

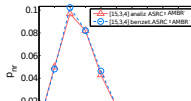
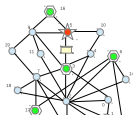
Kümeleme Tabanlı Dağıtık Depolama Sistemi ile Dosya Tamiri

File Repair with Cluster Based Distributed Storage System

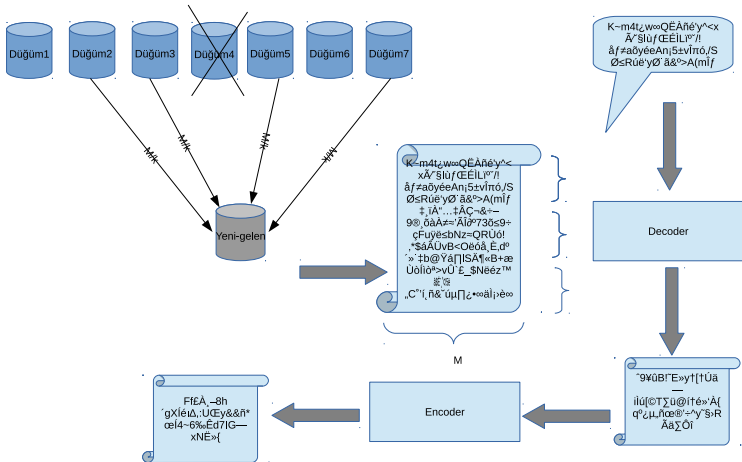
Elif HAYTAOĞLU¹
Mehmet Emin DALKILIÇ²

¹Bilgisayar Mühendisliği, Pamukkale Üniversitesi, Denizli
²Uluslararası Bilgisayar Enstitüsü, Ege Üniversitesi, İzmir

18 Mayıs 2017



Silinti Kodlama Kullanan Dağıtık Depolama Sistemlerinde Düğüm Tamiri



Şekil: Silinti Kodlama Kullanan Dağıtık Depolama Sistemleri

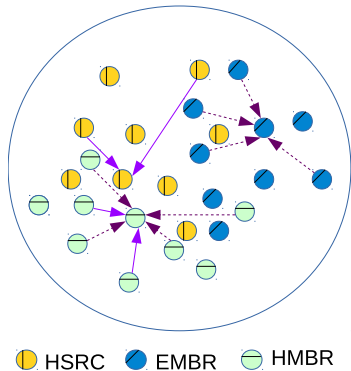
Motivasyon

- Dağıtık depolama sistemleri her zaman birbirlerine eş olan düğümleri barındırmazlar.
- Bant genişliği, CPU süresi, I/O kullanımı, toplam süre ve tamir başarımı maliyetleri farklı düğümler için farklı olabilmektedir.
- Bu maliyetler düğümden düğüme değişebileceği gibi bir düğüm içinde de zamanla değişebilmektedir.

Motivasyon (Dvm)

- Bir dağıtık depolama sistemini kaynak maliyetlerine göre kümeleyip her kümeye uygun silinti kodunu atamak bu sistemlerde düğüm tamiri işlem performansını artırabilir.
- Bu bildiride dağıtık depolama sistemini oluşturan düğümler kaynak maliyetlerine verdiği önceliklere göre kümelendirilmiştir.
- Ardından, HSRC [7], E-MBR [8] ve HMBR [9] silinti kodlarını uygun kümelere atanarak, düğüm tamiri maliyetleri incelenmiştir.
- Düğüm tamiri maliyetlerinin kümeleme tabanlı dağıtık depolama sistemlerinde %6 ila %44 arasında düştüğü görülmüştür.

Kümeleme Tabanlı Dağıtık Depolama Sistemi



Şekil: Kümeleme tabanlı dağıtık depolama sistemi.

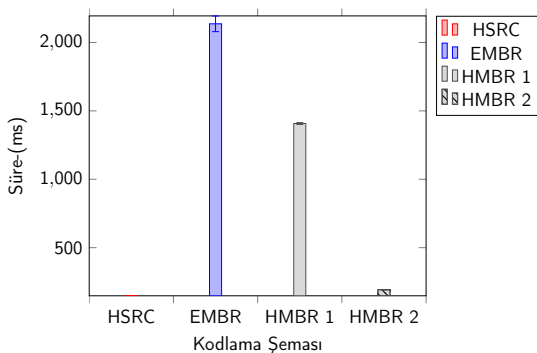
Benzetim Ayarları

- ns-3 [6] benzetim programı
- BRITE [11] topoloji üretme aracı, Waxman [10] topolojisi,
- 100 MB dosya yedekli olarak depolanıyor.
- HSRC [7] kümesini oluşturan düğümlerde CPU kullanım süresi diğer kümelere göre daha maliyetlidir.
 - $[n = 15, k = 3]$ olarak alınmıştır.
 - Bu kümede, bant genişliği 12 Mbps - 120 Mbps arasından eş dağılımlı atanmış, yayılım gecikmesi ise her bir hat için 80ms'dir.
- EMBR [8] kümesinde bant genişliği diğer kümelere göre daha kısıtlı olan düğümler bulunmaktadır,
 - Bant genişlikleri 9 Mbps- 90 Mbps değerleri arasından eş dağılımlı olarak atanmıştır.
 - Hatların yayılım gecikmesi ise 50 ms'dir
 - $[n = 15, k = 3, d = 4]$ olarak alınmıştır.

Benzetim Ayarları Dvm

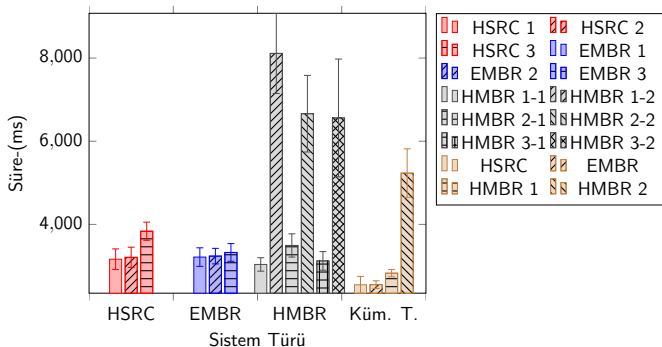
- Üçüncü kümede ise kullanılan kaynakların maliyeti zaman içerisinde değişebilmektedir. Bu kaynaklar CPU, I/O kullanımı ve bant genişliğidir.
- Bu kümede $[n = 15, k = 3, d = 4]$ HMBR kodları [9] kullanılmıştır.
- HMBR1, HMBR2
- Bu kümedeki hatların bant genişliği değerleri 10 Mbps-100 Mbps arasından eş dağılımlı olarak atanmıştır
- Hatların yayılım gecikmesi ise 50 ms'dir

Dosya Tamiri için Kullanılan Toplam CPU Süresi



Şekil: Dosya tamiri için kullanılan toplam CPU süreleri.

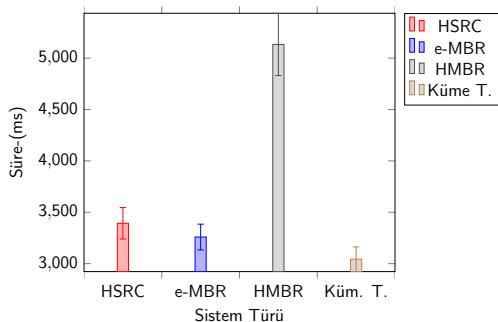
Dosya Tamiri için Geçen Toplam Süre



Şekil: Dosya tamiri için geçen süreler.

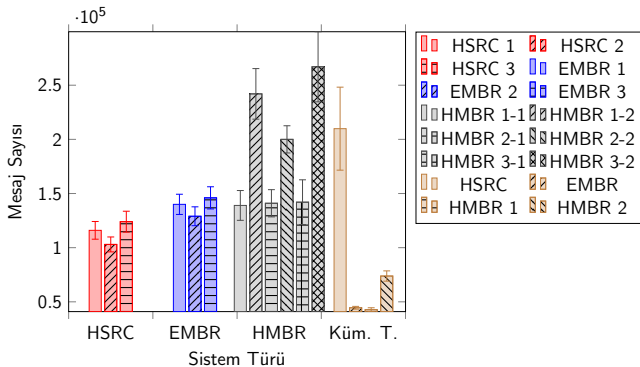
$$t_{\text{tamir}} = t_{\text{iletim}} + t_{\text{sonPARÇA}} \quad (1)$$

Kümeleme tabanlı olmayan dağıtık depolama sistemlerinde 100 MB dosya tamiri için geçen sürelerin ortalamaları



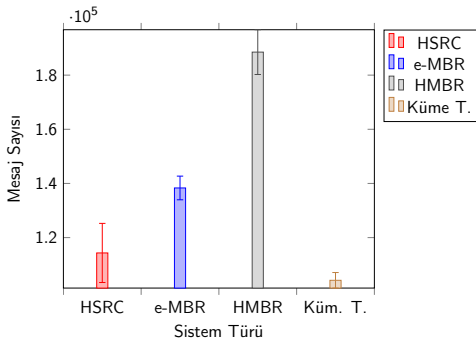
Şekil: 100 MB dosya tamiri için geçen sürelerin ortalamaları.

Dosya Tamiri için Geçen Toplam Süre



Şekil: Dosya tamiri için kullanılan mesaj sayısı.

Kümeleme tabanlı olmayan dağıtık depolama sistemlerinde 100 MB dosya tamiri için kullanılan mesajların ortalamaları



Şekil: 100 MB dosya tamiri için kullanılan mesajların ortalamaları.

Sonuç

Sistemlerde düğüm tamiri işlemi için kullanılan toplam sürelerin ortalamalarına bakıldığında, kümeleme tabanlı dağıtık depolama sisteminin kümeleme tabanlı olmayan dağıtık depolama sistemlerine göre sırasıyla %10, %6 ve %40 daha az süre kullanımı gerçekleştirdiği görülmektedir.

Sonuç Dvm

Tüm sistemlerde kullanılan ortalama mesaj sayılarına bakıldığında ise kümeleme tabanlı dağıtık depolama sisteminde bir dosya tamiri için ortalama kullanılan mesaj sayısında sadece HSRC kullanılan sisteme göre %9, sadece EMBR kodları kullanılan sisteme göre %25 ve sadece HMBR kodları kullanılan sisteme göre %44 düşüş gözlemlenmiştir.

Kümeleme tabanlı olan ve olmayan dağıtık depolama sistemlerinin maliyet analizi

Tablo: Maliyet faktörlerinin aldığı değerler

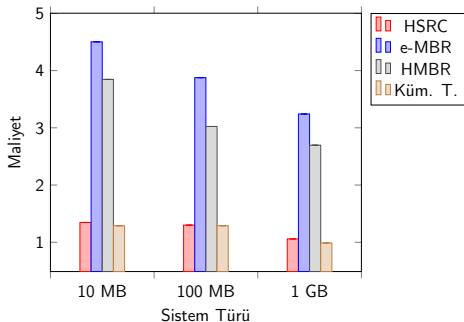
| Parametre | Değeri | Parametre | Değeri |
|---------------|--------|----------------|--------|
| t_1 | 1 | $t_{g1,1}$ | 1.82 |
| t_2 | 1.5 | $t_{g1,2}$ | 0.5 |
| t_3 | 1 | $t_{g2,1}$ | 1.42 |
| $t_{3,1}$ | 1.6 | $t_{g2,2}$ | 0.9 |
| $t_{3,2}$ | 0.4 | $t_{g3,1}$ | 1.9 |
| t_{ort} | 1.16 | $t_{g3,2}$ | 0.42 |
| β_1 | 1 | $\beta_{g1,1}$ | 1 |
| β_2 | 0.9 | $\beta_{g1,2}$ | 0.6 |
| β_3 | 0.5 | $\beta_{g2,1}$ | 1.1 |
| $\beta_{3,1}$ | 0.7 | $\beta_{g2,2}$ | 0.5 |
| $\beta_{3,2}$ | 0.3 | $\beta_{g3,1}$ | 0.9 |

Kümeleme tabanlı olan ve olmayan dağıtık depolama sistemlerinin maliyet analizi

Tablo: Maliyet faktörlerinin aldığı değerler (dvm)

| Parametre | Değeri | Parametre | Değeri |
|-----------------|--------|----------------|--------|
| β_{ort} | 0.8 | $\beta_{g3,2}$ | 0.7 |
| t_{cpu1} | 1.4 | $t_{gcpu1,1}$ | 0.4 |
| t_{cpu2} | 0.3 | $t_{gcpu1,2}$ | 1.42 |
| t_{cpu3} | 0.75 | $t_{gcpu2,1}$ | 0.7 |
| $t_{cpu_{ort}}$ | 0.81 | $t_{gcpu2,2}$ | 0.92 |
| $t_{cpu3,1}$ | 0.2 | $t_{gcpu3,1}$ | 0.5 |
| $t_{cpu3,2}$ | 1.1 | $t_{gcpu3,2}$ | 1.12 |

Dosya tamiri için sistemlerin maliyetleri



Şekil: 10 MB, 100 MB ve 1 GB dosya tamiri için sistemlerin maliyetleri.

Düğüm tamiri başarımlı olasılığına ile göreceli maliyet

$$c_{sistem'_{NR}}(x) = \delta c_{\widehat{sistem}_{NR}}(x) + \gamma(1 - p_{nr_{sistem}}) \quad (2)$$

$$c_{HSRC'_{NR}}(100 \text{ MB}) = \delta c_{\widehat{HSRC}_{NR}}(100 \text{ MB}) + \gamma(1 - p_{nr_{HSRC}}) \quad (3)$$

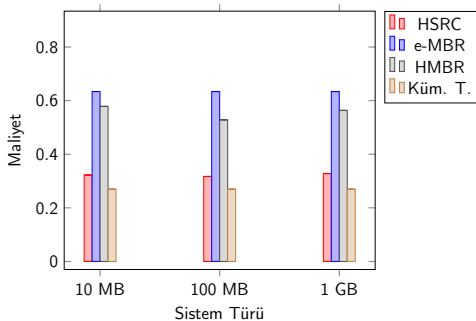
$$c_{EMBR'_{NR}}(100 \text{ MB}) = \delta c_{\widehat{EMBR}_{NR}}(100 \text{ MB}) + \gamma(1 - p_{nr_{EMBR}}) \quad (4)$$

$$c_{HMBR'_{NR}}(100 \text{ MB}) = \delta c_{\widehat{HMBR}_{NR}}(100 \text{ MB}) + \gamma(1 - p_{nr_{HMBR}}) \quad (5)$$

$$c_{Kum'_{NR}}(100 \text{ MB}) = \delta c_{\widehat{Kum}_{NR}}(100 \text{ MB}) + \gamma(1 - p_{nr_{Kum}}) \quad (6)$$

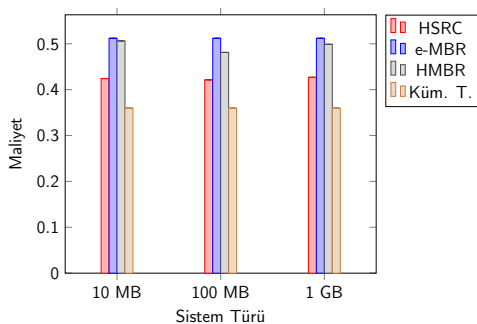
Düğüm tamirinde göreceli maliyetler I

Sistemlerde bulunan düğümlerin çalışır durumda olma olasılıklarının 0.5'den küçük ve eşit olduğu durum için sistemlerin düğüm tamiri işlemindeki göreceli maliyetleri, $\delta = 0.4$ ve $\gamma = 0.6$ değerleri kullanıldığında aşağıdaki gibi çıkmıştır.



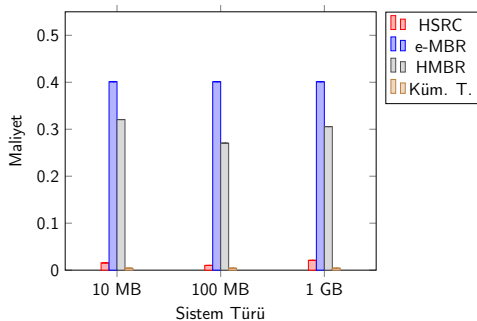
Şekil: 10 MB, 100 MB ve 1 GB dosya tamiri için sistemlerin göreceli maliyetleri ($\delta = 0.4$, $\gamma = 0.6$ ve $p_{saglam} \leq 0.5$ için).

Düğüm tamirinde göreceli maliyetler II








Şekil: 10 MB, 100 MB ve 1 GB dosya tamiri için sistemlerin göreceli maliyetleri ($\delta = 0.2$, $\gamma = 0.8$ ve $p_{saglam} \leq 0.5$ için).

Düğüm tamirinde göreceli maliyetler III $p_{saglam} > 0.5$ için




Şekil: 10 MB, 100 MB ve 1 GB dosya tamiri için sistemlerin göreceli maliyetleri ($\delta = 0.4$, $\gamma = 0.6$ ve $p_{saglam} > 0.5$ için).


TEŞEKKÜRLER


-  [Online]. Facebook's erasure coded hadoop distributed file system (HDFS-RAID). Available: <https://github.com/facebook/hadoop-20>. accessed Feb. 3, 2017
-  Huang, C., Simitci, H., XU, Y., Ogus, A., Calder, B., Gopalan, P., Li, J., and Yekhanin S., "Erasure Coding in Windows Azure Storage", In Proceedings of the 2012 USENIX Conference on Annual Technical Conference , USENIX ATC'12, USENIX Association, pp. 15–26, 2012.
-  Dimakis, A. G., Godfrey, P. B., Wu, Y. and Wainwright, M. J. and Ramchandran, K., "Network Coding for Distributed Storage Systems", Information Theory, IEEE Transactions on, 56(9): 4539-4551, 2010.
-  Reed I. S., Solomon G., "Polynomial codes over certain finite fields", Journal of the Society for Industrial and Applied Mathematics, 8(2):300 -304, 1960.
-  Zhu, Y. and Lee, P. P. C., Xiang, L., Xu, Y. and Lingling G., "A Cost-based Heterogeneous Recovery Scheme for Distributed Storage Systems with RAID-6 Codes", Proceedings of the 42nd Annual

IEEE/IFIP International Conference on Dependable Systems and Networks, pp. 1-12,2012.

 [Online]. The network simulator ns-3. Available: <http://www.nsnam.org/>. accessed Feb. 3, 2017

 Oggier, F. and Datta, A., "Self-repairing Homomorphic Codes for Distributed Storage Systems", INFOCOM, Proceedings IEEE, pp. 1215–1223, 2011.

 Rashmi, K. V., Shah, N. B. and Kumar, P. V., "Optimal Exact-Regenerating Codes for Distributed Storage at the MSR and MBR Points via a Product-Matrix Construction", Information Theory, IEEE Transactions on, 57(8): 5227-5239,2011.

 Haytaoglu, E. and Dalkilic, M. E., "A new hybrid coding scheme: homomorphic minimum bandwidth repairing codes", Computing, 2017, doi:10.1007/s00607-017-0542-0

 Waxman, B. M., "Routing of Multipoint Connections", IEEE Journal on Selected Areas in Communications 6(9): 1617–1622, 1988.



Medina, A., Matta, I. and Byers, J., "On the Origin of Power Laws in Internet Topologies", SIGCOMM Comput. Commun. Rev., 30(2):18–28, 2000.