First Exceed Level 이론이 적용된 확률적 예측 가능한 Hitless 라우터 설계

김 송 규
삼성전자 정보통신총괄, 경기도 수원시 영통구 매탄동 416번지
amang.kim@samsung.com

Abstract

본 연구는 hitless-restart기능과 확률적인 예측을 할 수 있는 발전된 형태의 Hitless 라우터(Router) 설계를 제안한다. Hitless-restart기능이라 함은 라우터가 reset 혹은 shutdown이 되더라도 forwarding path와 네트워크 구조는 유지하는 것을 말한다. 그러나 현재 hitless-restart의 가장 큰 문제점은 라우터가 restart를 할 때를 대비하여 항상 active한 상태로 유지시켜야 한다는 것이다. 확률적 예측이 가능한 Hitless 라우터는 restart할 시점을 확률적으로 예측함으로써 라우터 시스템을 보다 효과적으로 운영할 수 있도록 한다. First Exceed Level이론은 라우터의 조건에 따라 restart가 필요한 시점을 확률 적으로 예측할 수 있도록 한다. 이러한 예측결과를 이용하여 우리는 라우터가 구조적인 한계를 넘어서기 전에 hitless-restart를 실시함으로써 라우터가 shutdown되는 것을 방지할 수 있다.

1. Introduction

라우터(Router)의 안정성(Reliability)는 네트워크 시스템에 있어서 가장 중요한 문제 가운데 하나이며, 라우터의 Queue가 overflow하는 것을 방지하는 것은 라우터의 안정성을 향상 시키는 좋은 방법이 될 수 있다. 그동안 네트워크 장비 사업자들은 라우터시스템이 어떠한 이유에서 새로 시작해야 하는 상황이 되더라도, hitless한 상태에서, 이용자들이 계속해서 서비스를 받을 수 있도록 제공하고자 노력 하였다 [1]. Hitless-restart(혹은 nop-stop forwarding)는 라우터의 시스템이 restart를 하더라도, 라우터의 forwarding path를 유지할 뿐 아니라, 네트워크의 Topology를 유지하도록 한다 [5]. 일반적으로 이러한 restart는 라우터의 queue가 수용할 수 있는 용량을 넘어설 경우 발생한다.

현재, 많은 업체들이 high-availability 혹은 carrier-class 시스템을 개발하고 있으며, 이러한 시스템들은 링크 layer가 보다 안정적이고, 네트워크의 fail over를 염두하여 만든다 [1]. Hitless-restart 라우터의 가장 큰 목적은 라우터의 restart가 네트워크에 영향을 주지 않도록 하고, 라우터 내의 트래픽이 계속해서 운용될 수 있도록 하는데 있다. 현재의 hitless-restart 라우터가 가지고 있는 가장 큰 문제점은 restart를 하는 시점에서 라우터가 동작이 불가능하게 되어서는 안 된다. 또한, hitless-restart가 시작되면, 라우터의 업그레이드나 유지, 보수와 같은 일들을 행하게 된다. 이러한 일련의 작업은 라우터가 동작불능인 상태가 되어서는 할 수 없는 작업들이다.

이 글은 이러한 기존의 hitless-restart 라우터의 기능을 개선하기 위해, first exceed level 이론[2]을 적용하여 확률적인 예측이 가능한 hitless-restart 라우터를 모델링 하고자 한다. First exceed level 이론은 특정하게 정해진 level에 대한 marked 포인트 프로세스에 대한 해석을 뜻 한다. Dshalalow[2]는 first exceed level 프로세스를 해석하여 first passage time, first passage index에 대한 joint function을 구해 내었다. 확률적 예측이 가능한 hitless-restart 라우터는 기본적으로 라우터의 queue가 overflow되기 전에 restart action을 취하는 라우터를 말한다. 만약, 라우터내의 트래픽(traffic)을 처리하는 queue의 크기가 특정 한계치 S를 넘어설 경우, 우리는 queue가 overflow한 것으로 간주하며, router는 이러한 queue의 overflow로 인해, shutdown이 될 수 있음을 의미 한다. 관찰 프로세스(observation process)는 무작위로 queue사이즈를 관찰하고, queue가 한계치 S를 넘기 전 시점을 예측하여, 미리 hitless-restart를 실행한다. 이러한 한계치 S를 hitless-prediction point라고 칭한다.

만약, k번째 관찰시점(observation point)에서 queue의 한계치 S를 넘어선 후 처음 관찰한 시점이라고 하면, k번째 관찰시점을 *first exceed observation point*라고 칭한다. 이 경우, ($k-1$)번째 관찰시점에서 hitless-restart가 행해진다. 라우터는 이 hitless-restart시점에는 non-stop forwarding과 동시에 프로그램의 업그레이드, 자가진단 등을 행할 수 있다. 이러한 일련의 과정들은 reset point까지 진행되며, 이 시점이 지나면 예측부터 non-stop forwarding까지의 과정이 반복되게 된다. 확률적 예측이 가능한 hitless-restart 모델은 위의 상황을 수학적으로 해석 할 수 있으며, hitless-restart 시점과, reset 시점까지의 평균 간격을 계산 할 수 있다.

이 논문은 다음과 같은 구성으로 되어 있다. 2장에서는 확률적 예측이 가능한 hitless-restart 라우터에 대한 모델을 hitless-prediction model이라 하고, 이 것의 기본이 되는 확률모델과 first exceed level이론의 적용방법을 다루고 있다. 3장에서는 실질적인 적용과정과 가상실험을 통한 발전된 hitless-restart 라우터의 효과에 대한 비교를 하였고, 마지막 5장에서는 결론과 향후 발전 방향에 대해서 언급하였다.

2. Stochastic Hitless-prediction Model

C_k는 queue에 있는 패킷의 수를 의미하고, D_k는 k번째 관찰간격을 의미(observation period)한다. τ_k은 k번째 관찰시점(observation point)이라고 하면, 관찰간격은 $[\tau_{k-1},\tau_k)$가 된다. 처음으로 한계치 S를 넘어선 것을 관찰한 τ_k시점을 *first passage time*이라고 칭한다. 또한 이 시점에서의 k의 값을 υ[2]로 정의하며, *termination index*라고 부른다. 이 경우, hitless-restart가 시작되는 시점은 τ_{k-1}가 된다. 만약, Ak를 다음과 같이 정의 하자.

$$(A,\tau)=\sum_{k\geq 0} C_k\cdot \varepsilon_{\tau_k} \qquad (1)$$

그리고, C_k에 대한 joint function은

$$\gamma(z,\theta)=\mathbb{E}\left[z^{C_k}e^{-\theta\chi_k}\right] \qquad (2)$$

으로 정의 한다. 앞에서 언급한 termination index는

$$\nu=min\{k:A_k\geq S\} \qquad (3)$$

가 된다. 확률적 예측이 가능한 hitless-prediction 모델의 joint function은

$$L_{S-1}(\xi,z,\theta,\vartheta)=\mathbb{E}\left[\xi^{\nu}z^{A_\nu}e^{-\theta\tau_\nu}e^{-\vartheta\tau_{\nu-1}}\right] \qquad (4)$$

이 된다. 여기서 $S-1$은 queue가 overflow하기 전의 관찰시점(observation point)을 뜻한다. 이러한 first level exceed 이론의 해법은 Dshalalow [2]가 제시한 방법으로 풀 수 있다. 만약, functional

$$G(z)=D_p f(p):=(1-z)\sum_{p\geq 0} f(p)z^p, \qquad (5)$$

이라고 하면,

$$f(p)=\mathfrak{D}_z^p G(z)$$

이 된다. 여기에 functional

$$\mathfrak{D}_x^k(\bullet)=\begin{cases}\frac{1}{k!}\lim_{x\to 0}\frac{\partial^k}{\partial x^k}\frac{1}{1-x}(\bullet), & k\geq 0,\\ 0, & k<0.\end{cases}$$

이 정의 된다 [3]. 여기서, (4)는

$$L_{S-1}(\xi,z,\theta,\vartheta)=$$

$$\mathfrak{D}_w^{S-1}\frac{\mu_0(wz,\theta+\vartheta)}{\gamma(wz,\theta+\vartheta)}\cdot\frac{\gamma(z,\theta)-\gamma(wz,\theta)}{1-\xi\gamma(wz,\theta+\vartheta)}$$

와 같이 된다. 여기서, τ_{k-1}의 확률분포에 대한 moment generating function과 υ에 대한 generating function은

$$\mathbb{E}\left[e^{-\vartheta\tau_{\nu-1}}\right]=L_{S-1}(1,1,0,\vartheta) \qquad (6)$$

과(와)

$$\mathbb{E}[\xi^{\nu}]=L_{S-1}(\xi,1,0,0). \qquad (7)$$

이 된다. 만약, $k-1$번째 관찰 시점의 평균이나 termination index의 평균을 구하고자 한다면, (6)과 (7)을 이용하여 구하면,

$$\mathbb{E}[\nu]=\lim_{\xi-1}\frac{\partial}{\partial\xi}L_{S-1}(\xi,1,0,0)=\mathfrak{D}_w^{S-1}\frac{\mu_0(w,0)}{1-\gamma(w,0)}$$

과(와)

$$\mathbb{E}[\tau_{\nu-1}] = \lim_{\vartheta \to 0} -\frac{\partial}{\partial \vartheta} L_{S-1}(1,1,0,\theta)$$

$$= \mathfrak{D}_{\omega}^{S-1} \frac{\gamma'(\omega,0)(1-\gamma(\omega,0))(1-2\gamma(\omega,0))}{(\gamma(\omega,0)-\gamma^2(\omega,0))^2}$$

이 된다.

3. Simulation of Hitless-Prediction Router

이번 장에서 우리는 Hitless-prediction router를 실제로 설계 하고자 한다. 이 장에서는 설계의 간략화를 위해서 $k-1$번째 관찰 시점의 평균을 구한다 [3]. 그 해는

$$\mathbb{E}[\tau_{\nu-1}] = \tilde{\mu}$$

$$+\tilde{\gamma}\left[\frac{S}{\tilde{\chi}\lambda} - 1 + \frac{1-(\lambda_0/\lambda)}{(1+\tilde{\chi}\lambda_0)} \cdot \frac{1-[\tilde{\chi}\lambda_0/(1+\tilde{\chi}\lambda_0)]^S}{1-[\tilde{\chi}\lambda_0/(1+\tilde{\chi}\lambda_0)]}\right]$$

가 되고, 이 값은 hitless-restart의 생성 주기이다.

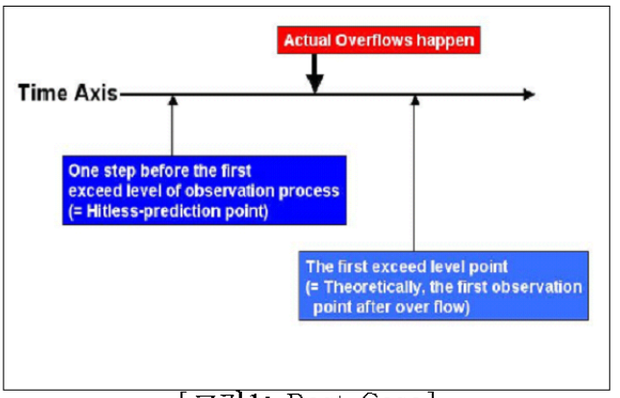
가상실험을 위해서 우리는 두개의 라우터를 고려한다. 한 가지는 예측기능이 없는 hitless-restart기능만을 가지는 라우터이며, 또 한 가지는 hitless-prediction model을 적용한 라우터 이다. 이론적으로 hitless-predict 라우터는 queue가 overflow되기 전에 non-stop forwarding과 같은 action을 하게 된다. 실질적으로 이러한 action은 overflow가 되기 전에 일어 날 수도 있고, overflow가 된 후에 일어날 수도 있다. 가상실험은 다음과 같은 전제조건을 가진다.

1) 관찰 프로세스는 queue가 overflow될 때까지를 한 주기로 갖는다.
2) 가상실험에서 한번의 시도는 queue가 overflow될 때까지로 한다.
3) Hitless-prediction은 크게 3가지 경우로 나누어 진다 [그림 1~3 참조].

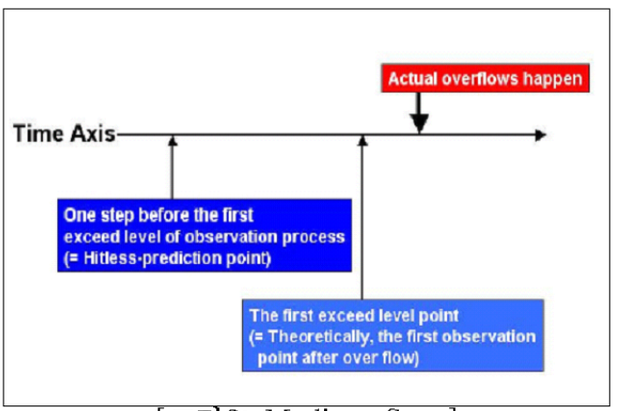
Best (그림1):
실질적인 overflow가 hitless-prediction 시점과 first exceed 시점 사이에서 일어난다.

Medium (그림2):
실질적인 overflow가 first exceed 시점 후에 일어난다.

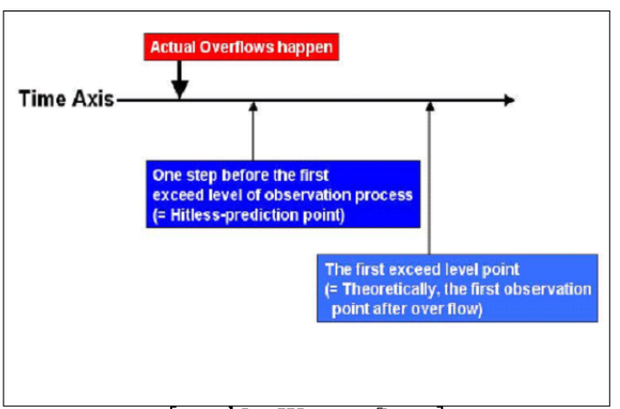
Worst (그림3):
실질적인 overflow가 hitless-prediction 시점보다 일찍 일어난다.



[그림1: Best Case]



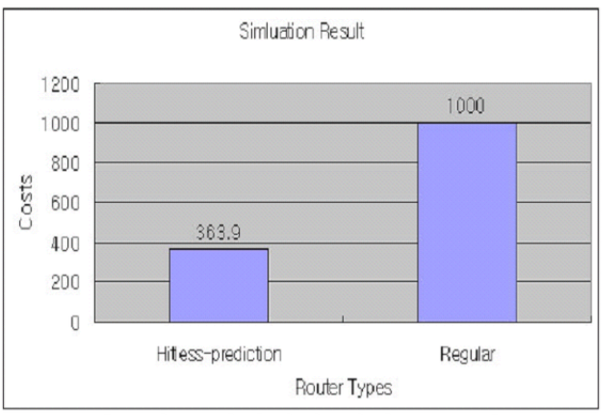
[그림2: Medium Case]



[그림3: Worst Case]

가상실험은 100의 시도 가운데, Best인 경우, medium인 경우 그리고, worst인 경우의 수를 conut한다. Best인 경우의 소요 비용을 1[unit원](여기서 unit은 백만이 될 수도 있고, 1000만이 될 수도 있다.), medium인 경우의 소요비용은 5[unit원] 그리고, worst인 경우의 소요 비용은 10[unit원]이라고 하자. 100번의 시도를 한 후, 소요비용에 대한 비교해보면 다음과 같다. 그림4와 같이, 일반 라우터의 경우, 총 소요비용은 1000[unit원]의 비용이

소요 되고, hitless-prediction 라우터의 경우는 363.9[unit원]의 소요비용이 필요하다.



[그림4: 가상실험 결과 Graph]

4, 결론

본 연구의 목표는 hitless-prediction 모델의 수학적인 증명과 응용에 있다. 이 모델은 확률적인 예측을 이용하여, 라우터가 crash되기전에 조치를 취할 수 있게 한다. 또한, 가상실험을 이용하여, 실제 라우터에 적용하였을 때의 performance를 비교하였다.

5. 참고문헌

[1] Cowburn, I., **Toward a Hitless Network: BGP Graceful Restart**, *Riverstone technology white paper* **134** (2002)
[2] Dshalalow, J. H., **First excess level process**, in *Advances in Queueing* (Edited by Dshahalow, J. H.), CRC Press, Boca Raton, FL, 244-261, 1995
[3] Kim, S. -K. and Dshalalow, J. H., **Stochastic disaster recovery systems with external resources**, *Mathematical and Computer Modelling* **36** (2002), 1235-1257
[4] Malbotra, R., *IP routing*, O' Reilly and Asso. Inc., Sebastopol, CA, 2002
[5] Moy, J., Padma, P. -E. and Lindem, A., **Hitless OSPF restart**, *IETF Draft* (2002)
[6] Shaikh, A., Dube, D. and Varma, A., **Avoiding Instability during Shutdown of OSPF**, *INFOCOM 2002 Proceedings*, New York, June 23-27, 2002