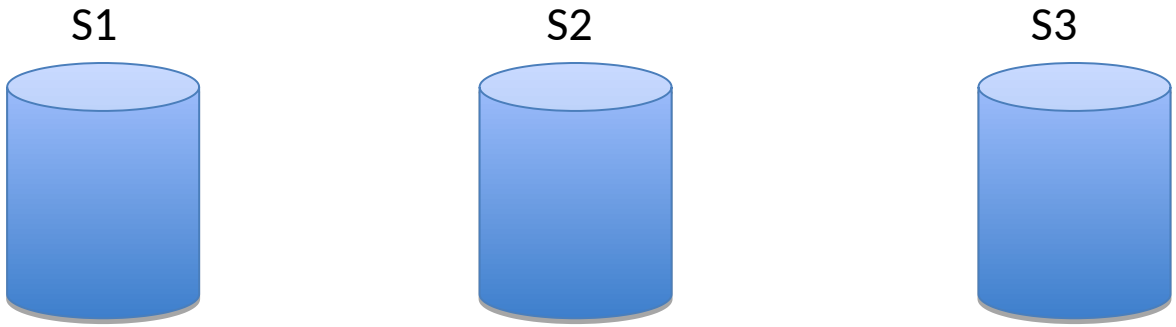
# Weighted Highest Random Weight (HRW) and its Applications

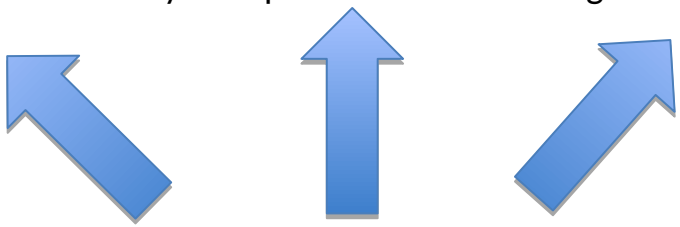
Satya R Mohanty
Mankamana Misra
Ali Sajassi
Acee Lindem

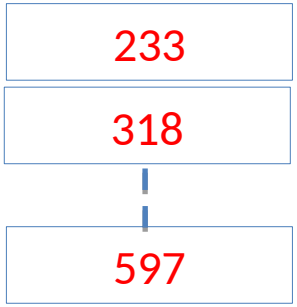
IETF 104 Prague

# The Load Balancing problem:




Given a set of servers and objects, devise a mapping scheme such that load is evenly spread and minimally disruptive in case of reassignments






Objects with object-id

Modulo-N Assignment: S = key%N

When one server goes down or comes up, a lot of reassignments!

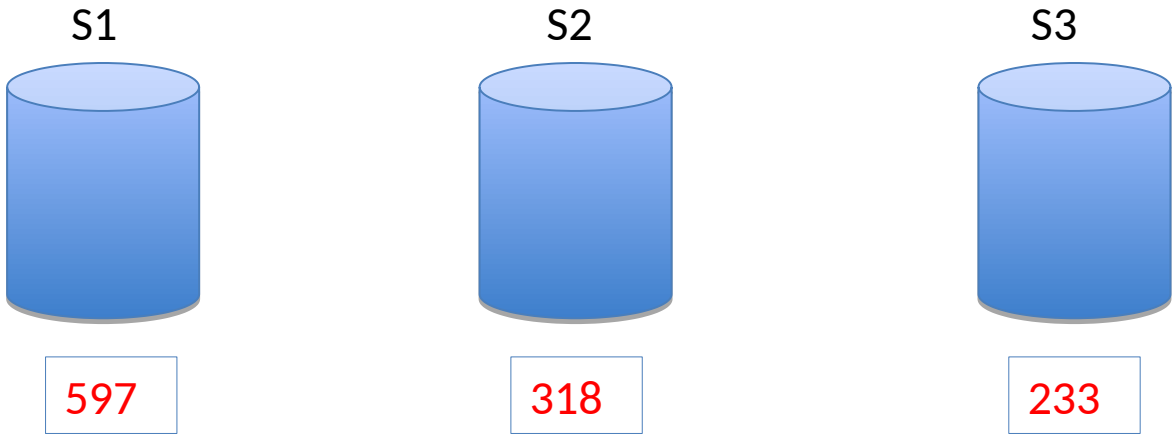
# Highest Random Weight (HRW)

- When the hash function is uniform (any good hash function should satisfy this) and as the load (number of objects) increases, It is proved [†] that
  - **The load is evenly balanced across the servers using HRW**
  - **Minimal disruption property: a server going up or down results in a minimal reassignment of impacted objects**

[†] Using name-based mappings to increase hit rates: Thaler et. al. IEEE Transactions on Networking, 1999

Hash(Srvr-id * Key) = Score

Highest score wins




H(S1* 233 ) = 457

H(S2* 233 ) = 317

H(S3* 233 ) = 512

H(S1* 318 ) = 471

H(S2* 318 ) = 513

H(S3* 318 ) = 172

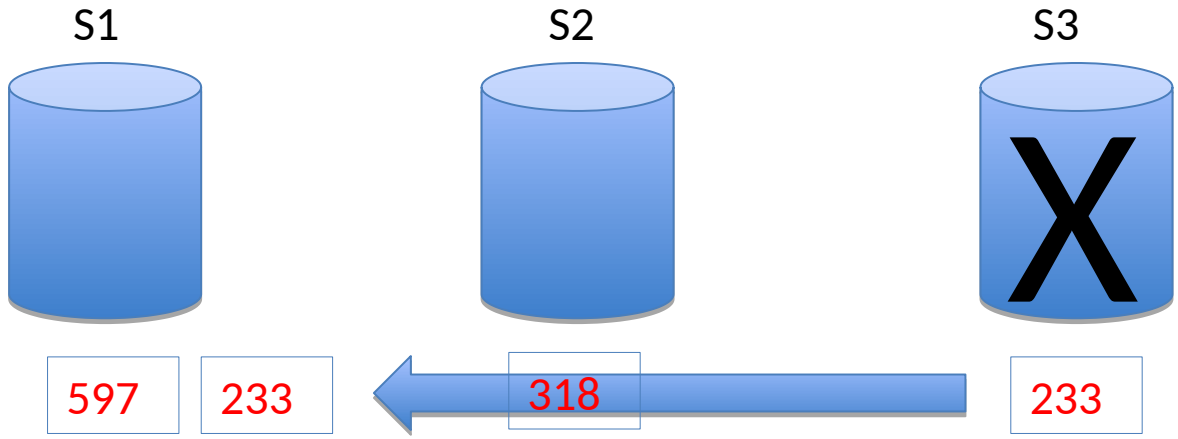
H(S1* 597 ) = 919

H(S2* 597 ) = 200

H(S3* 597 ) = 706

Hash(Srvr-id * Key) = Score
Highest score wins

S3 goes down!



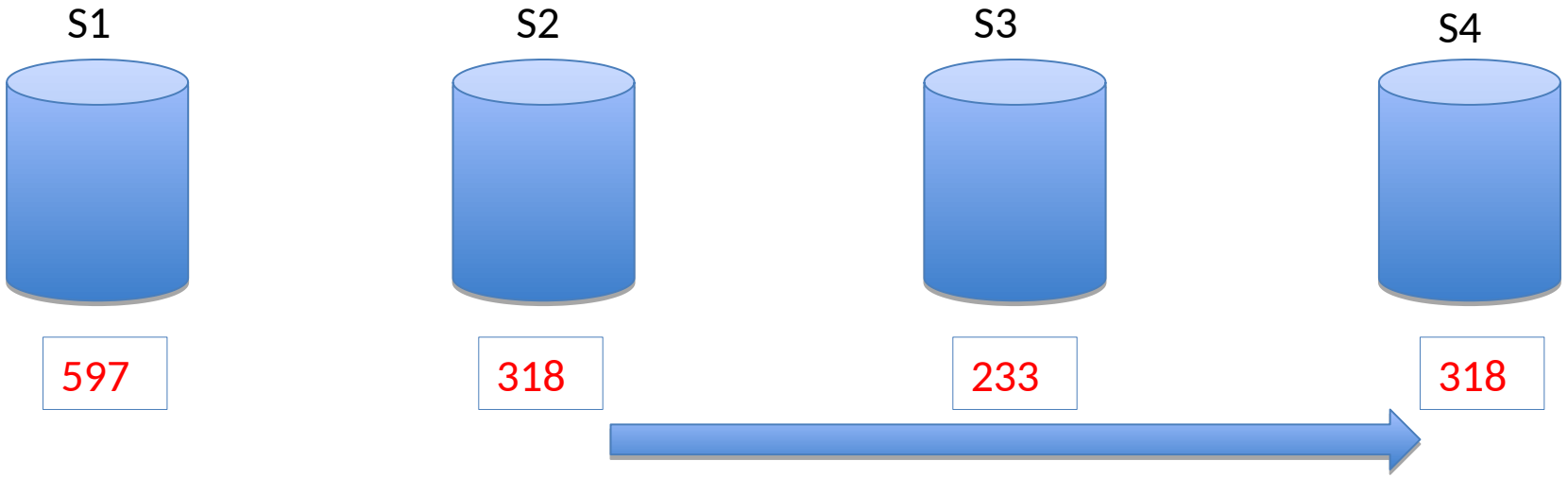

H(S1* 233 ) = 457
H(S2* 233 ) = 317
H(S3* 233 ) = 512

H(S1* 318 ) = 471
H(S2* 318 ) = 513
H(S3* 318 ) = 172

H(S1* 597 ) = 919
H(S2* 597 ) = 200
H(S3* 597 ) = 706

Hash(Srvr-id * Key) = Score
Highest score wins

S4 comes up!




H(S1* 233 ) = 457
H(S2* 233 ) = 317
H(S3* 233 ) = 512
H(S4* 233 ) = 236

H(S1* 318 ) = 471
H(S2* 318 ) = 513
H(S3* 318 ) = 172
H(S4* 318 ) = 672

H(S1* 597 ) = 919
H(S2* 597 ) = 200
H(S3* 597 ) = 706
H(S4* 597 ) = 234

# Weighted HRW

- ***What happens*** when the Servers are ***not*** of equal capacities or weights?

- One approach: Take the weighted score:

  $f_i$ * Hash(Srvr-id * Key); where fi is wi/sum(wj), j=1,..,

- Microsoft: Cache Array Routing Protocol (CARP)
  - https://tools.ietf.org/html/draft-vinod-carp-v1-03

$f_i$ * Hash(Srvr-id * Key) = Score

Highest score wins


H(S1* 233 ) * 0.5 = 457 * 0.5

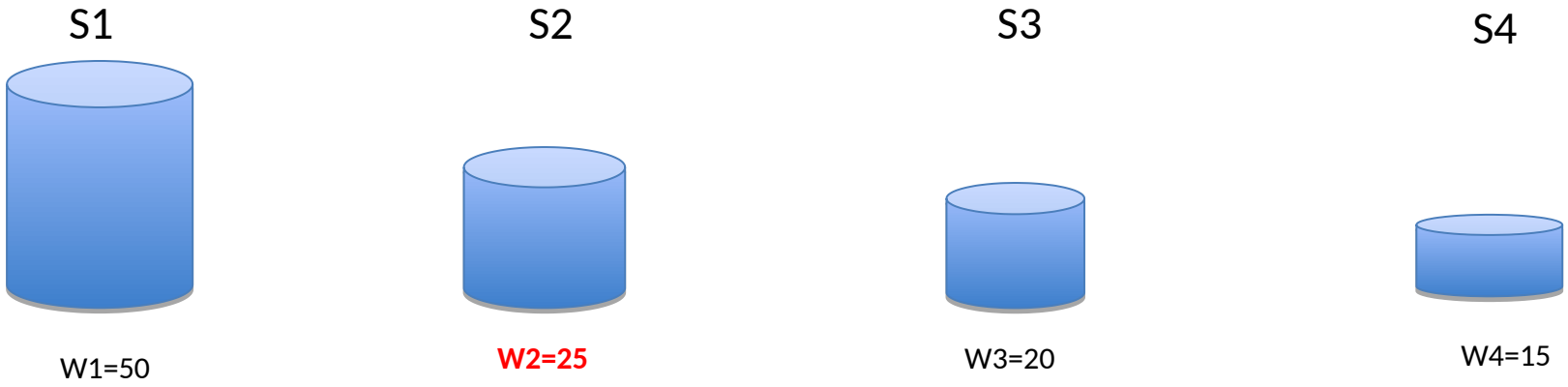
H(S2* 233 ) * 0.15 = 317 * 0.15

H(S3* 233 ) * 0.2 = 512 * 0.2

H(S4* 233 ) * 0.15 = 236 * 0.15

$f_i$ * Hash(Srvr-id * Key) = Score

Highest score wins




H(S1* 233 ) * 0.456

H(S2* 233 ) * 0.227

H(S3* 233 ) * 0.182

H(S4* 233 ) * 0.136

- The weight of S2 only changed.
- But load factors changed everywhere!
- This will result in **re-computation** and **re-assignment** in a potentially disruptive manner
- **Does not** satisfy HRW desirable properties
- CARP does not have this property

# Weighted HRW

- Taking the weighted score is not efficient

  $f_i$ * Hash(Srvr-id * Key); where fi is wi/sum(wj), j=1,.., N

- Take the score as: $-w_i/\ln(\text{Hash}(\text{Srvr-id} * \text{Key})/\text{Hmax})$

  Jason Resch. "New Hashing Algorithms for Data Storage [Storage Developer Conference, Santa Clara, 2015]

- **Only need to re-compute** the score for the server whose weight changed. Other's scores **do not** change

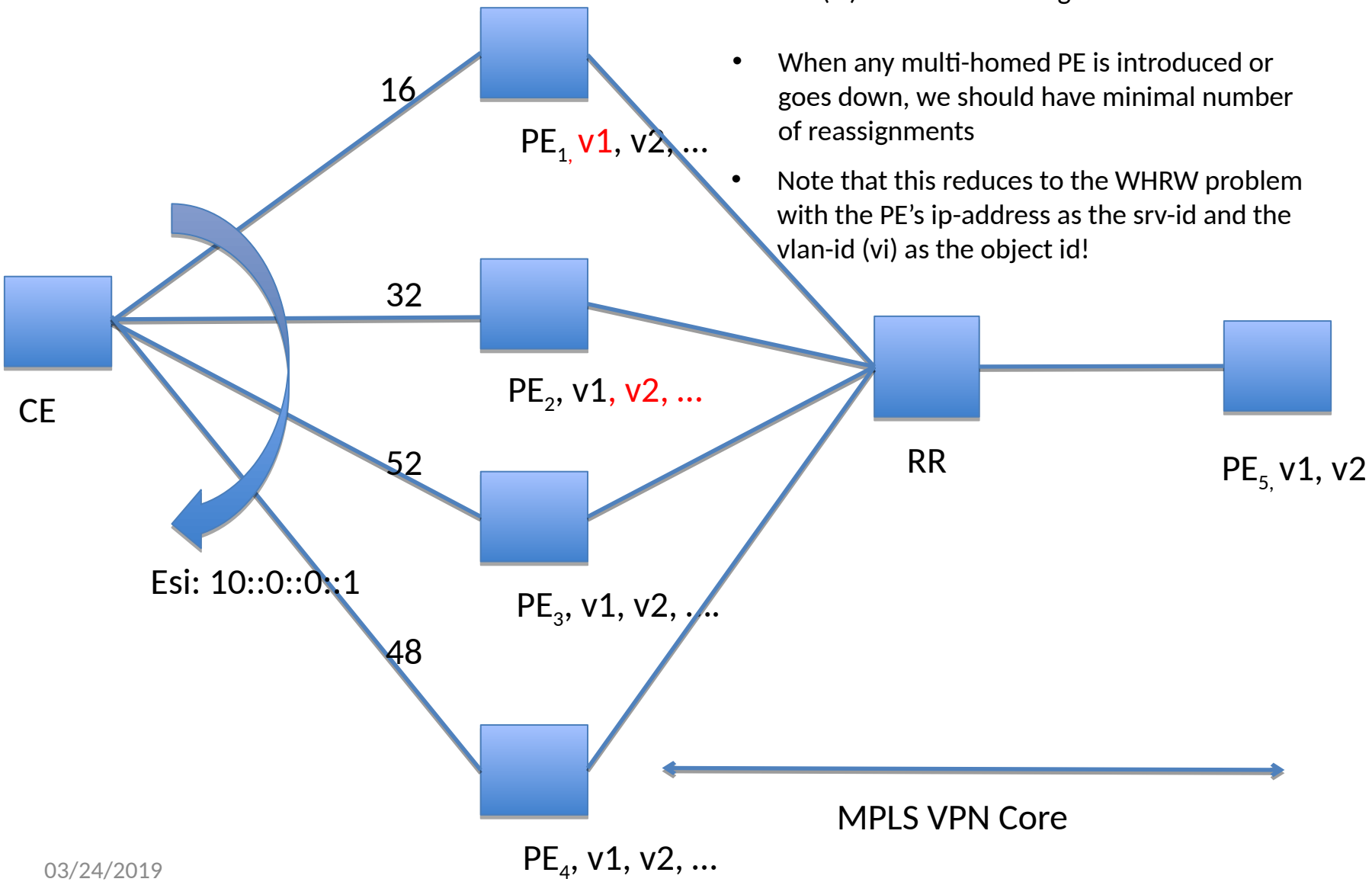
- Obeys the **minimal disruption** properties of the HRW
  - When a server is added/removed or changed, only the scores for that node change.
  - It may win some keys (if score increases)
  - It may lose some keys (if score decreases)
  - And it does so with **minimal disruption**

# Applications

- EVPN DF
  - Different link Bandwidth on lag
    https://tools.ietf.org/html/draft-ietf-bess-evpn-unequal-lb-00
- Resilient Hashing
  - LAG
  - Unequal cost multipath
- Multicast
  - Unequal B/W towards receivers
  - DR elections when access bandwidth is different for attach points in the last hop network

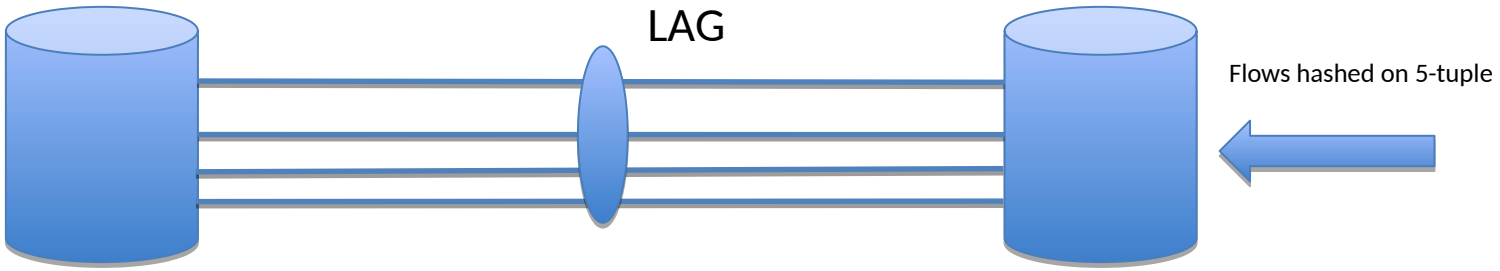
# EVPN DF Election in A/A Deployments with DMZ link bandwidth)

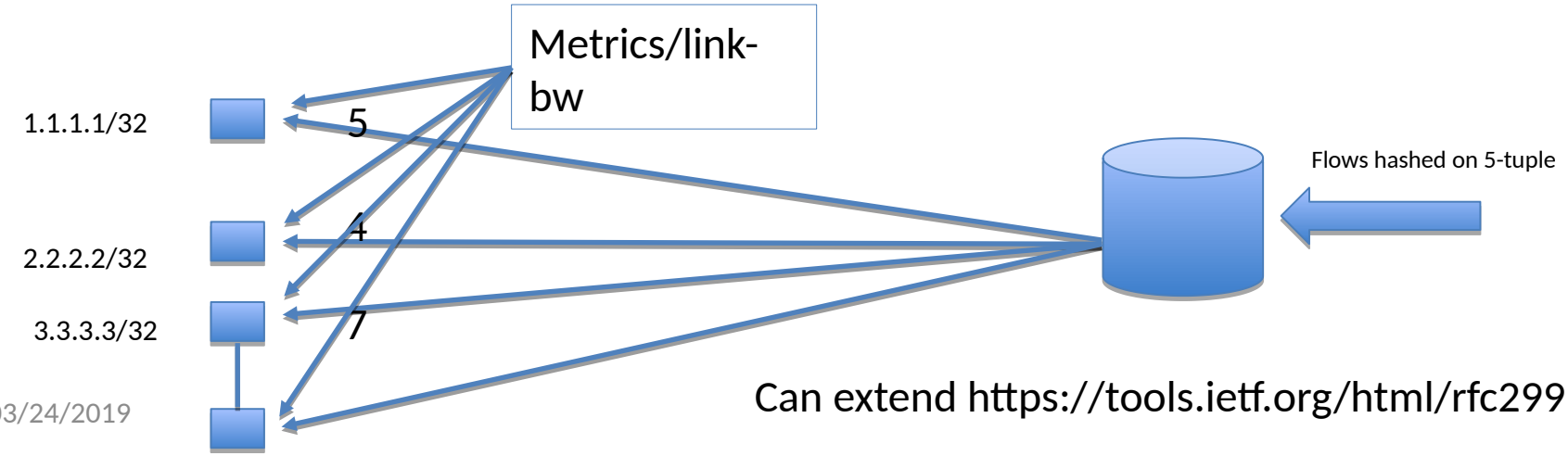
- Goal is to have different DFs (PEs) for different EVI (vi) for load balancing
- When any multi-homed PE is introduced or goes down, we should have minimal number of reassignments
- Note that this reduces to the WHRW problem with the PE's ip-address as the srv-id and the vlan-id (vi) as the object id!



# Resilient Hashing

- Minimize flow remapping in Trunk/ECMP Groups in FIB
  - Many vendors…..
  - But nothing on UCMP?




Can extend https://tools.ietf.org/html/rfc2991

Thanks!!!