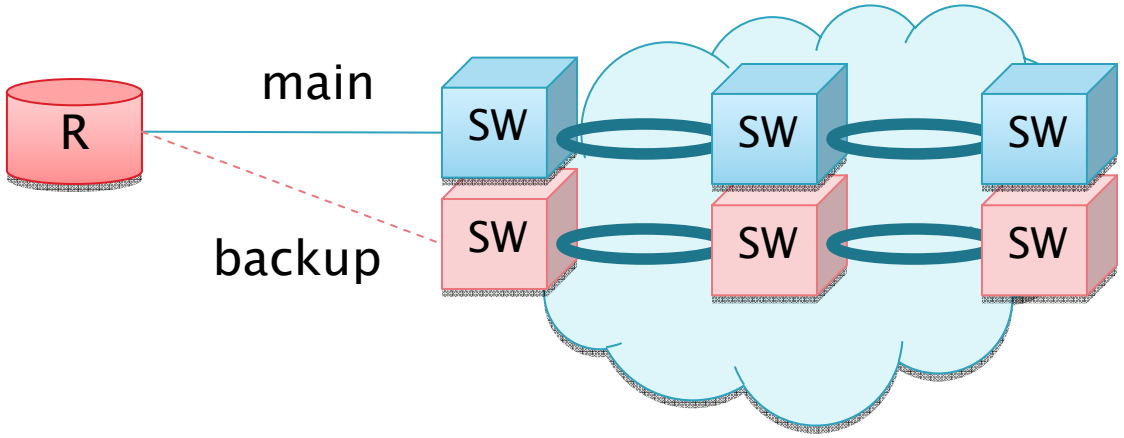


Outline

- Brief introduction of JPNAP
- An availability model of exchange point
- JPNAP solution for improving availability
- JPNAP case study

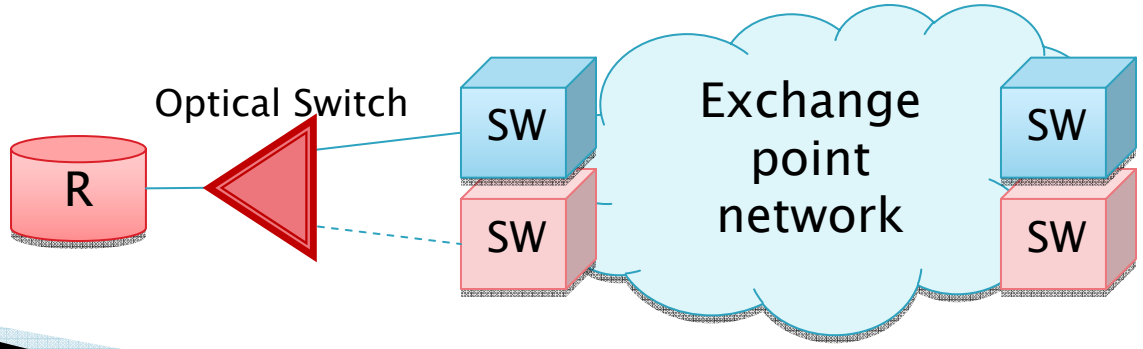
How does JPNAP improve availability?

- (1) Switches and network are redundant
  - We provide main and backup ports to all customers



How does JPNAP improve availability?

- (2) Optical switch for customer port
  - It instantaneously changes troubled port to backup in tens of milliseconds.
  - This means most routers do not sense link down, consequently the BGP sessions are not down.
  - Also maintenance can be done by changing some of customer ports to backup.



Outline

- Brief introduction of JPNAP
- An availability model of exchange point
- JPNAP solution for improving availability
- JPNAP case study

Troubles and maintenance works at JPNAP Tokyo I

From June 2007 to May 2008 (1)

	Date	Start	Recovered	Reason	Impact	Detailed reason / Workaround
1	2007/7/6	8:26:28	8:27:00	switch trouble	potentially all the customers	one line card reset (for backbone)
2	2007/7/7	3:00:00	6:36:00	maintenance	Approx. 10 customers, and their peers	some line cards replaced, software upgraded, and rebooted
3	2007/7/15	3:00:00	6:48:00	maintenance	Approx. 20 customers, and their peers	some line cards replaced, software upgraded, and rebooted
4	2007/7/29	3:00:00	8:00:00	maintenance	no impact	switching fabric modules replaced (hot swapped)
5	2007/11/21	14:20:57	14:25:06	human error	two specific customers	forgot to change configuration (regarding port security)
6		15:36:17	15:42:32			
7	2007/12/1	3:00:00	4:58:00	maintenance	Approx. 5 customers, and their peers	some line cards replaced
8	2007/12/28	18:50:00	19:49:00	switch trouble	one specific customer, and its peers	one port failed partially

Troubles and maintenance works at JPNAP Tokyo I

From June 2007 to May 2008 (2)

	Date	Start	Recovered	Event	Impact	Reasons or Workarounds
9	2008/1/26	7:24:52	8:22:00	Switch trouble	Onespecific customer and its peers	Under investigation: One possibility is that switch sent unnecessary frames (due to software bug)
10	2008/2/11	12:19:00	12:24:00			
11	2008/3/1	3:00:00	8:00:00	Maintenance	Approx. 30 customers	Upgrade switch software, and reboot
12	2008/3/15	15:59:46	15:59:47	Switch trouble	Potentially all the customers	Unexpectedchange to backup links in backbone
13	2008/3/22	3:00:00	6:09:00	Maintenance	Approx. 20 customers	Change backbone topology (to add new switches)
14	2008/3/29	3:00:00	6:08:00	Maintenance	Approx. 20 customers	Change backbone topology (to add new switches)
15	2008/4/4	12:32:57	12:32:57	Human error	Two specific customers and their peers	unintended power cycle of transmission device
16	2008/4/16	1:39:53	1:40:24	Switch trouble	Potentially all the customers	Unexpected change to backup links in backbone
17	2008/4/26	3:00:00	8:00:00	Maintenance	Approx. 10 customers	Replace hardware and reboot
18	2008/5/27	6:00:00	6:55:00	Maintenance	Approx. 10 customers	Replace transmission device

Focusing on maintenance work (1) "*no redundancy*" case

- If we did not have redundant network, the IX service would not be available during the maintenance window:
- 26 hours and 34 minutes

	Date	Start	Recovered	**Impact time (1)**	Impact on	Detailed reason
2	2007/7/7	3:00:00	6:36:00	**3:36:00**	Approx. 10 customers	some line cards replaced, software upgraded, and rebooted
3	2007/7/15	3:00:00	6:48:00	**3:48:00**	Approx. 20 customers	some line cards replaced, software upgraded, and rebooted
4	2007/7/29	3:00:00	8:00:00	0:00:00	no impact on all customers	switching fabric modules replaced (hot swapped)
7	2007/12/1	3:00:00	4:58:00	**1:58:00**	Approx. 5 customers	some line cards replaced
11	2008/3/1	3:00:00	8:00:00	**5:00:00**	Approx. 30 customers	Software upgraded, and rebooted
13	2008/3/22	3:00:00	6:09:00	**3:09:00**	Approx. 20 customers	add new switches
14	2008/3/29	3:00:00	6:08:00	**3:08:00**	Approx. 20 customers	to add new switches
17	2008/4/26	3:00:00	8:00:00	**5:00:00**	Approx. 10 customers	Replace hardware and reboot
18	2008/5/27	6:00:00	6:55:00	**0:55:00**	Approx. 10 customers	Replace transmission device

Focusing on maintenance work (2) "*switch network redundancy*"

- If the switch network was redundant but customer cables had to be removed and inserted to backup port manually,
- each customer's downtime would be estimated 30 seconds (remove and insert).
- 62 minutes 30 second

	Date	Start	Recovered	**Impact time (2)**	Impact on	Detailed reason
2	2007/7/7	3:00:00	6:36:00	**0:05:00**	Approx. **10**customers	some line cards replaced, software upgraded, and rebooted
3	2007/7/15	3:00:00	6:48:00	**0:10:00**	Approx. **20**customers	some line cards replaced, software upgraded, and rebooted
4	2007/7/29	3:00:00	8:00:00	0:00:00	no impact on all customers	switching fabric modules replaced (hot swapped)
7	2007/12/1	3:00:00	4:58:00	**0:02:30**	Approx. **5**customers	some line cards replaced
11	2008/3/1	3:00:00	8:00:00	**0:15:00**	Approx. **30** customers	Software upgraded, and rebooted
13	2008/3/22	3:00:00	6:09:00	**0:10:00**	Approx. **20** customers	add new switches
14	2008/3/29	3:00:00	6:08:00	**0:10:00**	Approx. **20** customers	to add new switches
17	2008/4/26	3:00:00	8:00:00	**0:05:00**	Approx.**10** customers	Replace hardware and reboot
18	2008/5/27	6:00:00	6:55:00	**0:05:00**	Approx.**10** customers	Replace transmission device

Focusing on maintenance work (3) "*customer port redundancy*"

- Optical switch is used for customer ports,
- Each customer's downtime is approximately tens of millisecond.
  - router interface does not sense link-down, accordingly bgp session is kept up.
- approximately 15 seconds

	Date	Start	Recovered	**Impact time (3)**	Impact on	Detailed reason
2	2007/7/7	3:00:00	6:36:00	**0:00:01**	Approx. **10**customers	some line cards replaced, software upgraded, and rebooted
3	2007/7/15	3:00:00	6:48:00	**0:00:02**	Approx. **20**customers	some line cards replaced, software upgraded, and rebooted
4	2007/7/29	3:00:00	8:00:00	0:00:00	no impact on all customers	switching fabric modules replaced (hot swapped)
7	2007/12/1	3:00:00	4:58:00	**0:00:01**	Approx. **5**customers	some line cards replaced
11	2008/3/1	3:00:00	8:00:00	**0:00:05**	Approx. **30** customers	Software upgraded, and rebooted
13	2008/3/22	3:00:00	6:09:00	**0:00:02**	Approx. **20** customers	add new switches
14	2008/3/29	3:00:00	6:08:00	**0:00:02**	Approx. **20** customers	to add new switches
17	2008/4/26	3:00:00	8:00:00	**0:00:01**	Approx.**10** customers	Replace hardware and reboot
18	2008/5/27	6:00:00	6:55:00	**0:00:01**	Approx.**10** customers	Replace transmission device

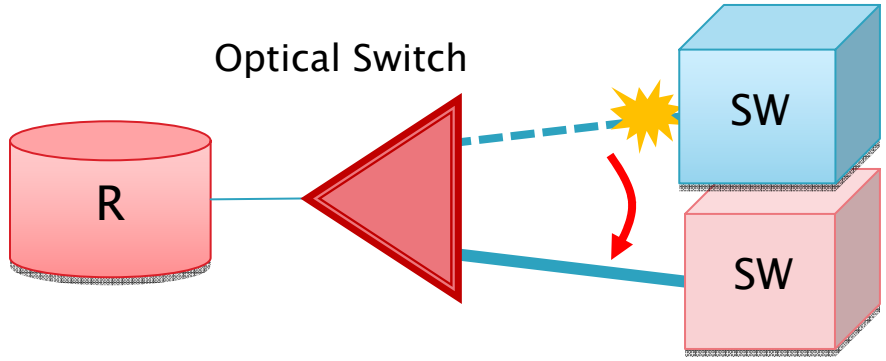
How much availability differs?

- Focusing only on these maintenance works,downtime/availability is in each case
  - (1) 26h 34m 0.30% (99.70%)
  - (2) 62m 30s 0.0033% (99.9967%)
  - (3) 15 s 0.000047% (99.999953%)
    - From Jun 1, 2007 to May 31, 2008 (366 days)

Focusing on human error

- A human error lead to a loss of light.

	Date	Start	Recovered	Event	Impact	Reasons or Workarounds
15	2008/4/4	12:32:57	12:32:57	Human error	Two specific customers and their peers	unintended power cycle of transmission device



Focusing on human error

- Bad scenario
  - Probably it takes about 30 seconds to notice the mistake and recover it…

	Date	Start	Recovered	impact	Impact	Reasons or Workarounds
15	2008/4/4	12:32:57	???	**30 seconds?**	Two specific customers and their peers	unintended power cycle of transmission device

- Actual story
  - Optical switch detected the failure of light and changed to the backup port in tens of millisecond!

	Date	Start	Recovered	Impact	Impact	Reasons or Workarounds
15	2008/4/4	12:32:57	12:32:57	0:00:01	Two specific customers and their peers	unintended power cycle of transmission device

Availability of JPNAP Tokyo I

- 99.976%(in this STRICT model)
  - Took time to determine whether failure on customer side or us
- 99.9998% (in usual sense)
  - 64sec per year (99.9998%)

	Date	Start	Recovered	Impact time	Reason	Impact	Detailed reason / Workaround
1	2007/7/6	8:26:28	8:27:00	**0:00:32**	switch trouble	potentially all the customers	one line card reset (for backbone)
5	2007/11/21	14:20:57	14:25:06	**0:04:09**	human error	two specific customers	forgot to change configuration (regarding port security)
6		15:36:17	15:42:32	**0:05:15**			
8	2007/12/28	18:50:00	19:49:00	**0:59:00**	switch trouble	one specific customer, and its peers	one port failed partially
9	2008/1/26	7:24:52	8:22:00	**0:57:08**	Switch trouble	Onespecific customer and its peers	Under investigation: One possibility is that switch sent unnecessary frames (due to software bug)
10	2008/2/11	12:19:00	12:24:00	**0:00:30**			
12	2008/3/15	15:59:46	15:59:47	**0:00:01**	Switch trouble	Potentially all the customers	Unexpectedchange to backup links in backbone
16	2008/4/16	1:39:53	1:40:24	**0:00:31**	Switch trouble	Potentially all the customers	Unexpected change to backup links in backbone

Summary

- Presented an availability model of exchange point.
- Using the model, JPNAP availability was calculated.

- Availability can be significantly improved at JPNAP by:
  (1) redundant switch network, and
  (2) optical switches.

Thank you!

- Any questions or comments?

- Contact: info jpnap.net