

XDP + NetEm = XnetEm



Stephen Hemminger
@networkplumber

NETDEV 2.2

8th Nov 2017

Introduction

- Motivation
- Implementation
- Usage
- Plans
- Problems
- Q&A



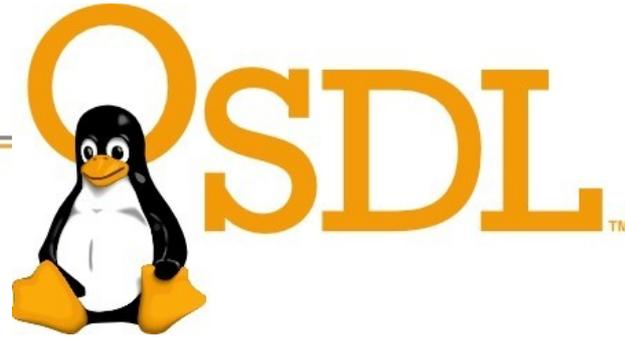


**START
SMALL**

BACK TO THE FUTURE



Spigot



Netem emulating real networks

<http://developer.osdl.org/shemminger/netem>

Stephen Hemminger

shemminger@osdl.org

[Linux.conf.au](http://linux.conf.au)

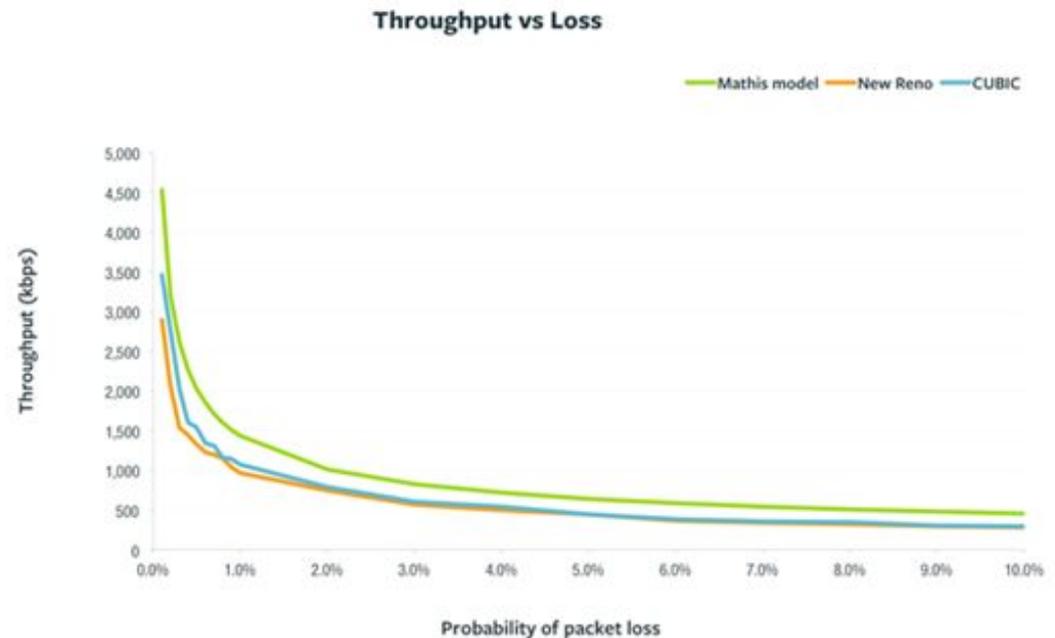
21 April 2005

TCP throughput

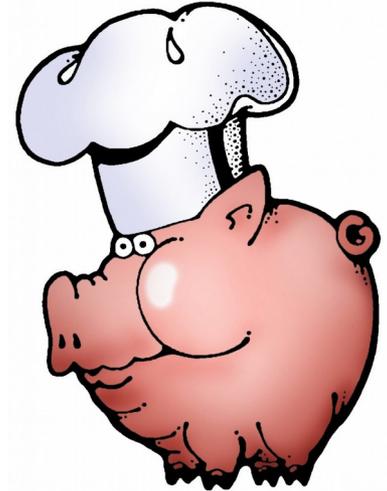
- following Mathis, et al (1997)

$$T = \frac{MSS \times C}{RTT \times \sqrt{p}}$$

- where,
- MSS – maximum segment size
- C – constant – lumpsum for several terms (TCP implementation, ACK strategy, loss mechanism)
- RTT – round-trip time (latency)
- p – packet-loss



Netem use cases







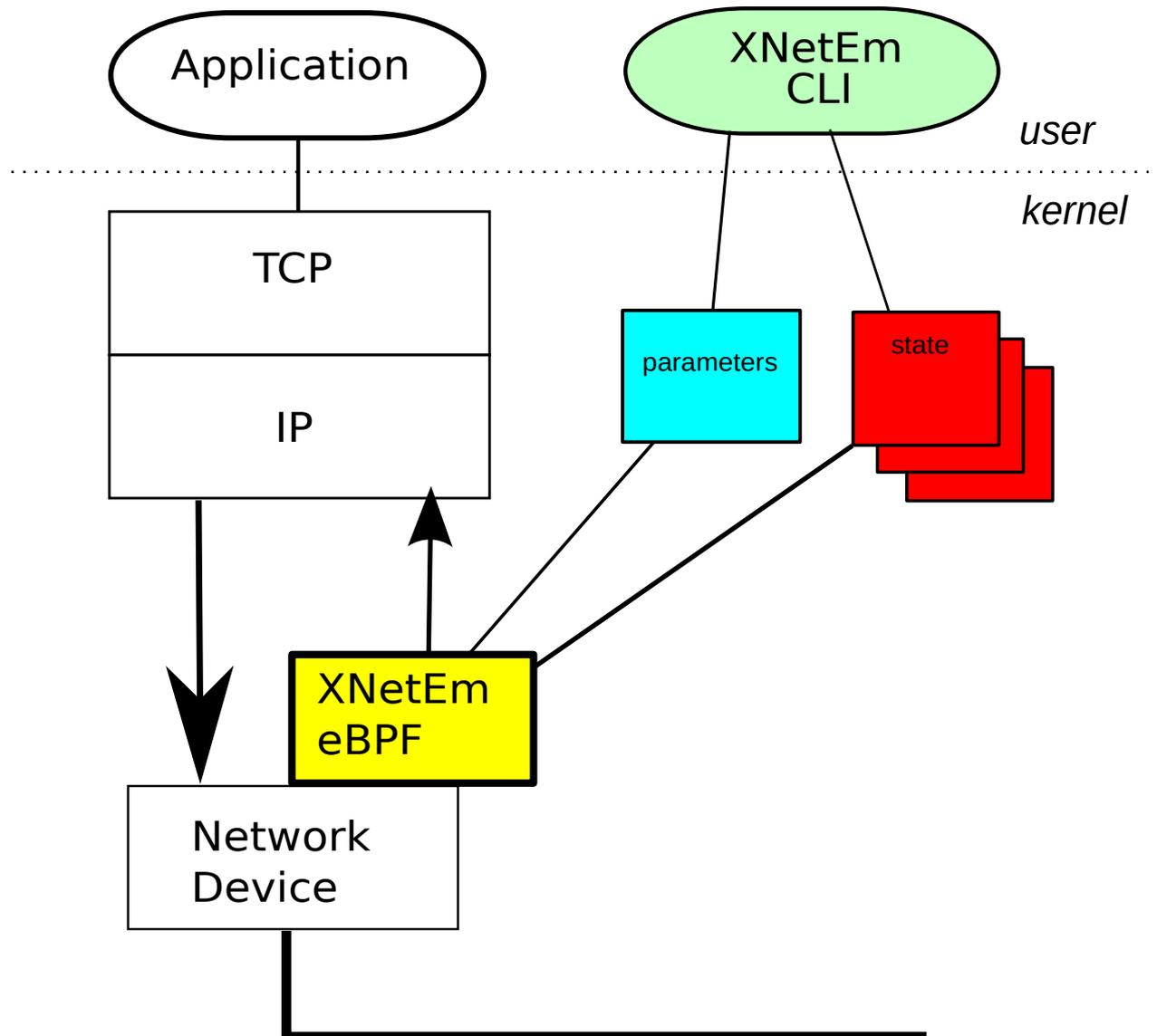




IF TAPE IS PERLED,
TORN OR CUT,
INSPECT CONTENTS
OR PACKAGE FIRST
TO AVOID DAMAGE

IF TAPE IS PERLED,
TORN OR CUT,
INSPECT CONTENTS
OR PACKAGE FIRST
TO AVOID DAMAGE

Architecture



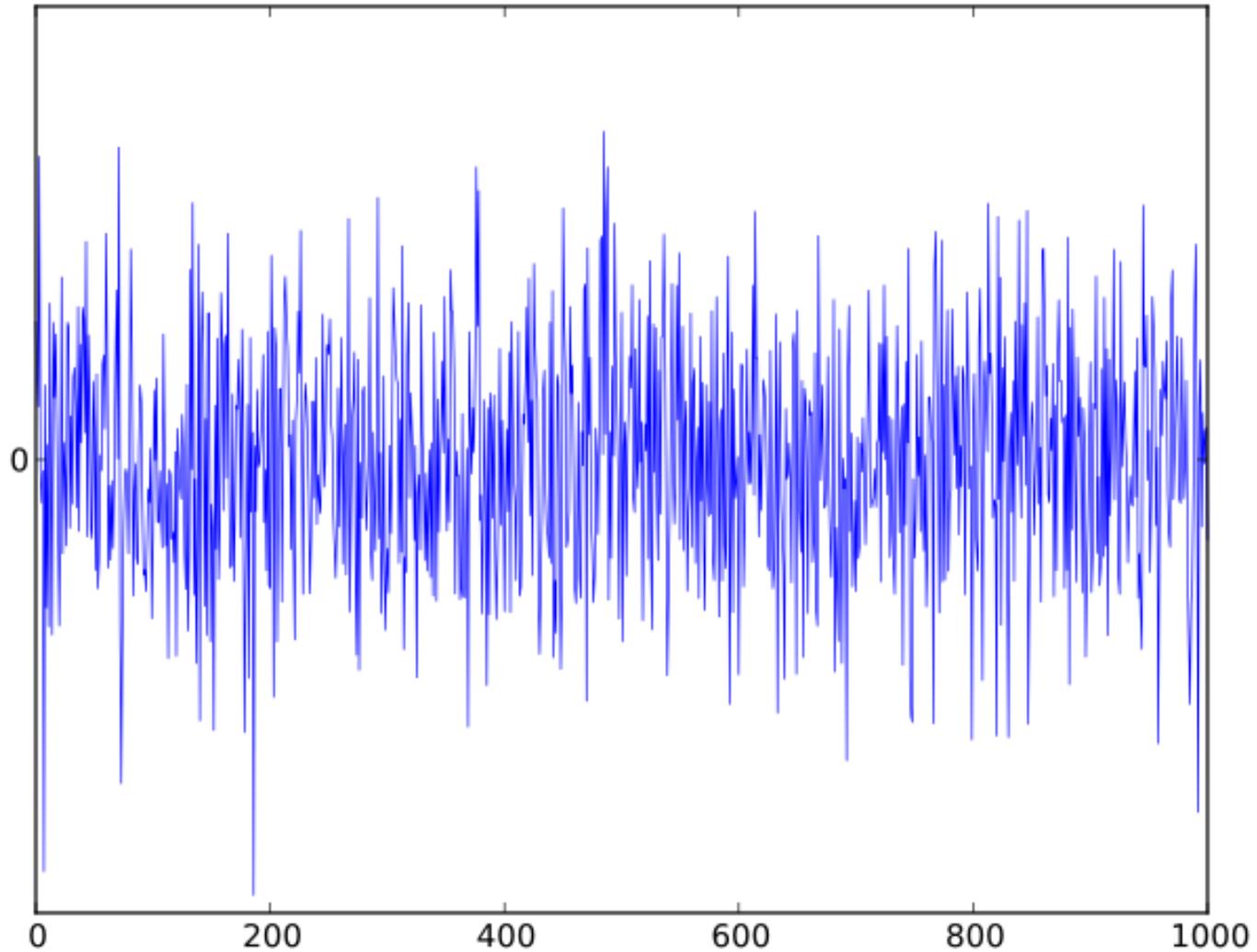
XnetEm CLI

```
# tc qdisc add dev eth0 root netem \  
    loss random 3% .25
```

```
# xnetem eth0 loss random 3% .25
```

Random

Anyone who attempts to generate random numbers by deterministic means is, of course, living in a state of sin. John von Neumann



Scaling Random Numbers



$\text{RANDOM32} = 0.. 4294967295$

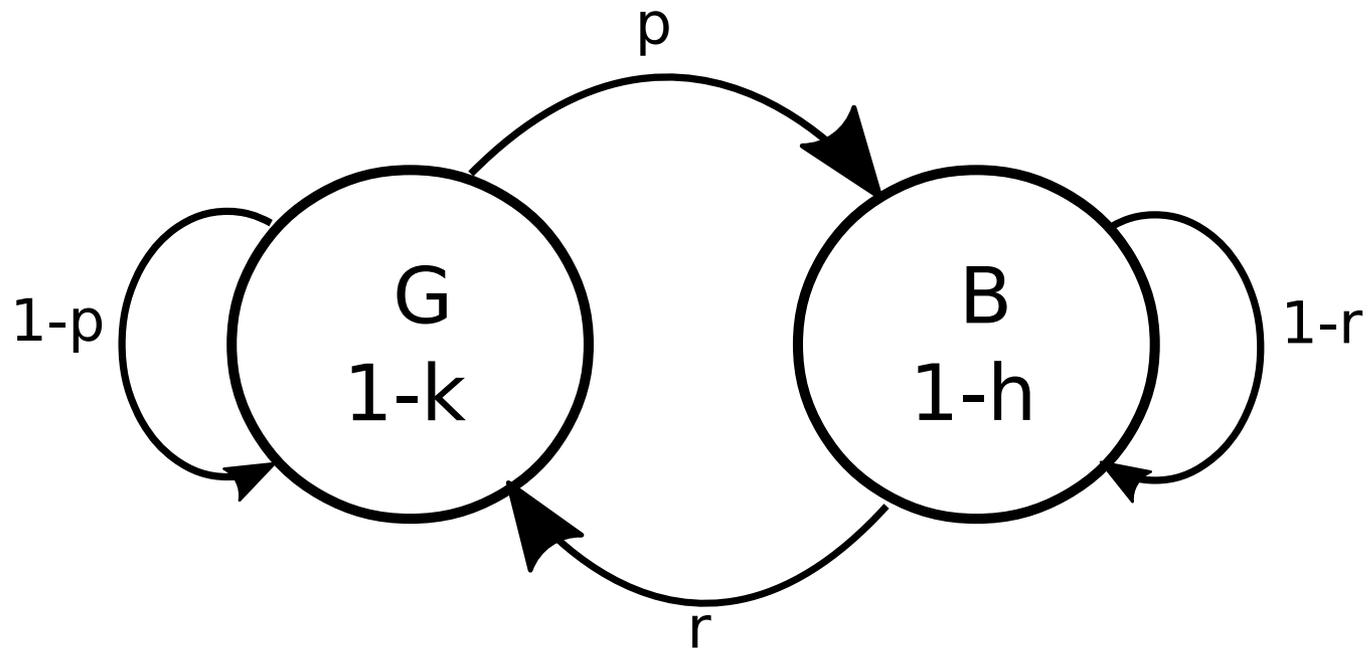
$100\% = 4294967295$

$1\% = 42949672$

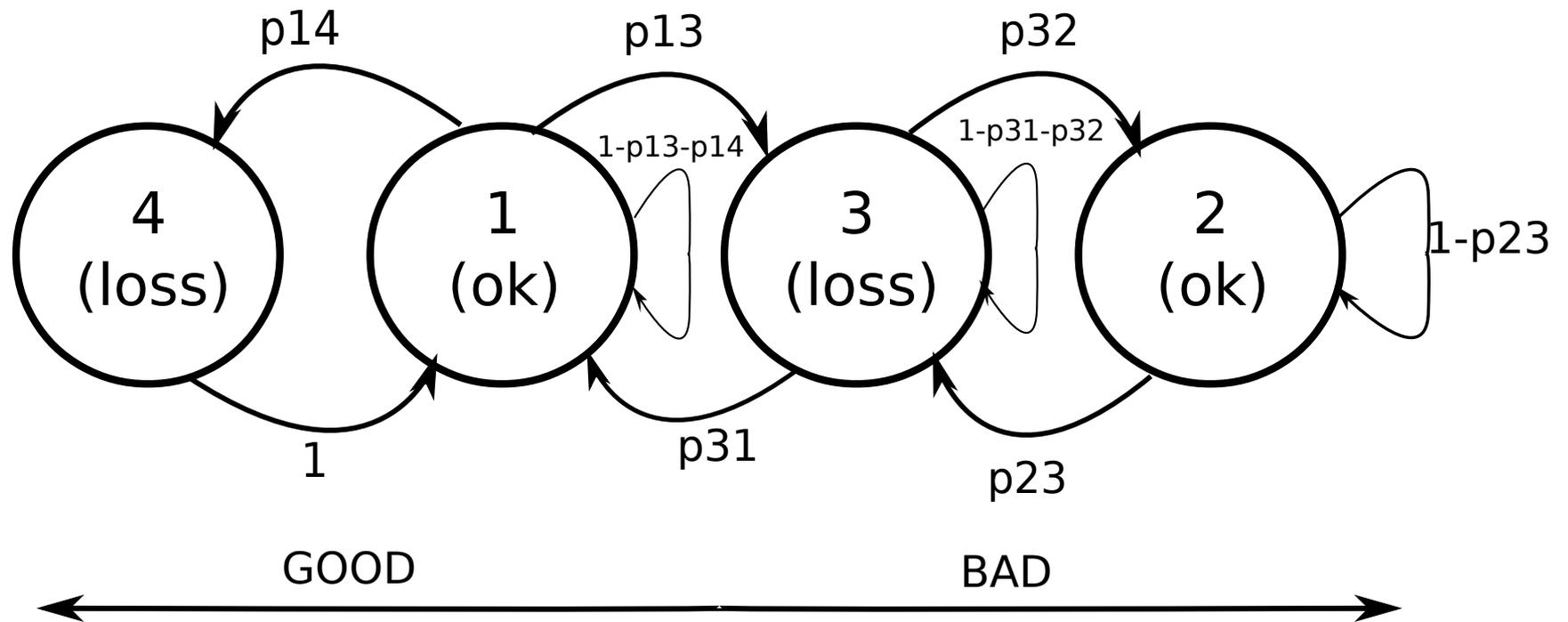
```
int xdp_loss_random_prog(struct xdp_md *ctx)
{
    int key = 0;
    u32 *val = bpf_map_lookup_elem(&options, &key);
    u32 prob = val ? *val : 0;
    int rc;

    if (prob >= bpf_get_prandom_u32())
        rc = XDP_DROP;
    else
        rc = XDP_PASS;
    return rc;
}
```

General Intuitive Model



4 State Loss model





```
int xdp_corrupt_prog(struct xdp_md *ctx)
{
    int key = 0;
    u32 *val = bpf_map_lookup_elem(&options, &key);
    u32 prob = val ? *val : 0;
    int rc;

    if (prob >= bpf_get_prandom_u32()) {
        u32 rnd = bpf_get_prandom_u32();
        u32 len = data_end - data;
        u32 offs = (rnd >> 3) % len;
        u8 mask = 1u << (rnd & 7);

        data[offs] ^= mask;
    }
    return XDP_PASS;
}
```

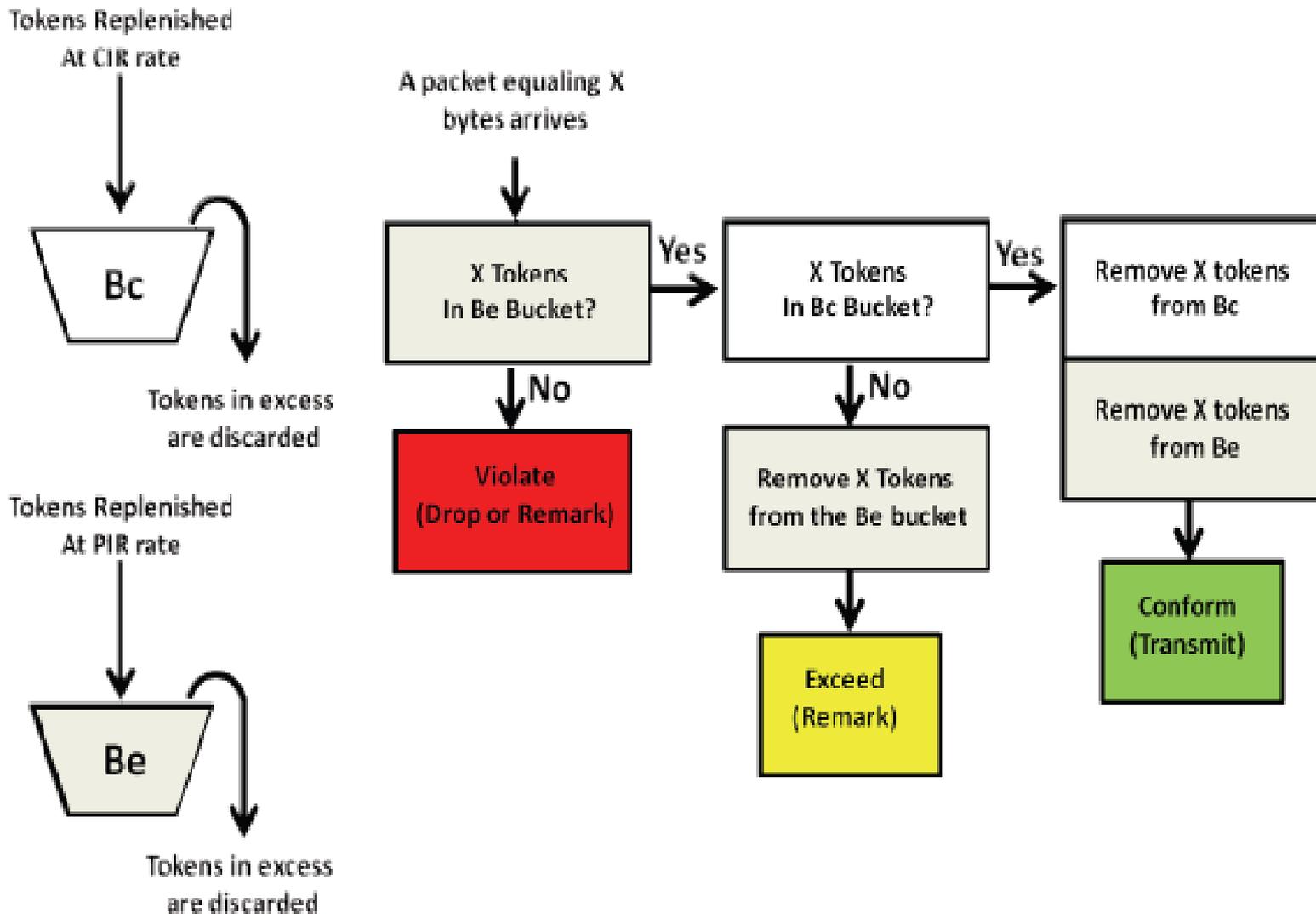
I will not say: "Since we are doing this, how about ..."



XNetEm Report card

	Status	Issues
Loss	POC	
Corruption	POC	
Combined	1.1?	
Policing	1.1?	Two color?
Explicit Congestion Notification	1.1?	
Delay	?	Defer and timers
Reordering	?	Needs delay
Duplication	?	Not work conserving

Two Color Policing



Delay



FUTURE





If we build it, they will come

Thank You

감사합니다

KamSahamnida

김밥 누나 :)
Kimbapnoona.tumblr