Measuring and Improving the Reliability of Wide-Area Cloud Paths

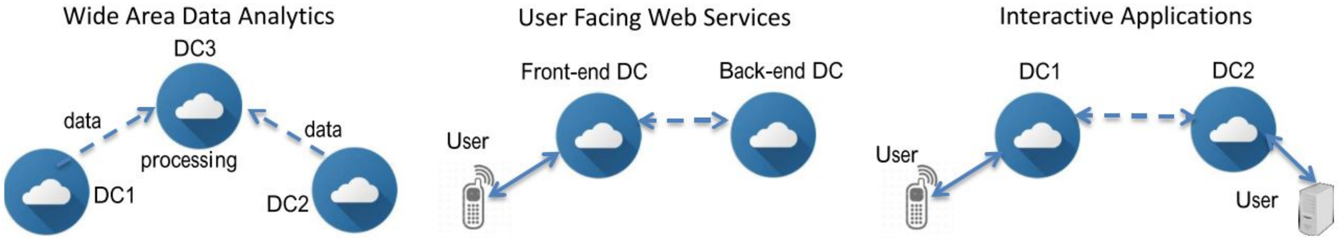
Osama Haq, Mamoon Raja, and Fahad R. Dogar - Tufts University

WWW 2017

Introduction & Motivation

- Cloud providers (CPs) are being used for a wide variety of Internet applications
- Performance and regulatory needs of customers has lead to deployment of many data centers across the globe
- Subset of applications benefit from multi-region deployments for:
  - Load balancing and localized content serving
  - Wide-area distributed applications such as wide-area data analytics [Vulimiri2015]
- CPs often utilize own network for inter-data center communication
- *Little is known about the characteristics of inter-data center paths compared to the public Internet*

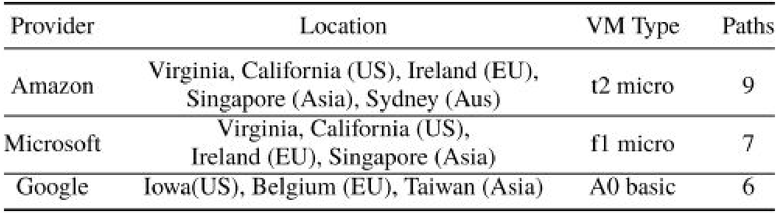
Background



- **Cloud-cloud communication:** typically utilize network of CP. Cost, latency and throughput are important factors
- **User-cloud communication:** most-prevalent use case. Geo-replicated cloud storage, online social network, etc. User forwarded to closes front-end server.
- **User-user communication:** includes VoIP, online gaming, etc. Google Hangout a good example.

Measurement Methodology

- Utilize three major CPs: Amazon, Google, Microsoft
- Deploy VM in each continent that CP offers service
- “*Path*” is defined between each pair of VM from same CP (**22** in total)



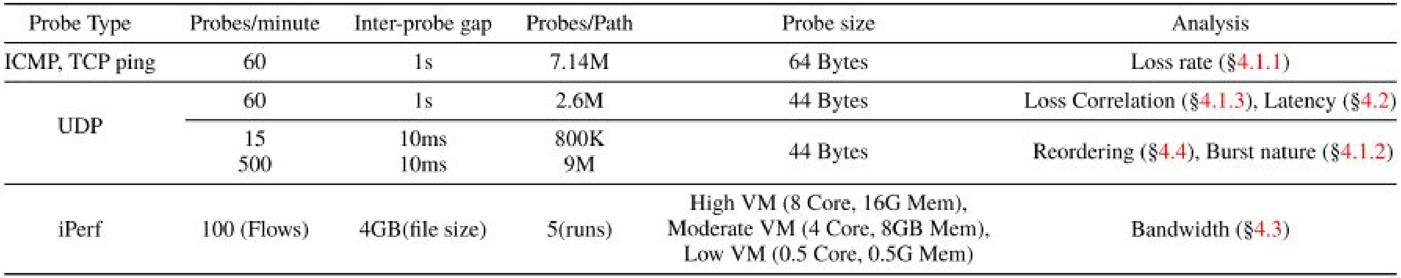
Provider	Location	VM Type	Paths
Amazon	Virginia, California (US), Ireland (EU), Singapore (Asia), Sydney (Aus)	t2 micro	9
Microsoft	Virginia, California (US), Ireland (EU), Singapore (Asia)	f1 micro	7
Google	Iowa(US), Belgium (EU), Taiwan (Asia)	A0 basic	6

Measurement Objectives

- **Bi-directional loss rate:** use ICMP **ping**
  - For Microsoft TCP **ping** was utilized since their network drops ICMP packets
- **Loss characterization:** send a burst of UDP packets for measuring random loss, outages, outage duration, and inter-arrival time of losses
- **AS path characterization:** utilize **traceroute** between VMs for Amazon’s network. For Google and Microsoft VM <-> Internet probes were utilized
- **Bandwidth:** inter-VM bandwidth is measured using **iperf3**
- Public Internet measurement and statistics are performed/gathered through **PlanetLab** and **PingER**

Measurement Campaign

- Performed over a sixteen week period which is spread over the span of eighteen months starting from November 2014 to June 2016



Probe Type	Probes/minute	Inter-probe gap	Probes/Path	Probe size	Analysis
ICMP, TCP ping	60	1s	7.14M	64 Bytes	Loss rate (§4.1.1)
UDP	60	1s	2.6M	44 Bytes	Loss Correlation (§4.1.3), Latency (§4.2)
	15 500	10ms 10ms	800K 9M	44 Bytes	Reordering (§4.4), Burst nature (§4.1.2)
iPerf	100 (Flows)	4GB(file size)	5(runs)	High VM (8 Core, 16G Mem), Moderate VM (4 Core, 8GB Mem), Low VM (0.5 Core, 0.5G Mem)	Bandwidth (§4.3)

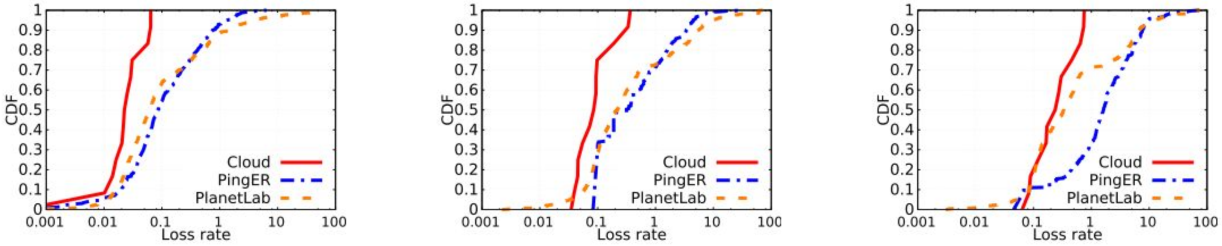
Results & Analyses

Loss Rate - Longitudinal Analysis

- For each path aggregate loss rate is measured using all probes sent through that path
  - 7 million probes per path

- Public Internet loss rate is measured using same set of probes from PlanetLab nodes corresponding to **1200** unique paths as well as **300** unique PingER paths
  - PingER probing frequency less than this study but sampling studies measurements at the same frequency produces same results

Loss Rate - Longitudinal Analysis



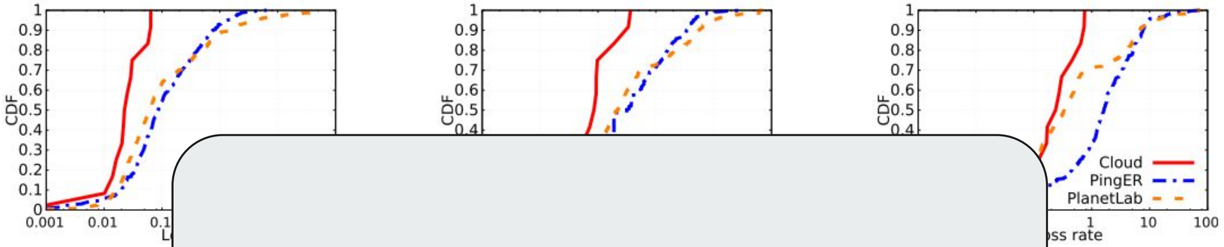
(a) Aggregate Loss Rate

(b) 95th Percentile (Per-Day)

(c) 99th Percentile (Per-Day)

- Aggregate loss rate comparable for best *paths* while cloud paths have a much lower loss rate for remainder of paths
- Looking into extreme cases of loss-rate (95th and 99th percentile) we observe much lower loss-rates for cloud paths
  - < **0.8%** for majority of cloud paths
  - **30%** of Internet paths have at least **0.1%** loss rate

Loss Rate - Longitudinal Analysis



(a) Aggrega

ntile (Per-Day)

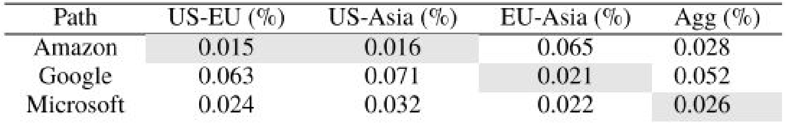
Cloud path loss-rate is more predictable compared to an average Internet path

- Aggregate loss ra h lower loss rate for remainder of pat
- Looking into extreme cases of loss rate (95th and 99th percentile) we observe much lower loss-rate for cloud
  - < **0.8%** for majority of cloud paths
  - **30%** of paths have at least **0.1%** loss rate

Loss Rate - Cross CP Comparison

- Average loss rate between VMs of each CP

- No clear winner, depending on deployment and utilization of various regions loss rates could vary



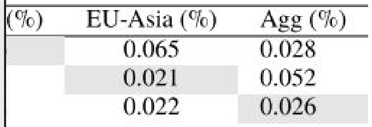
Path	US-EU (%)	US-Asia (%)	EU-Asia (%)	Agg (%)
Amazon	0.015	0.016	0.065	0.028
Google	0.063	0.071	0.021	0.052
Microsoft	0.024	0.032	0.022	0.026

Loss Rate - Cross CP Comparison

- Average loss rate between VMs of each CP

- No clear winner, and utilization of could vary

Best choice is dependent on employed regions



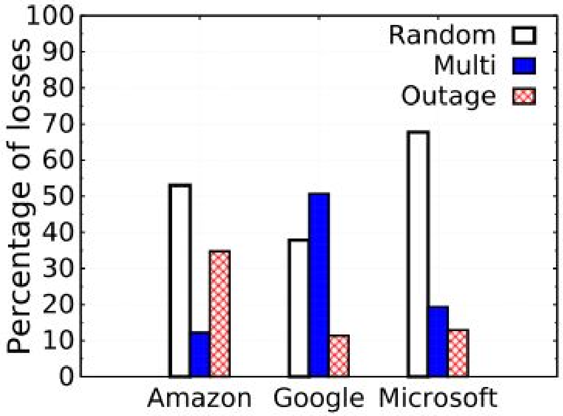
(%)	EU-Asia (%)	Agg (%)
	0.065	0.028
	0.021	0.052
	0.022	0.026

Loss Characteristics

- Send burst of UDP packets for **5** seconds after every **1** minute
- Each burst divided into buckets of **15** packets corresponding to a period of **150** ms
- **Loss episode:** a bucket containing at least one packet loss
    - **Random:** if only 1 packet is lost
    - **Multi:** if between 2 and 14 packets are lost
    - **Outage:** if all 15 packets are lost
- *No statistics on the percentage of buckets that experience a loss episode is provided!*

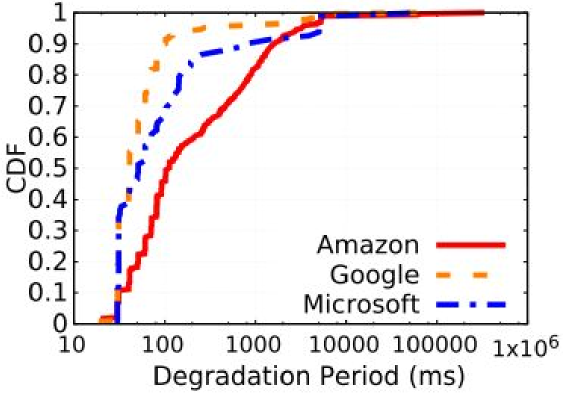
Loss Characteristics - Loss Episodes

- All CPs experience at least **35%** random loss in their loss episodes
    - MS and Google > **50%**
- Amazon experiences more outages while having less multi packet loss periods
- **Degraded periods:** loss episodes which are not random



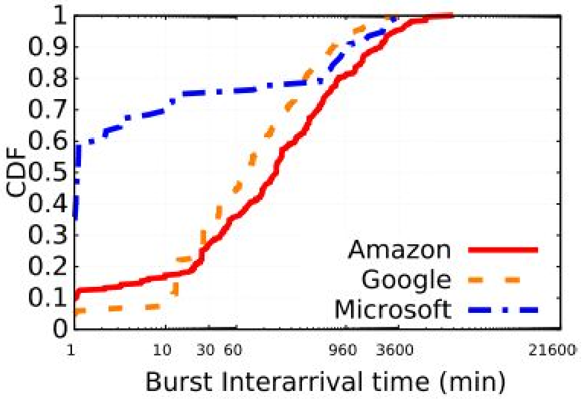
Loss Characteristics - Degraded Periods

- Degraded periods that span consecutive bursts are combined to measure total duration
  - *interpolating network degradation using 5 second measurements with 55 second gaps in between doesn't seem reasonable!*
- Degradation could last up to minutes but majority (**70%**) are less than a second
- Amazon has longest degraded periods



Loss Characteristics - Inter-arrival Time

- Measure time between degraded periods
- **70%** of inter-arrival times for Microsoft are less than **10** minutes
- **50%** of inter-arrivals for Google and Amazon are less than **2** hours



Loss Correlation

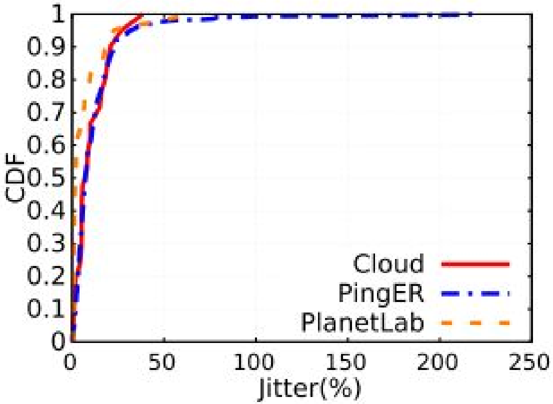
- Investigate whether loss events are correlated between CPs or not, i.e. if they share any inter-data center paths
- Perform uni-directional UDP probes for measuring losses on a per minute basis
- Use Pearson correlation coefficient to compare losses on two paths
- Losses are independent:
    - Forward and reverse path independent, US-EU and EU-US for Amazon has a correlation of **0.015**
    - Losses across paths of the same CP are independent. Correlation for US-Asia and US-EU of Microsoft is **0.0061**
    - Losses for paths of different CPs are independent. Correlation for US-EU for Amazon and Microsoft is **0.001**

Latency

- Measure latency variation (jitter in RTT) for CPs
- Rely on **ping** probes for cloud paths and compare them against PlanetLab and PingER
- Jitter is defined as percentage difference between **95th** percentile and **median** of RTT for each path
- Differences in latency for forward and reverse path are measured using uni-directional UDP probes
    - Clocks for VMs are synchronized

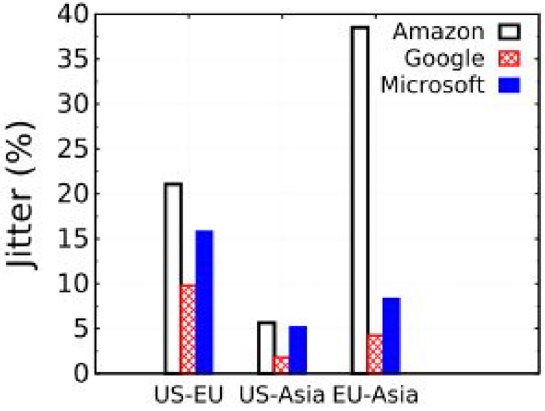
Jitter - Cloud vs Internet

- Cloud and Internet have relatively similar jitter
- Majority of paths have less than **30%** jitter
- Internet paths have longer tail in distribution
    - **1%** of paths have more than **100%** jitter



Jitter - Cross CP Comparison

- Google offering best performance
- Majority of paths have a jitter less than **20%**
- Only one Amazon path between EU-Asia had unexpected jitter
  - Forward path had latency of **100ms**
  - Reverse path had latency of **100-160ms**
  - Only path were Amazon utilizes external service provider

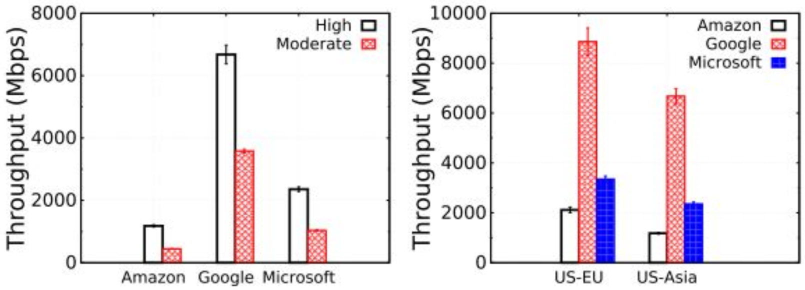


Bandwidth - Measurement

- Use **iperf3** to measure bandwidth between VMs
- Sender uses **100** TCP flows and transfers **4GB** of data
- Each measurement is repeated for **5** runs
- Two types of VMs were used: *moderate* and *high*

Bandwidth - Results

- Bandwidth is dependent on the hardware Tier and is rate limited
- Bandwidth could be higher than **1Gbps** and could reach up to **9Gbps**
- US-EU paths exhibit higher bandwidth compared to US-Asia
- Google offers highest bandwidth



(a) Bandwidth across machine types

(b) Bandwidth across regions

Packet Reordering

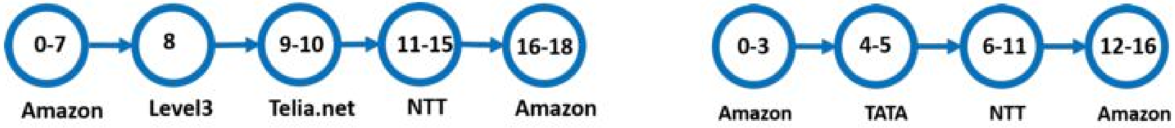
- Use Paxon’s definition of packet reordering
  - Count late arrivals rather than early arrivals
  - If packet 4 arrives before packets 1-3 we count 3 out of order packets
- Use UDP probes to measure packet reordering in both directions
  - Between cloud nodes
  - Between PlanetLab nodes
- Overall negligible amount of packet reordering was observed for both CP and PlanetLab nodes
  - Google had greatest packet reordering **< 0.02%**
- Internet packet reordering on the decline and **< 1%** in recent studies

Cloud vs Internet AS Path

- Use **traceroute** to probe VMs
  - Amazon being the only network to allow ICMP probes from VM
  - For Google and Microsoft VMs were probed from PlanetLab nodes
- All Amazon paths except for EU (Ireland) to Asia (Singapore) being handled within Amazon's own network

Cloud vs Internet AS Path

- Use **traceroute** to probe VMs
  - Amazon being the only network to allow ICMP probes from VM
  - For Google and Microsoft VMs were probed from PlanetLab nodes
- All Am azon’s own netwo



(a) Forward Path

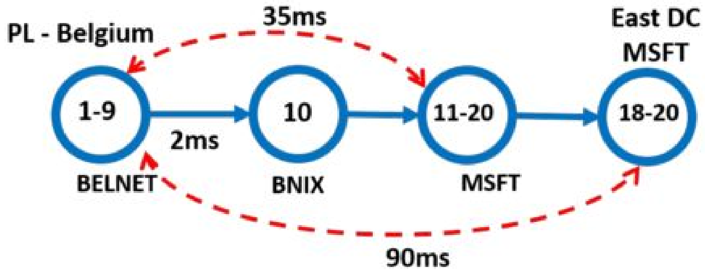
(b) Reverse Path

Cloud vs Internet AS Path

- Use **traceroute** to probe VMs
  - Amazon being the only network to allow ICMP probes from VM
  - For Google and Microsoft, VMs were probed from PlanetLab nodes
- All Amazon paths except for EU (Ireland) to Asia (Singapore) being handled within Amazon's own network
- Microsoft and Google handle inter-continental traffic by themselves
  - PlanetLab probe handed off to nearest datacenter

Cloud vs Internet AS Path

- Use **traceroute** to probe VMs
  - Amazon being the only network to allow ICMP probes from VM
  - For Google and Mi
- All Amazon paths exce dled within Amazon’s own network
- Microsoft and Google
  - PlanetLab probe h

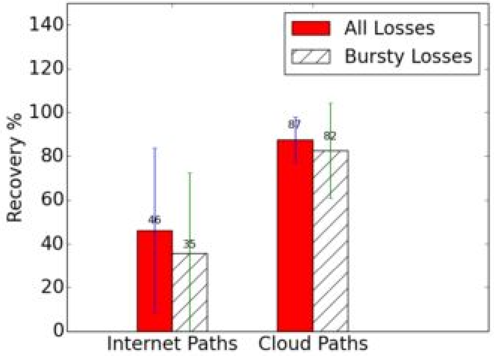


Case Study

- Investigate effects of known loss mitigation techniques to improve cloud path reliability
    - Detour Routing
    - Forward Error Correction (FEC)
- **Detour routing:**
    - Two week measurement period
    - PlanetLab node in Europe set as detour node
    - Modified UDP probes to send duplicate packets, one through normal path and another through a detour
    - Packet loss reported if none of the duplicate packets reach the destination
- **FEC:**
    - What if scenario, no actual measurement is performed
    - For every burst of **15** packets consider **4** different FEC levels: **1, 2, 4, 8** FEC packets

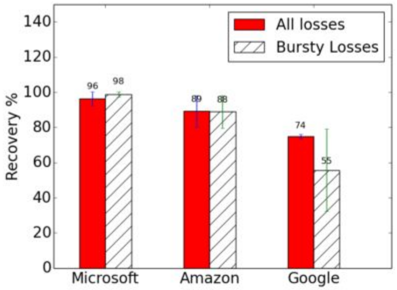
Detour Routing

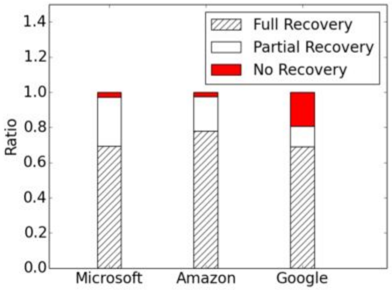
- Detour routing more effective for cloud paths **87%** compared to **46%** for the Internet
- Detour routing less effective for bursty losses, since duplicates are sent in succession



Detour Routing - Cont

- Microsoft and Amazon have higher recovering rates **> 90%** while Google can recover **74%** of loss episodes
- Loss episodes divided into three categories: full, partial and no recovery
- Microsoft and Amazon rarely have episode which doesn’t benefit from detour routing while Google has about **20%** *no recovery* episodes

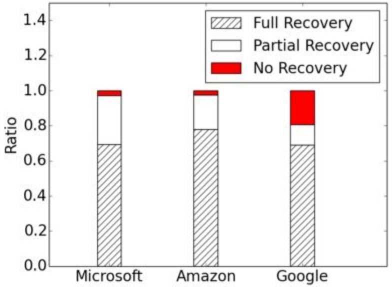




Detour Routing - Cont

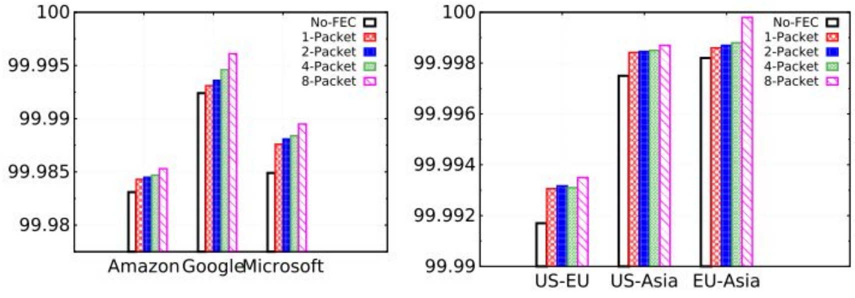
- Microsoft and Amazon have higher recovering rates > **90%** while Google can recover **74%** of lo
- Loss episodes div full, partial and n
- Microsoft and An which doesn't be while Google has episodes

Detour routing effective in preventing loss.
For Microsoft we can reach five 9's of availability.



FEC

- Recover from all random losses
- Gain **99.99%** availability with less than **10%** overhead
- Google benefits the most due to bursty nature of losses
- High levels of FEC provide no gain for MS and Amazon (losses mostly random or outage)



(a) Aggregated results for all providers

(b) Amazon results

Thank You!