Alcatel·Lucent

IMS Presence Server: Traffic Analysis & Performance Modelling

Caixia Chi
chic@alcatel-lucent.com

Ruibing Hao
rbhao@yahoo.com

Dong Wang
wangd01@gmail.com

Z. Cao
caozhenzhen1983@gmail.com

Presented

By

Zhenhua Liu

Computer Science & Technology, Tsinghua University, China

Outline

- Background Introduction
- Presence Server Traffic Load Analysis
- Traffic Process Modelling of a Presence Server
    - An Accurate Model
    - A Simplified Model
- Conclusion

Outline

- Background Introduction
- Presence Server Traffic Load Analysis
- Traffic Process Modelling of a Presence Server
    - An Accurate Model
    - A Simplified Model
- Conclusion

Presence Service Introduction

- Presence is a service that allows a user to be informed about the reachability, availability, and willingness of communication of another user.
- Able to indicate users' status: online or not; Idle or busy; communication means and capabilities: SMS, MMS, 3G,2G Phone….
- A key enabler for many popular applications: push-to-talk (PTT),instant messaging (IM).
- 3G IP Multimedia Subsystem (IMS) already has presence service well supported in its architecture.

SIP Presence Architecture

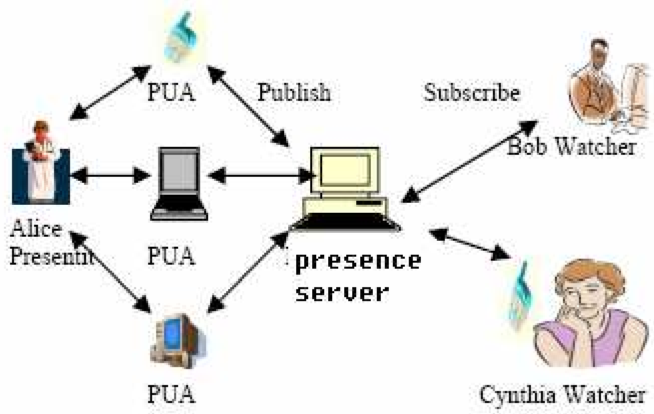
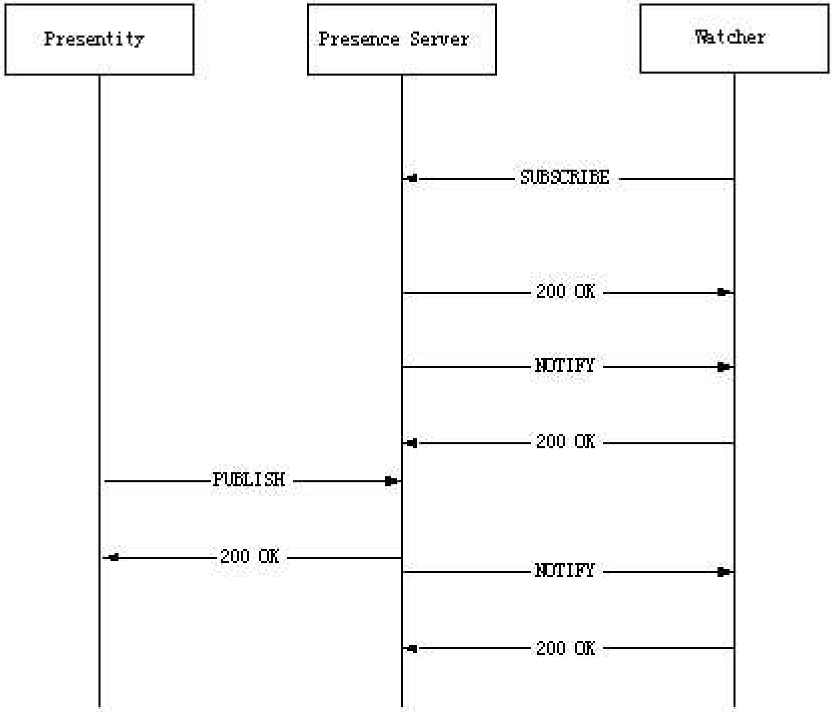


Figure 1-SIP Presence Architecture

- Alice is the publisher whose presence information is published to a presence server by PUA.
- PUA is the presence user agent of a user.
- Presence Agent is a presence server that is responsible for managing presence information.
- Bob/Cynthia are watchers who subscribes to information from PS on Alice's presentity.

Alcatel·Lucent

SIP Presence Service Message Flow



Alcatel·Lucent

Outline

- Background Introduction
- Presence Server Traffic Load Analysis
- Traffic Process Modelling of a Presence Server
  - An Accurate Model
  - A Simplified Model
- Conclusion

User Behavior & Traffic Characterization

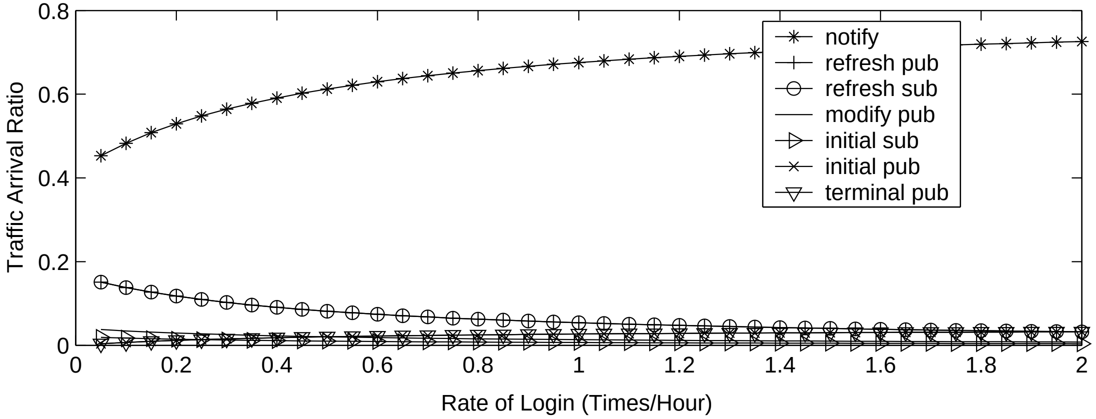
- Login and Logout
  - A user's login generates an initial PUBLISH message to the PS
  - Refresh PUBLISH messages are generated periodically
  - Terminating PUBLISH message is sent to the PS upon the user's logout
- Presence Subscription
  - Subscription of a user's presentity results in an initial SUBSCRIBE message being sent to the PS,
  - Refresh SUBSCRIBE messages will be sent to the PS periodically
  - Terminating SUBSCRIBE message un-subscribes the other user's presentity.
- Presence Status Updates
  - Change of a user's status results in a modifying PUBLISH message,

Traffic Types

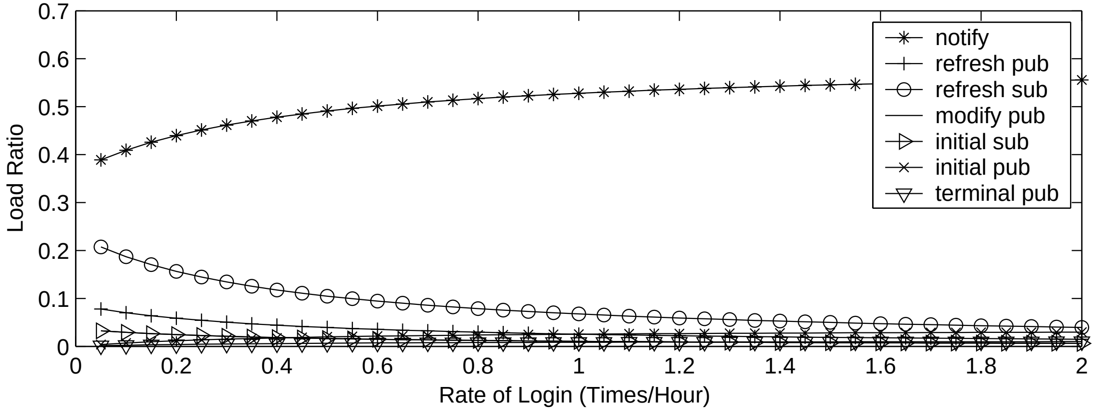
- Traffic to a PS is divided into eight types:
  - initial/refresh/modifying/terminal publish
  - initial/refresh/terminal subscribe
  - notify
- Process Time for each Traffic type is Different (a reference implementation)
  - $t\_initial\_pub = 6ms$ $t\_refresh\_pub = 3ms$ $t\_terminal\_pub = 2ms$
  - $t\_modify\_pub = 5ms$ $t\_notify = 5ms$ $t\_initial\_sub = 10ms$
  - $t\_refresh\_sub = 8ms$ $t\_terminal\_sub = 3ms$

Alcatel·Lucent

Ratio of Traffic Load vs. User Login Rate , *online watcher* = 10



(1)

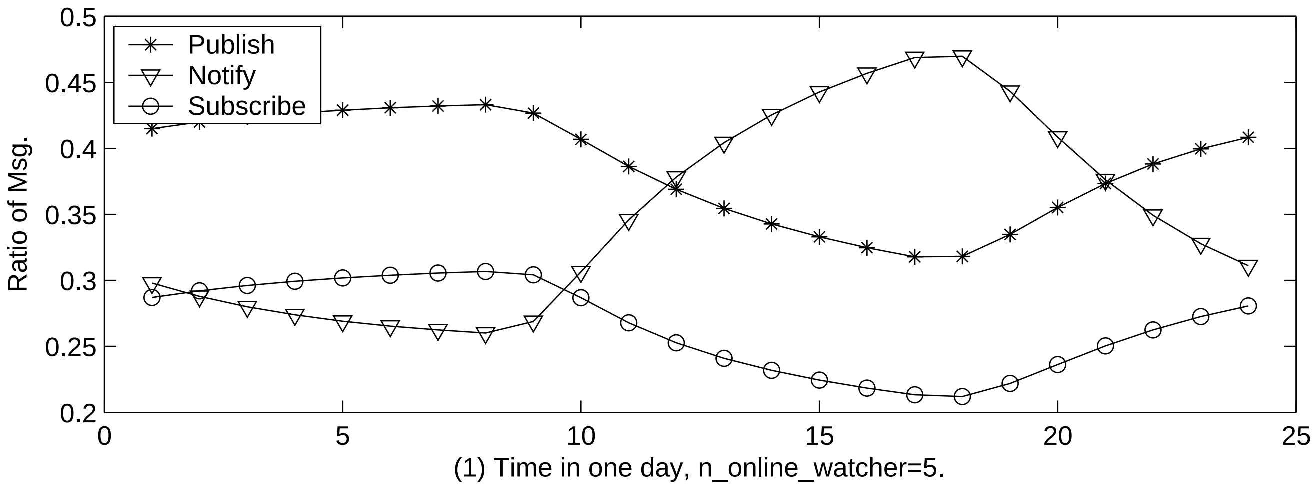


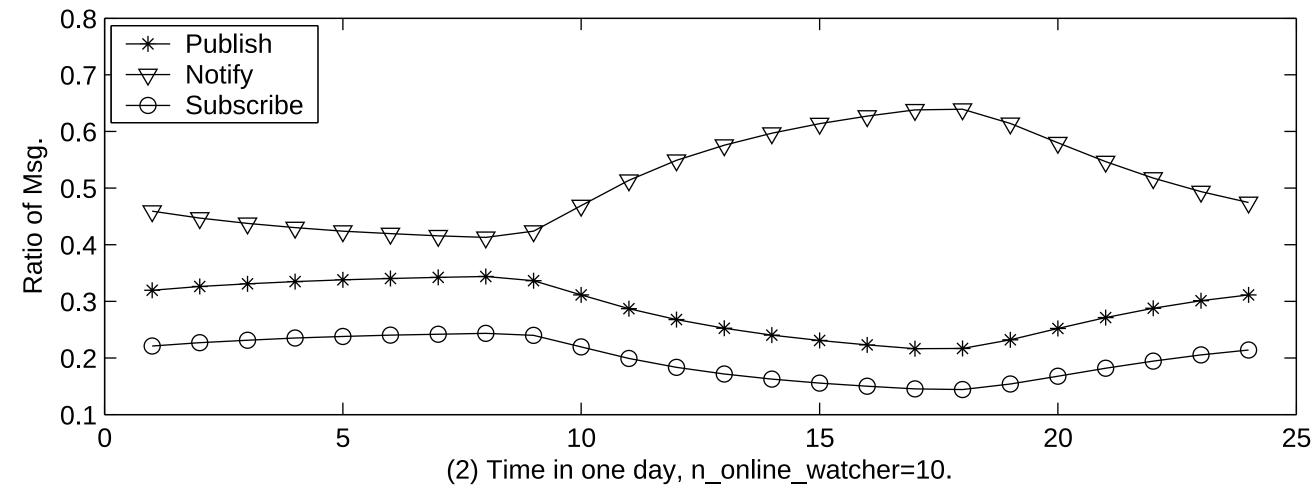
(2)

- Traffic Load increases with login frequence increases.
- Notify messages is the largest part of the traffic load.
- Refresh PUBLISH and refresh Subscribe has the same traffic arrival rate, but Refresh Subscribe has more heavier load.

Alcatel·Lucent

Messages Ratio Distribution During the Time of a Day.





- Traffic to a PS varies during the time of a day
- Notify messages are the largest part of all the traffic load when online watcher number greater than 10

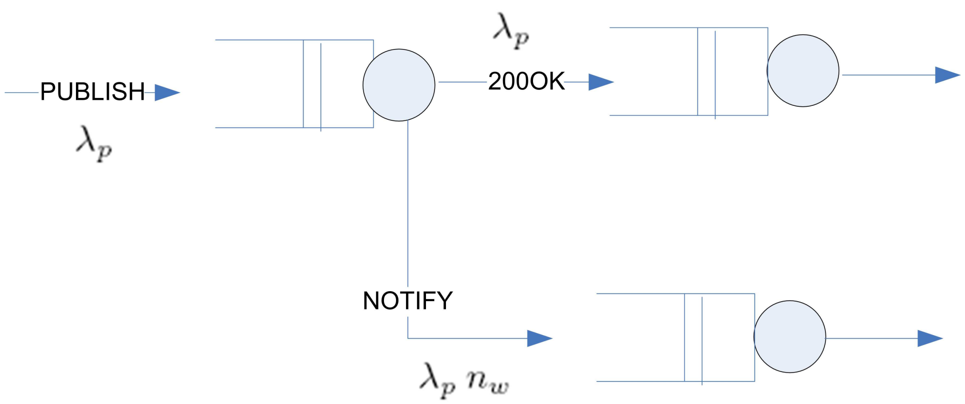
Presence Server Traffic Characterization

- NOTIFY messages have great impact on traffic load to a PS
- Traffic rate to a PS varies greatly with the time of a day.
- With each message to/from PS will be relayed by IMS Core network, NOTIFY message process will be critical for whole IMS network.
- NOTIFY message process mechanism in a PS is studied in the following.

Outline

- Background Introduction
- Presence Server Traffic Load Analysis
- Traffic Process Modelling of a Presence Server
  - An Accurate Model
  - A Simplified Model
- Conclusion

Publish and Notify Queues in a PS

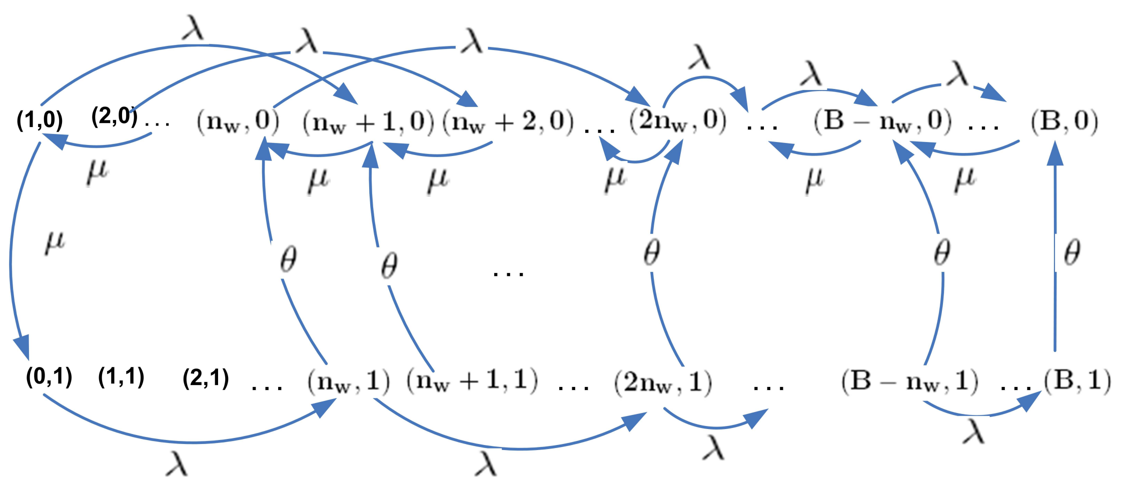


- A PUBLISH message arrives a PS, n_w NOTIFY messages are generated.
- 200OK messages are sent out as soon as possible to prevent retransmission
- NOTIFY messages are buffered in a queue and controlled by a scheduling mechanism.

Note: nw is the number of online watchers

- *Notify messages queue is modelled as a Queueing System with Controlled Vacation and Batch Poisson Arrival*

State Transition Graph of NOTIFY Queue



- Based on the state transition graph, balance equations can be used to get probability of system states and the relationship between vacation time ($T = \frac{1}{\theta}$) and state probability.

Optimization Problem from the Analysis

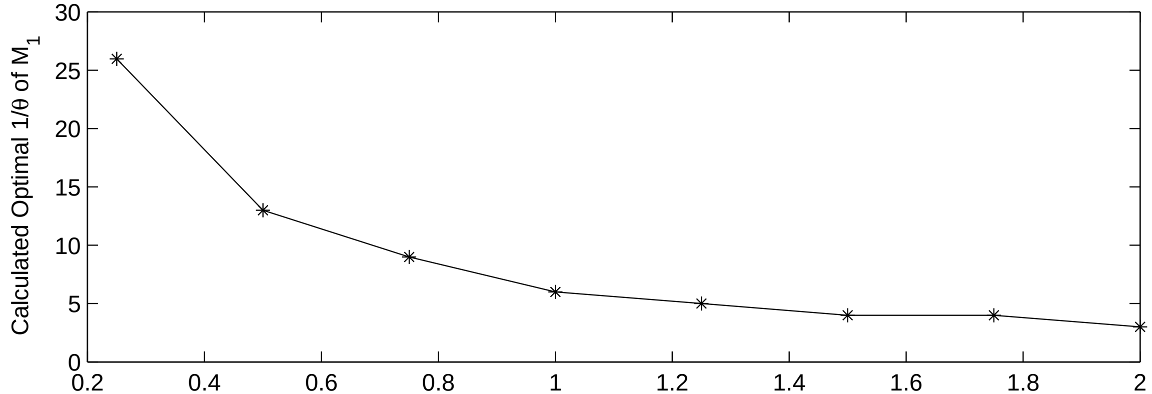
$$\min \theta \qquad (34)$$

$$s.t.$$

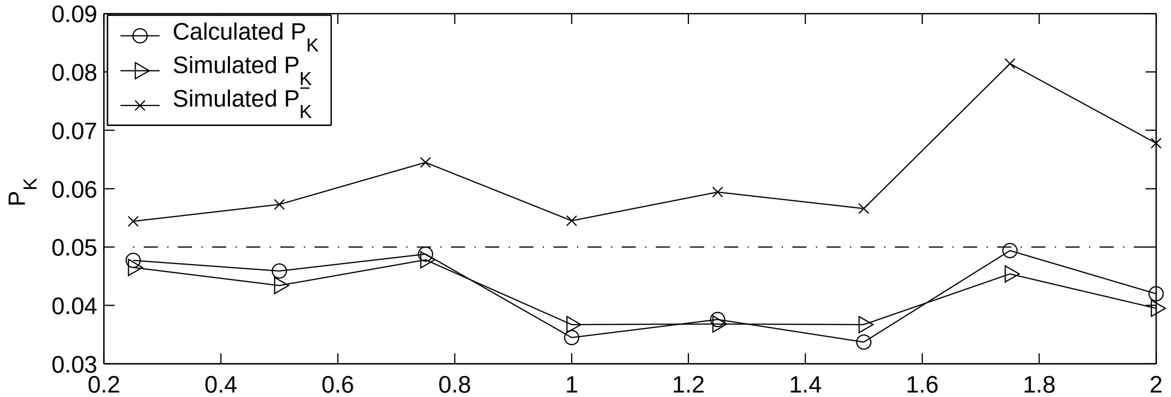
$$P_K \leq q \qquad (35)$$

- Minimize the busy time of the Presence Server such that
- The probability of message queue length greater than *K* should be less than *q*,

Optimal Notify Timers



(1) λ (μ=6, K=80%B, q=0.05)



(2) λ (μ=6, K=80%B, q=0.05)

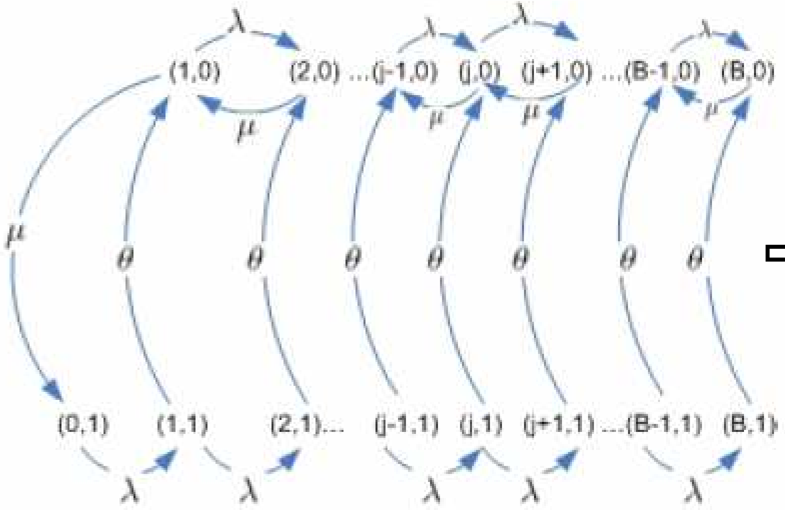
- A higher arrival rate needs a shorter vacation time.
- Timer is not very sensitive to the traffic arrival rate
- Traffic load can be divided into several levels and each level is associated with a control timer value.
- Optimal timer value, the loss probability of NOTIFY queue can vary for difference traffic load as indicated in Fig.10(2),but they all satisfies the constraint.

Alcatel·Lucent

Outline

- Background Introduction
- Presence Server Traffic Load Analysis
- Traffic Process Modelling of a Presence Server
  - An Accurate Model
  - A Simplified Model
- Conclusion

A Queue System with Controlled Vacation and Poisson Arrival (A Simplified Model)



$$
\begin{aligned}
\lambda p_{0,1} &= \mu p_{1,0} \\
(\lambda+\mu)p_{1,0} &= \mu p_{2,0} + \theta p_{1,1} \\
(\lambda+\mu)p_{j,0} &= \lambda p_{j-1,0} + \mu p_{j+1,0} + \theta p_{j,1}, \\
& \quad j = 2,\cdots,B-1. \\
\mu p_{B,0} &= \lambda p_{B-1,0} + \theta p_{B,1} \\
(\theta+\lambda)p_{j,1} &= \lambda p_{j-1,1}, j = 1,\cdots,B-1. \\
\theta p_{B,1} &= \lambda p_{B-1,1} \\
\sum_{j=1}^{B} p_{j,0} + \sum_{j=0}^{B} p_{j,1} &= 1
\end{aligned}
$$

$$
\begin{aligned}
p_{0,1} &= \frac{\theta}{\lambda}\left(1+\frac{\theta}{\lambda}\right)^{B-1} p_{B,1} \\
p_{j,1} &= \frac{\theta}{\lambda}\left(1+\frac{\theta}{\lambda}\right)^{B-1-j} p_{B,1} \\
& j = 1,\cdots,B-1. \\
p_{1,0} &= \frac{\lambda}{\mu} p_{0,1} \\
p_{j,0} &= \frac{\lambda}{\mu} p_{j-1,0} + \frac{\lambda}{\mu} p_{0,1} - \frac{\theta}{\mu}\sum_{k=1}^{j-1} p_{k,1}, \\
& j = 2,\cdots,B.
\end{aligned}
$$

$$
\begin{aligned}
p_{B,1} &= \frac{\mu-\lambda}{\mu(1+\frac{\theta}{\lambda})^B + \mu\times J - \lambda\times(1+\frac{\theta}{\lambda})^B - I\lambda} \\
I &= \frac{\theta}{\lambda}\left(1+\frac{\theta}{\lambda}\right)^{B-1}\frac{\lambda\left(1-(\frac{\lambda}{\mu})^B\right)}{\mu-\lambda} \\
&\quad - \frac{\theta}{\mu}\frac{\theta}{\lambda}\sum_{j=0}^{B-2}\left(\frac{\lambda}{\mu}\right)^j\sum_{k=1}^{B-j-1}\left(1+\frac{\theta}{\lambda}\right)^{B-k-1} \\
J &= \frac{\lambda(B-1)}{\mu}\frac{\theta}{\lambda}\left(1+\frac{\theta}{\lambda}\right)^{B-1} \\
&\quad - \frac{\theta}{\mu}\sum_{j=1}^{B-1}(B-j)\frac{\theta}{\lambda}\left(1+\frac{\theta}{\lambda}\right)^{B-j-1}
\end{aligned}
$$

$$\min \theta$$

$$s.t.$$

$$P_K \leq q$$

$$P_K = \sum_{j=K+1}^{B} (p_{j,1} + p_{j,0})$$

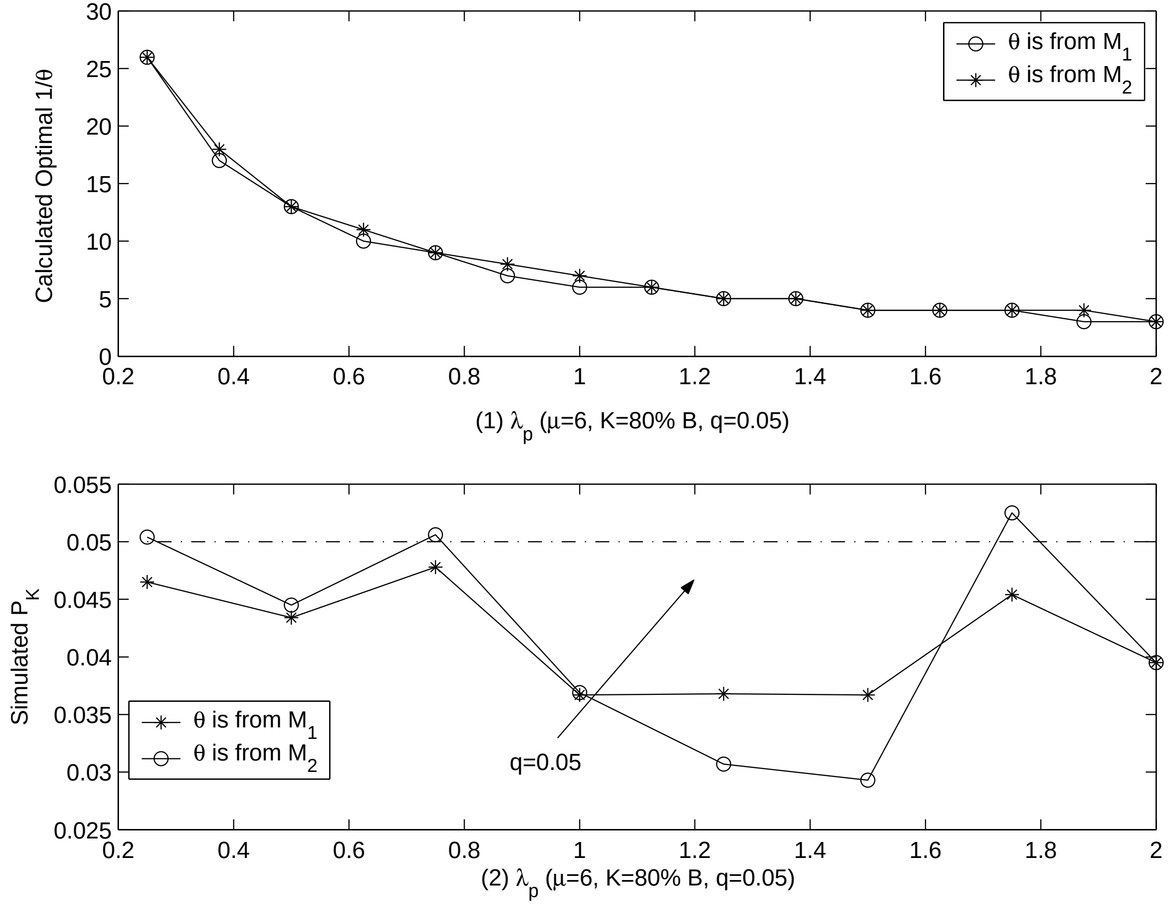
Buffer Control Optimization Algorithm 1:

begin

1. Given PUBLISH traffic arrival rate λ_p, number of online watchers n_w, QoS parameter K, q and timer interval $I = [L, U]$,$i = 1, maxI = L$;
2. for i=L:U /* loop from L to U */
   $\theta = 1/i$,
   Solve equations array M_1 to get $p_{j,0}, p_{j,1}, j = 0, \cdots, B$
   Get P_K from equation $P_K = \sum_{j=K+1}^{B} (p_{j,1} + p_{j,0})$
   if $P_K \leq q$ and $i > maxI$, $maxI = i$;
   end
3. $\theta = \frac{1}{maxI}$,Return

end

Experimental Result for Simplified Model



- The simplified model can satisfy the constraint at most of the time

Conclusion

- NOTIFY messages account for the largest portion of traffic load to a PS and their processing overhead has great impact on the quality of service of the PS.
- Queuing system with Controlled Vacation and Batch Poisson Arrival can be used to model Notify message queue.
- Simplified queuing system with Controlled Vacation and Poisson Arrival can be used to optimize vacation timer.
- Further work is needed to apply the analysis to the problems such as network sizing, traffic admission control.

For any questions, please contact

Caixia Chi	chic@alcatel-lucent.com
Ruibing Hao	rbhao@yahoo.com
Dong Wang	wangd01@gmail.com
Z. Cao	caozhenzhen1983@gmail.com

Thank you!