

Seawall: Performance Isolation for Cloud Datacenter Networks



Alan Shieh

Cornell University

Srikanth Kandula

Albert Greenberg

Changhoon Kim

Microsoft Research

Cloud datacenters: Benefits and obstacles

- Moving to the cloud has manageability, costs & elasticity benefits
- Selfish tenants can monopolize resources
- Compromised & malicious tenants can degrade system performance
- Problems already occur

Runaway client overloads storage



Spammers on AWS

Security Fix
Brian Krebs on Computer Security

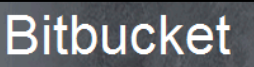
[About This Blog](#) | [Archives](#) | [Security Fix Live: Web Chats](#) | [E-Mail Brian Krebs](#)

Amazon: Hey Spammers, Get Off My Cloud!

I am accustomed to receiving e-mail from **Amazon.com**, as I am a fiercely loyal customer who shops there quite frequently. But it took me by surprise this weekend to discover that mounds of porn spam and junk e-mail laced with computer viruses are actively being blasted at my digital real estate leased to the e-commerce giant.

I wasn't the only one who spotted it. **Websense Security Labs** is [an alert](#) about the spam attacks on Monday, but it didn't name Amazon as the source. The advisory rightly noted that it had discovered "a substantial number of spam messages utilizing a reliable social engineering trick." The junk mail claims to have been sent from Microsoft, and urges the recipient to install an attached security update.

Windows users who fall for the ruse will have their systems infected with a [backdoor Trojan horse](#) program that gives the attackers easy access with which to control the infected machine from afar or upload



Bitbucket
We're here to serve

On our extended downtime, Amazon and what's coming

As many of you are well aware, we've been experiencing some serious downtime the past couple of days. Starting Friday evening, our network storage became virtually unavailable to us, and the site crawled to a halt.

We're hosting everything on Amazon EC2, aka. "the cloud", and we're also using their EBS service for storage of everything from our database, logfiles, and user data (repositories.)

Amazon EBS is a persistent storage solution for EC2, where you get high-speed (and free) connectivity from your instances, while it's also replicated. That gives you a lot for free, since you don't have to worry about hardware failure, and you can create periodic "snapshots" of your volumes easily.

While we were down, it was unknown to us what exactly the problem was, but it was almost certainly a problem with the EBS store. We've been working closely with Amazon the past 24 hours resolving the issue, and this post will outline what exactly went wrong, and what was done to remedy the problem.

Symptoms

Bitbucket DoS attack

Goals

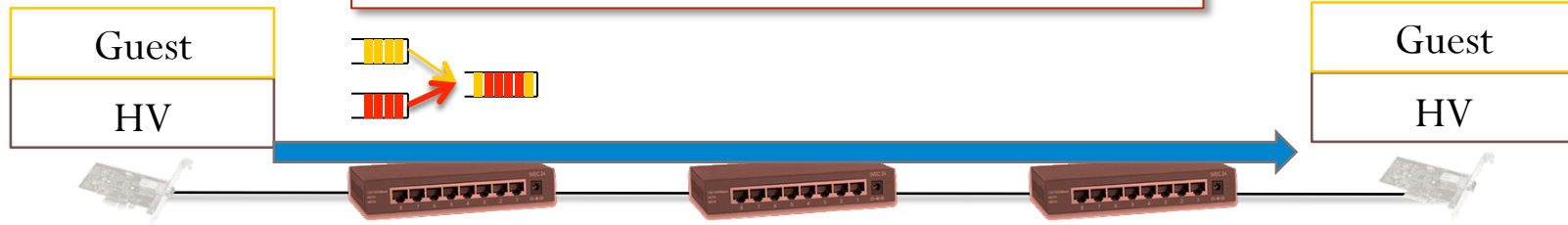
- Isolate tenants to avoid collateral damage
- Control each tenant's share of network
- Utilize all network capacity
- Constraints
 - Cannot trust tenant code
 - Minimize network reconfiguration during VM churn
 - Minimize end host and network cost

Existing mechanisms are insufficient for cloud

Existing mechanisms are insufficient

- In-network queuing and rate limiting

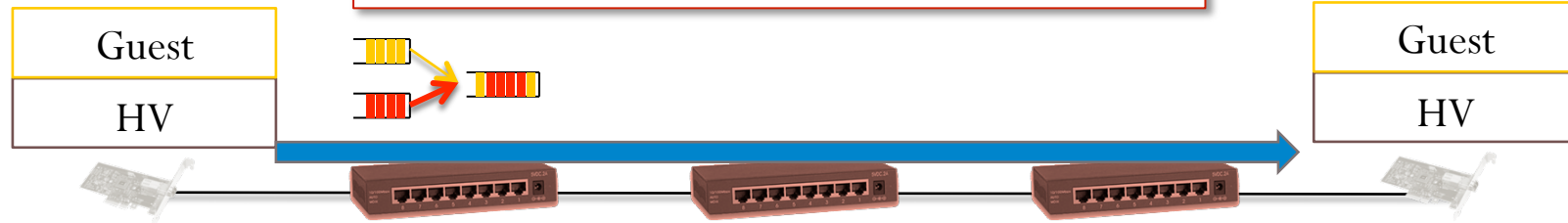
Not scalable. Can underutilize links.



Existing mechanisms are insufficient

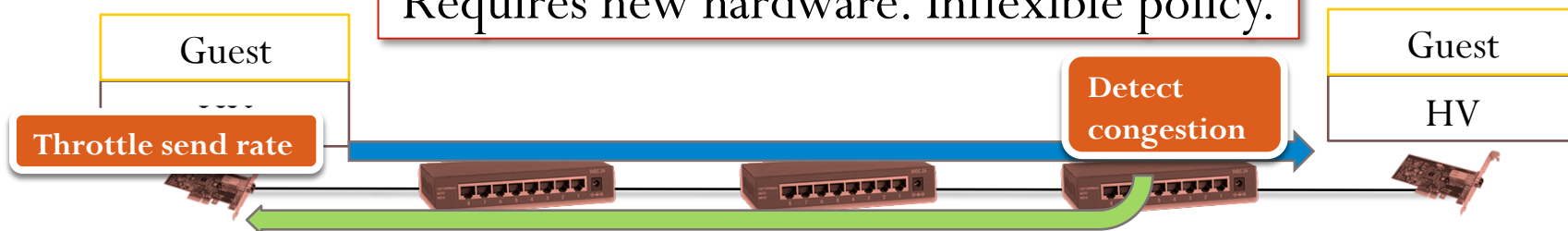
- In-network queuing and rate limiting

Not scalable. Can underutilize links.



- Network-to-source congestion control (Ethernet QCN)

Requires new hardware. Inflexible policy.



Existing mechanisms are insufficient

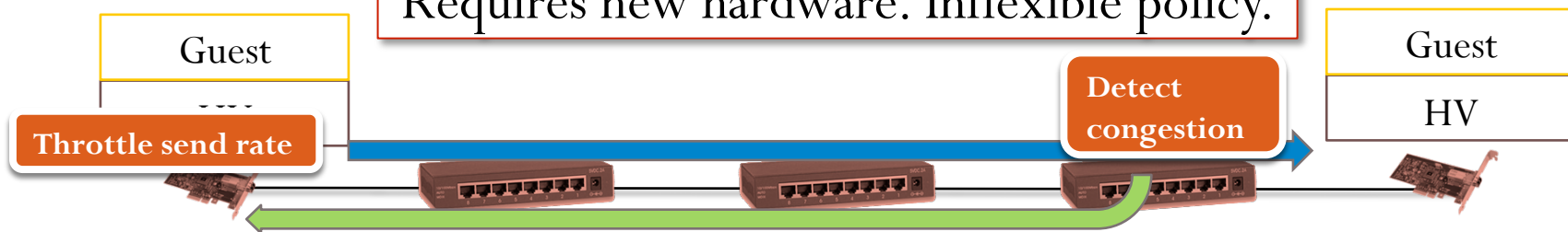
- In-network queuing and rate limiting

Not scalable. Can underutilize links.



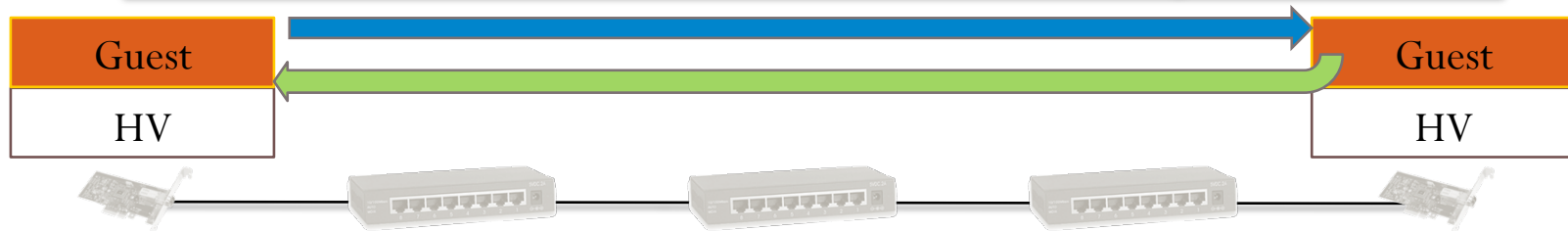
- Network-to-source congestion control (Ethernet QCN)

Requires new hardware. Inflexible policy.

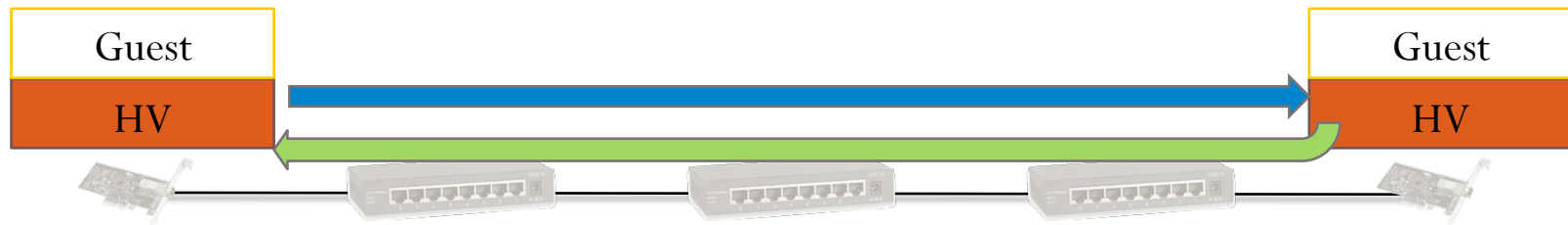


- End-to-end congestion control (TCP)

Poor control over allocation. Guests can change TCP stack.



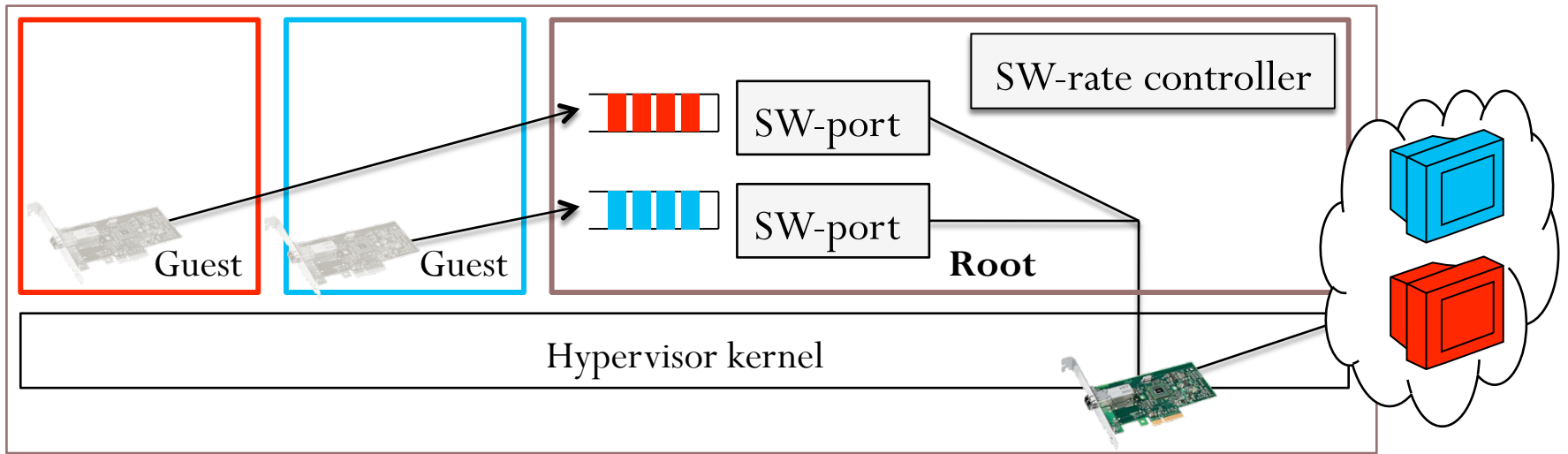
Seawall = Congestion controlled, hypervisor-to-hypervisor tunnels



Benefits

- Scales to # of tenants, flows, and churn
- Don't need to trust tenant
- Works on commodity hardware
- Utilizes network links efficiently
- Achieves good performance
(1 Gb/s line rate & low CPU overhead)

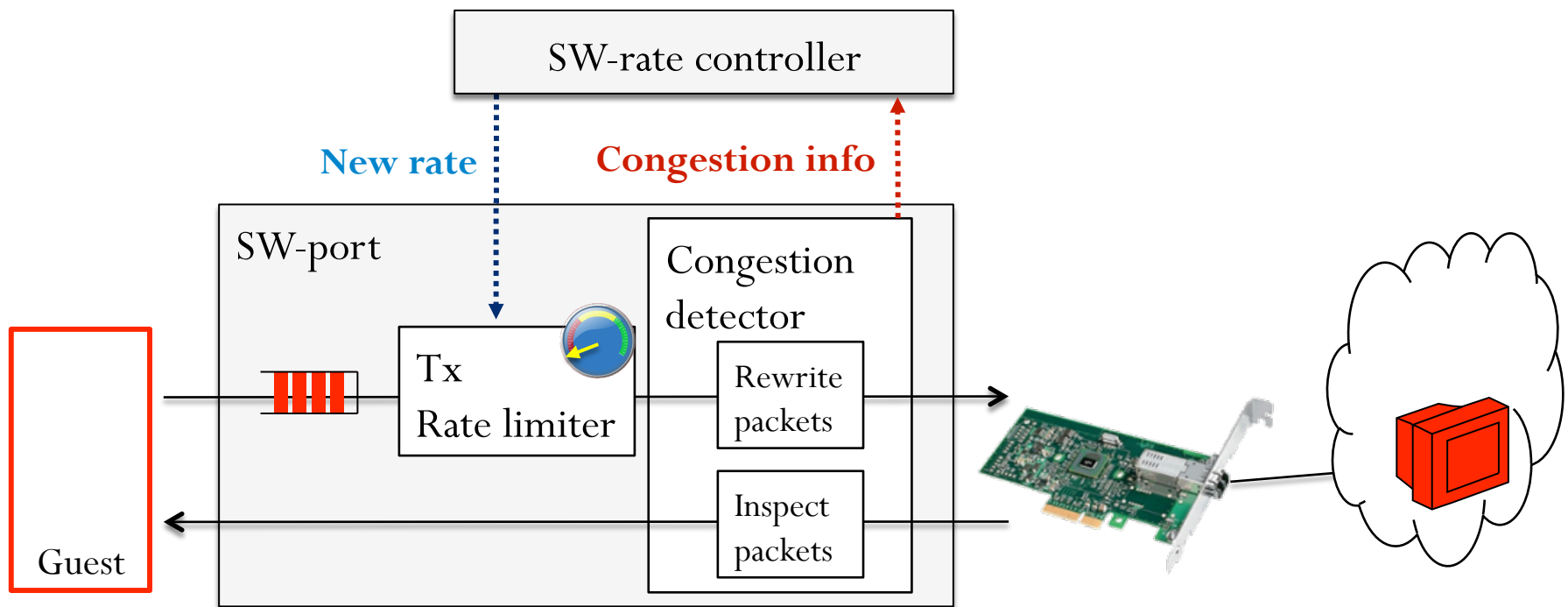
Components of Seawall



- Seawall rate controller allocates network resources for each output flow
 - Goal: achieve utilization and division
- Seawall ports enforce decisions of rate controller
 - Lie on forwarding path
 - One per VM source/destination pair

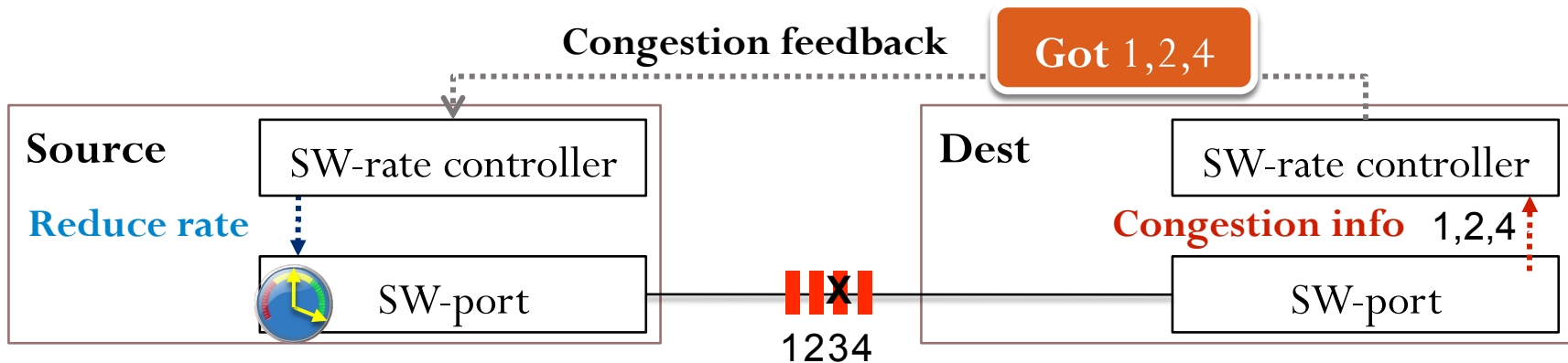
Seawall port

- Rate limit transmit traffic
- Rewrite and monitor traffic to support congestion control
- Exchanges congestion feedback and rate info with controller

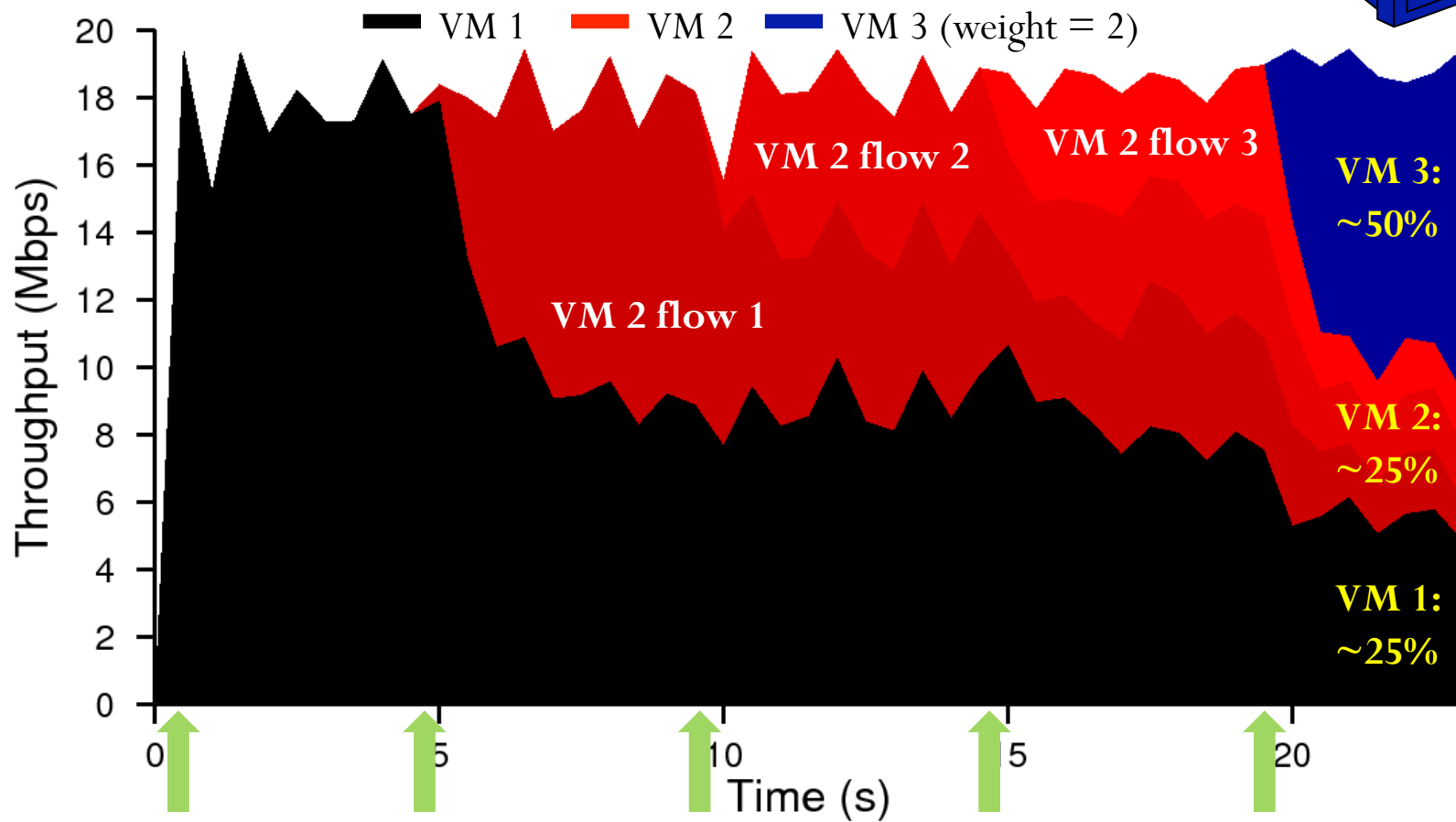
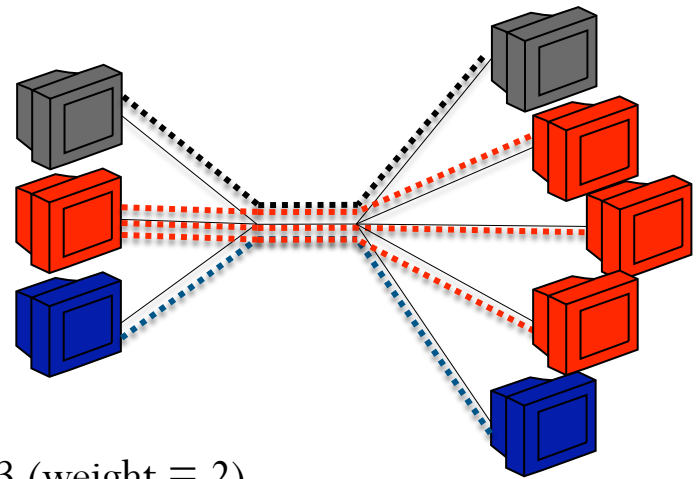


Rate controller: Operation and control loop

- Rate controller adjusts rate limit based on presence and absence of loss

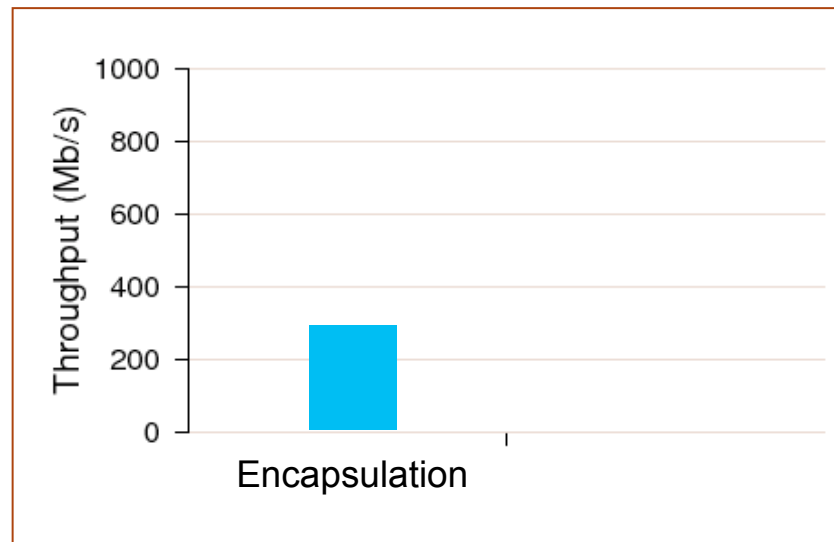


- Algorithm divides network proportional to weights & is max/min fair
 - Efficiency: AIMD with faster increase
 - Traffic-agnostic allocation:
Per-link share is same regardless of # of flows & destinations



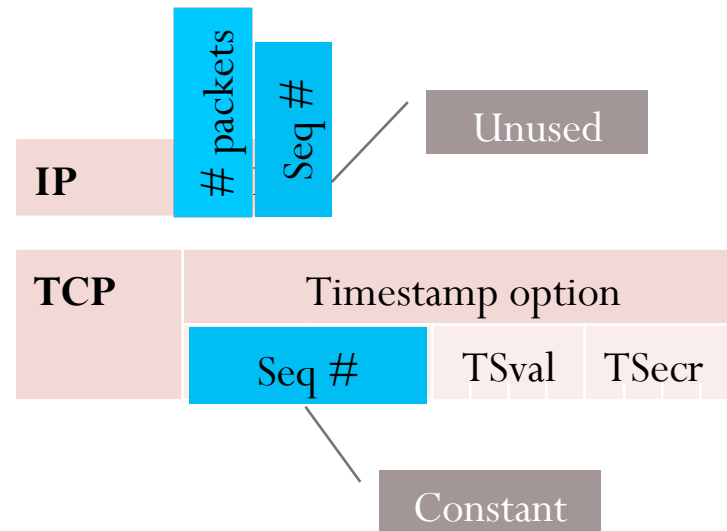
Improving SW-port performance

- How to add congestion control header to packets?
- Naïve approach: Use encapsulation, but poses problems
 - More code in SW-Port
 - Breaks hardware optimizations that depend on header format
 - Packet ACLs: Filter on TCP 5-tuple
 - Segmentation offload: Parse TCP header to split packets
 - Load balancing: Hash on TCP 5-tuple to spray packets (e.g. RSS)

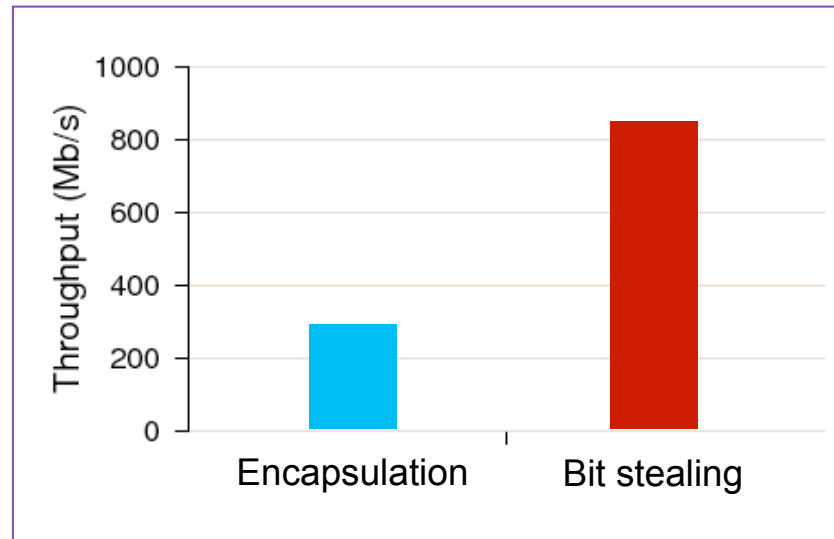


“Bit stealing” solution: Use spare bits from existing headers

- Constraints on header modifications
 - Network can route & process packet
 - Receiver can reconstruct for guest
- Other protocols: might need paravirtualization.



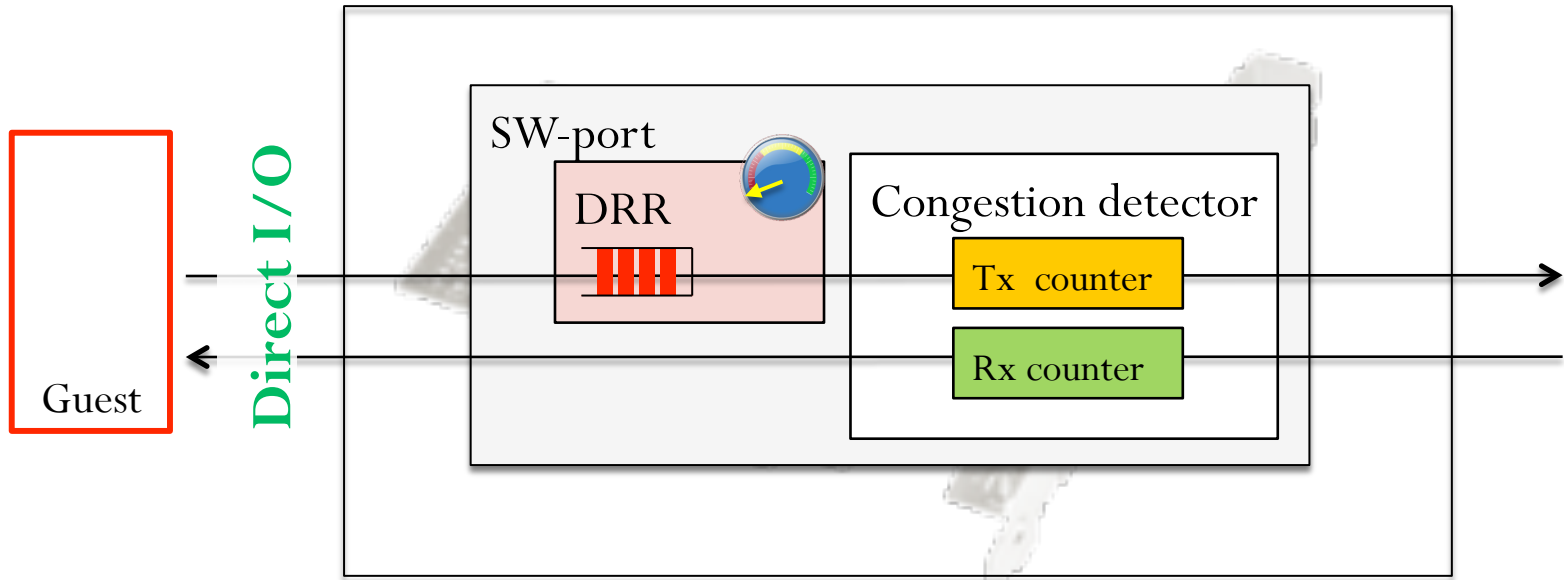
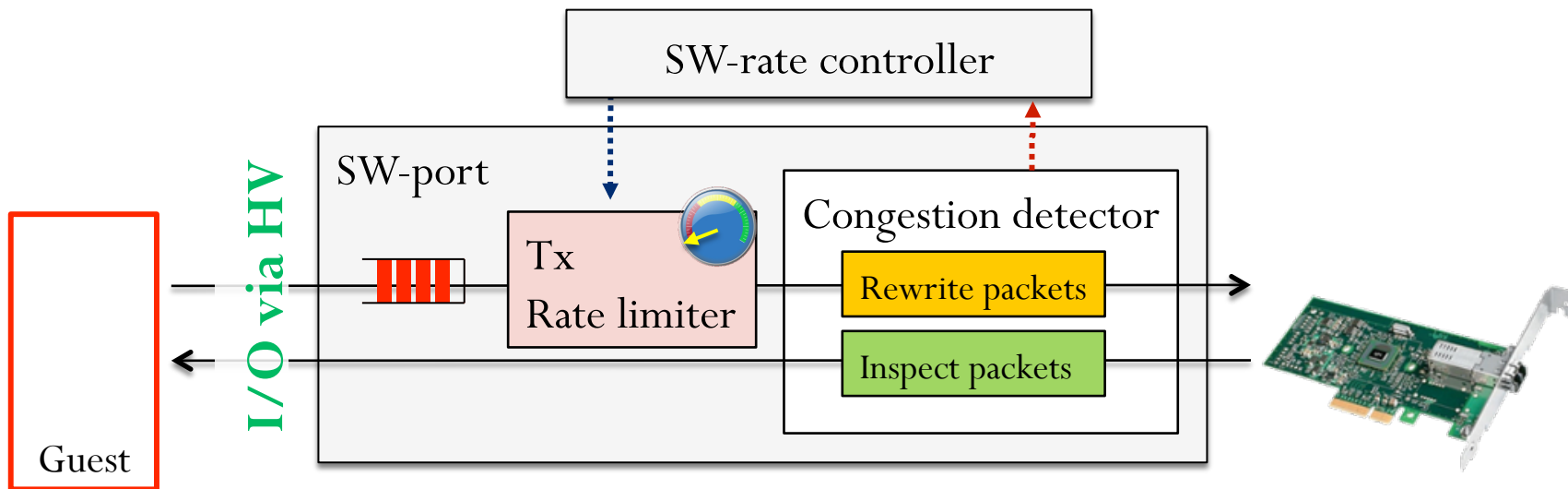
“Bit stealing” solution: Performance improvement



Throughput: 280 Mb/s \Rightarrow 843 Mb/s

Supporting future networks

- Hypervisor vSwitch scales to 1 Gbps, but may be bottleneck for 10 Gbps
- Multiple approaches to scale to 10 Gbps
 - Hypervisor & multi-core optimizations
 - Bypass hypervisor with direct I/O (e.g. SR-IOV)
 - Virtualization-aware physical switch (e.g. NIV, VEPA)
- While efficient, currently direct I/O loses policy control
- Future SR-IOV NICs support **classifiers, filters, rate limiters**

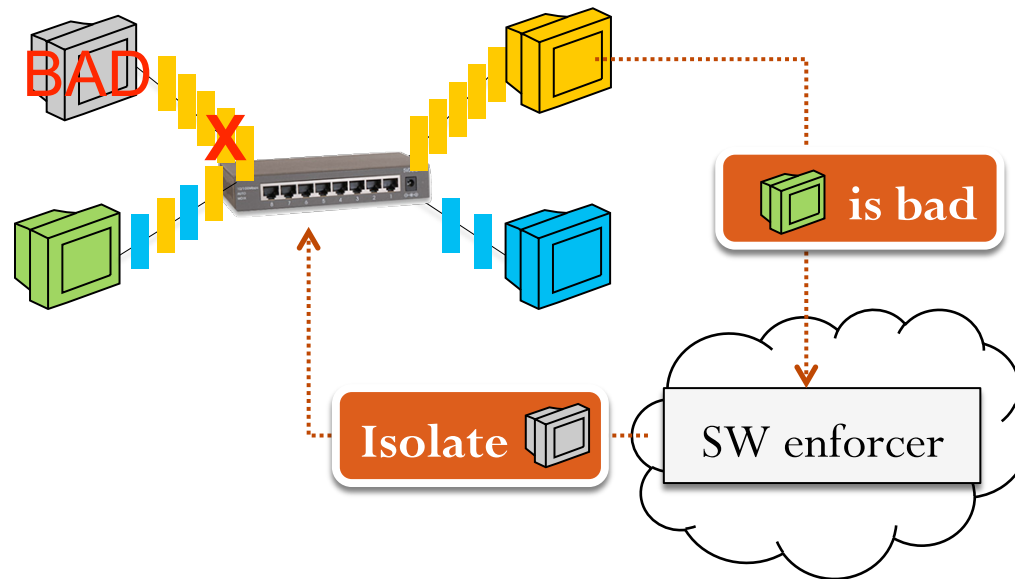


Summary

- Without performance isolation, no protection in cloud against selfish, compromised & malicious tenants
- Hypervisor rate limiters + end-to-end rate controller provide **isolation, control, and efficiency**
- Prototype achieves performance and security on commodity hardware

Preserving performance isolation after hypervisor compromise

- Compromised hypervisor at **source** can flood network
- Solution:
Use network filtering to isolate sources that violate congestion control
 - Destinations act as detector



Preserving performance isolation after hypervisor compromise

- Pitfall: If **destination** is compromised, danger of **DoS** from **false accusations**
- Refinement: Apply least privilege (i.e. fine-grained filtering)

