Non-cooperative Diagnosis of Submarine Cable Faults

Edmond W. W. Chan, Xiapu Luo, Waiting W. T. Fok, Weichao Li, and Rocky K. C. Chang

Department of Computing

The Hong Kong Polytechnic University

19 January 2011

This talk based on

- Xiapu Luo, Edmond W. W. Chan and Rocky K. C. Chang, "Design and Implementation of TCP Data Probes for Reliable Network Path Monitoring," *Proc. USENIX Annual Tech. Conf.*, June 2009.
- Rocky K. C. Chang, Edmond W. W. Chan, Weichao Li, Waiting W. T. Fok, and Xiapu Luo, "Could Ash Cloud or Deep-Sea Current Overwhelm the Internet?" *Proc. USENIX HotDep*, October 2010.
- Edmond W. W. Chan, Xiapu Luo, Waiting W. T. Fok, Weichao Li, and Rocky K. C. Chang, "Non-cooperative Diagnosis of Submarine Cable Faults", *Proc. PAM 2011*, March 2011.

Outline

1. OneProbe measurement
2. Ash Cloud or Deep-Sea Current?
3. An impact analysis of a submarine cable fault
4. Conclusions and future work

1. OneProbe measurement

Our design principles

- Use normal data packet to measure data-path quality.
- Use normal and basic data transmission mechanisms
- Integrated into normal application sessions.

Our design principles

- Use normal TCP data packet to measure data-path quality.
- Use normal and basic TCP data transmission mechanisms specified in RFC 793.
- Integrated into normal HTTP application sessions.

Our design principles

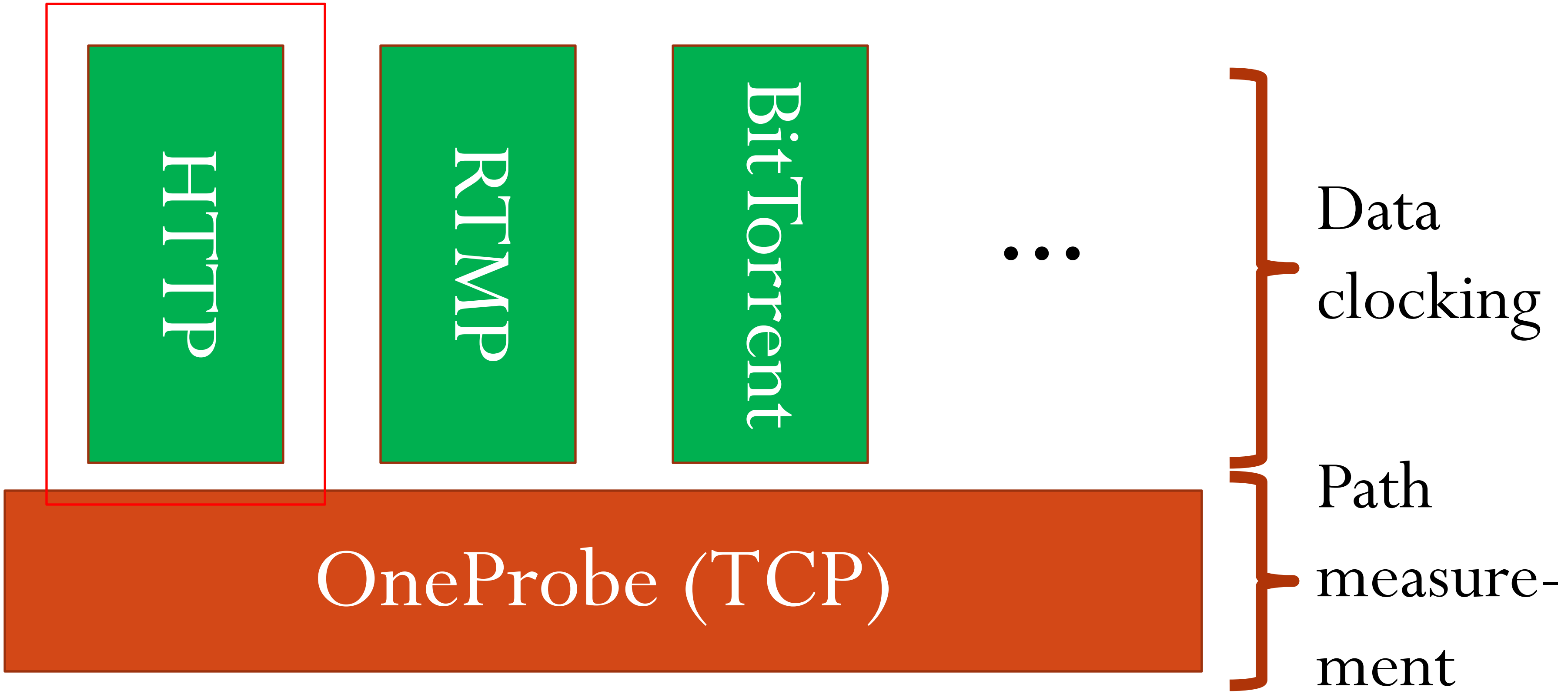
- Use normal TCP data packet to measure data-path quality.
- Use normal and basic TCP data transmission mechanisms specified in RFC 793.
- Integrated into normal HTTP application sessions.

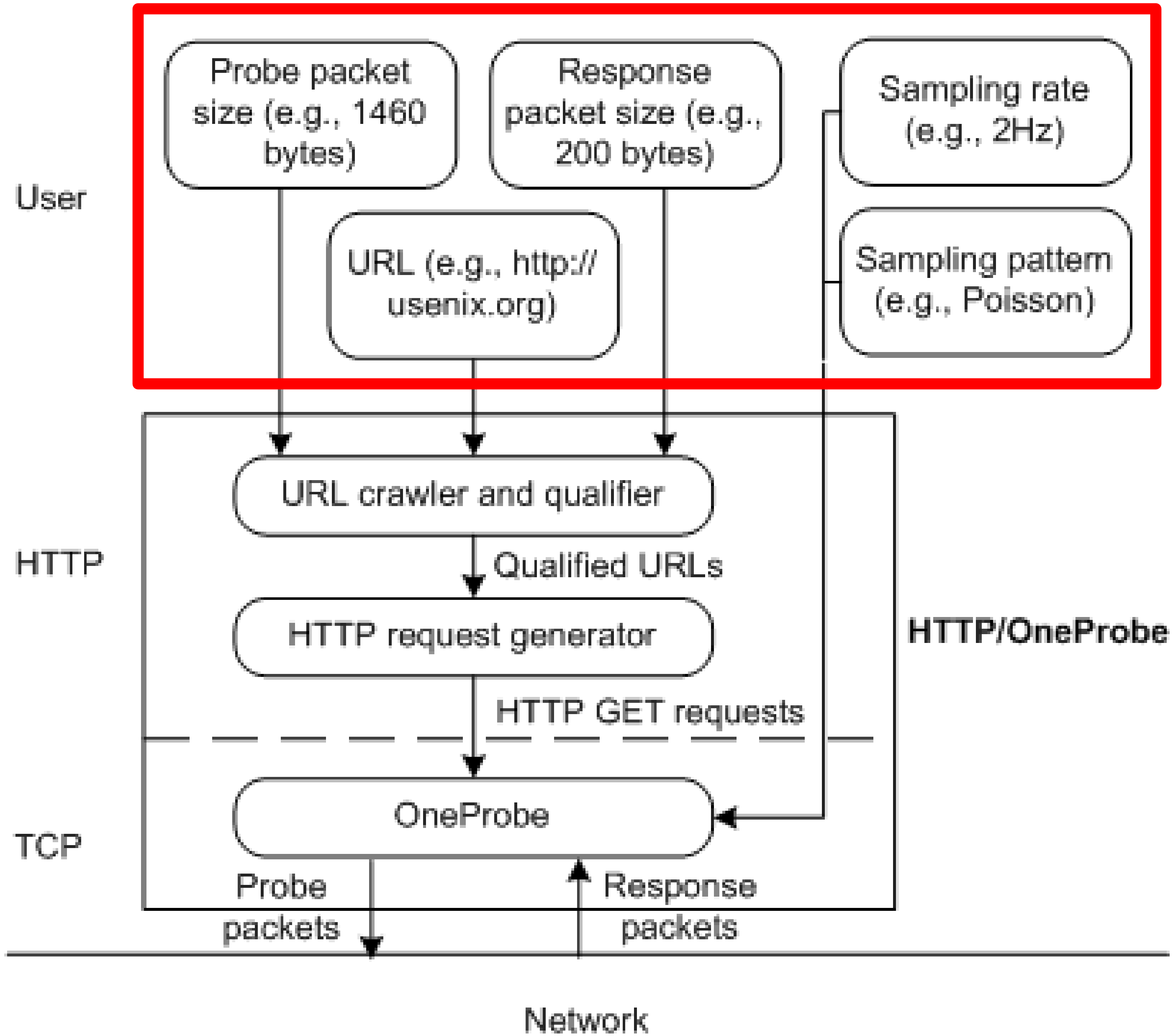
Reliable measurement

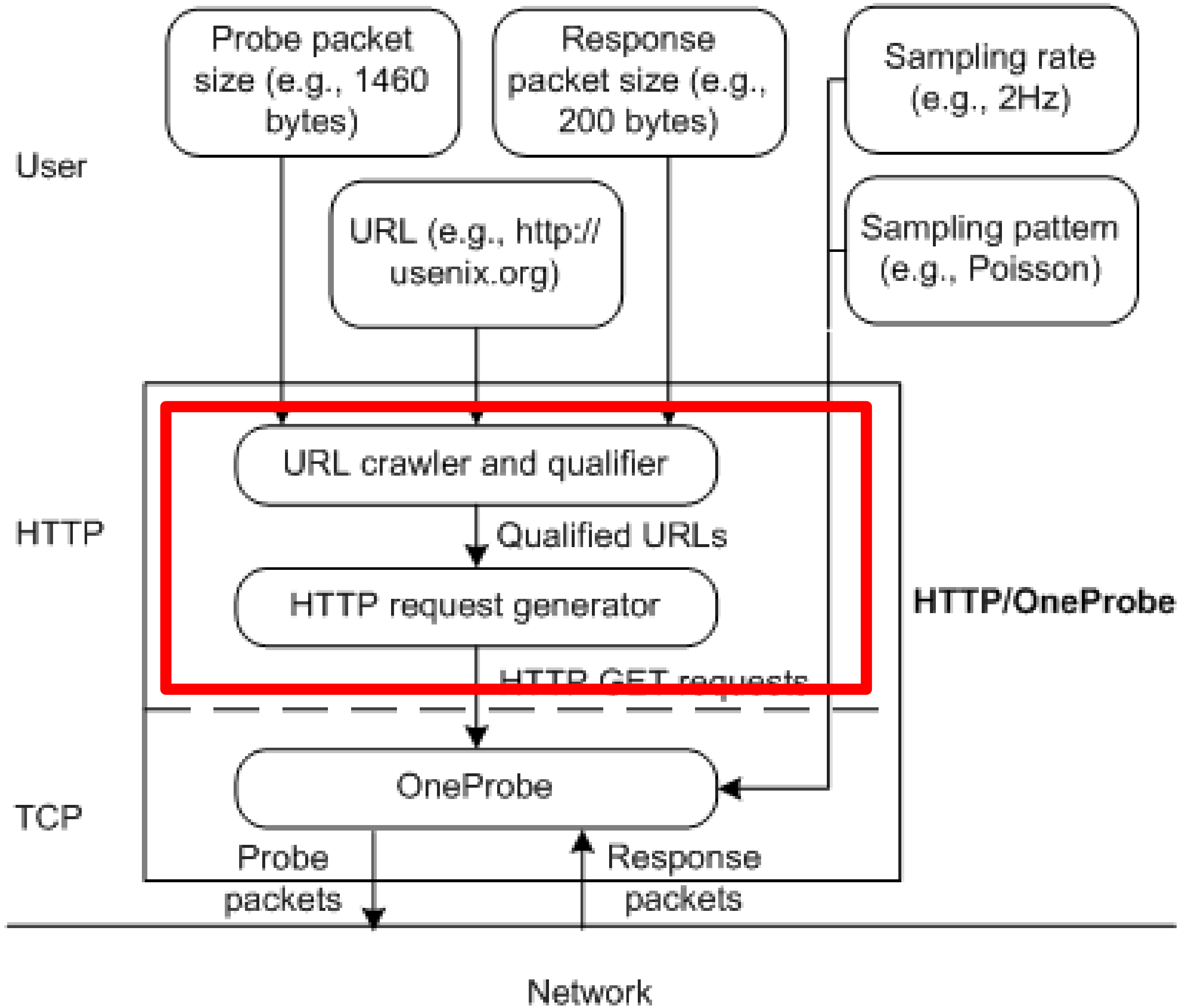
What does HTTP/OneProbe offer?

- Continuous path monitoring in an HTTP session (stateful measurement)
- All in one:
  - Round-trip time
  - Loss rate (uni-directional)
  - Reordering rate (uni-directional)
  - Capacity (uni-directional)
  - Loss-pair analysis
  - …

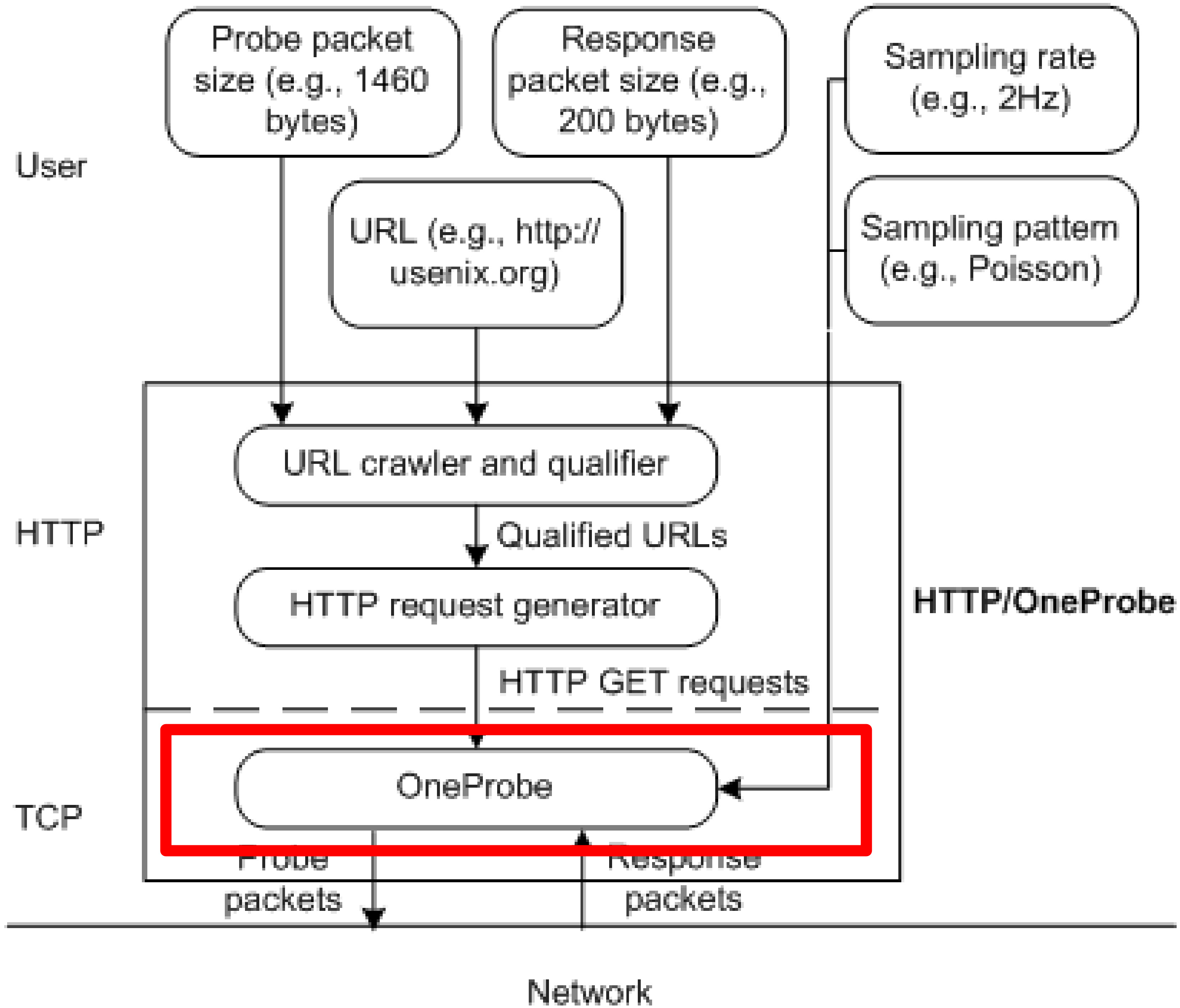
HTTP/OneProbe




Probe packet size (e.g., 1460 bytes)
Response packet size (e.g., 200 bytes)
Sampling rate (e.g., 2Hz)
User
URL (e.g., http:// usenix.org)
Sampling pattern (e.g., Poisson)
URL crawler and qualifier
HTTP
Qualified URLs
HTTP request generator
HTTP/OneProbe
HTTP GET requests
OneProbe
TCP
Probe packets
Response packets
Network

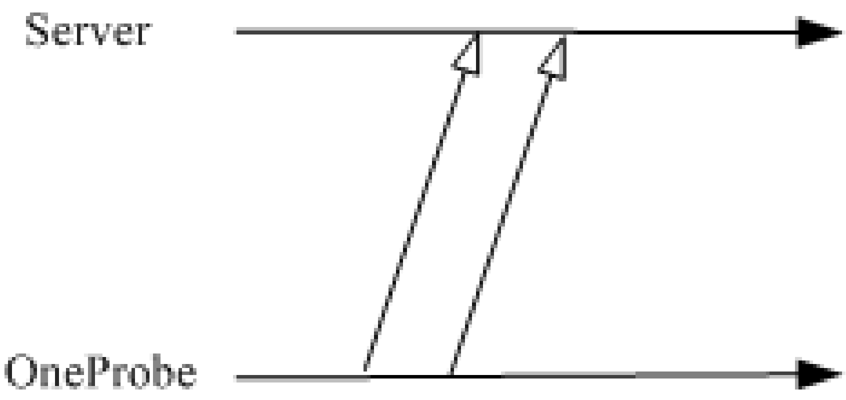


Probe packet size (e.g., 1460 bytes)
Response packet size (e.g., 200 bytes)
Sampling rate (e.g., 2Hz)
Sampling pattern (e.g., Poisson)
User
URL (e.g., http://usenix.org)
URL crawler and qualifier
HTTP
Qualified URLs
HTTP request generator
HTTP/OneProbe
HTTP GET requests
OneProbe
TCP
Probe packets
Response packets
Network


Probe packet size (e.g., 1460 bytes)
Response packet size (e.g., 200 bytes)
Sampling rate (e.g., 2Hz)
User
URL (e.g., http://usenix.org)
Sampling pattern (e.g., Poisson)
URL crawler and qualifier
HTTP
Qualified URLs
HTTP request generator
HTTP/OneProbe
HTTP GET requests
OneProbe
TCP
packets
packets
Network

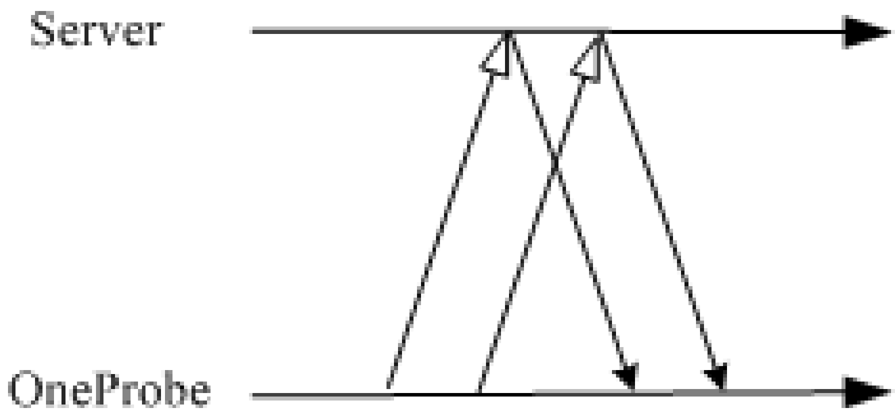
The probe design

- Send two back-to-back probe data packets.
  - Capacity measurement based on packet-pair dispersion
  - At least two packets for packet reordering
  - Determine which packet is lost.



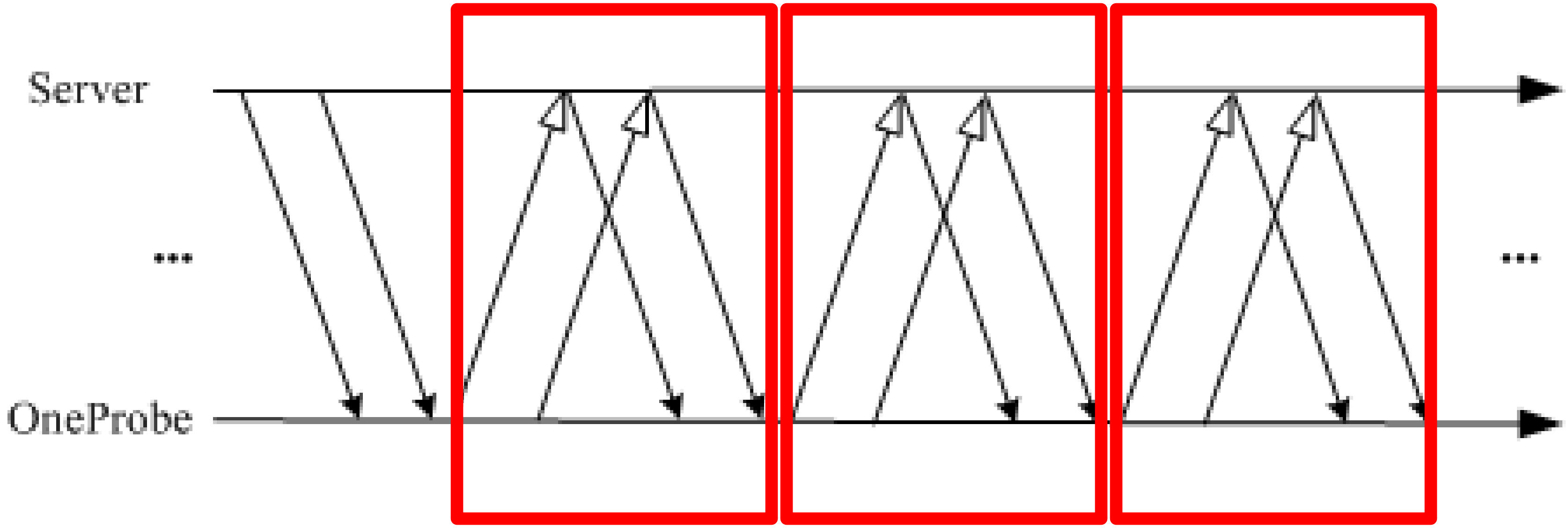
The probe design (cont'd)

- Similarly for the response packets

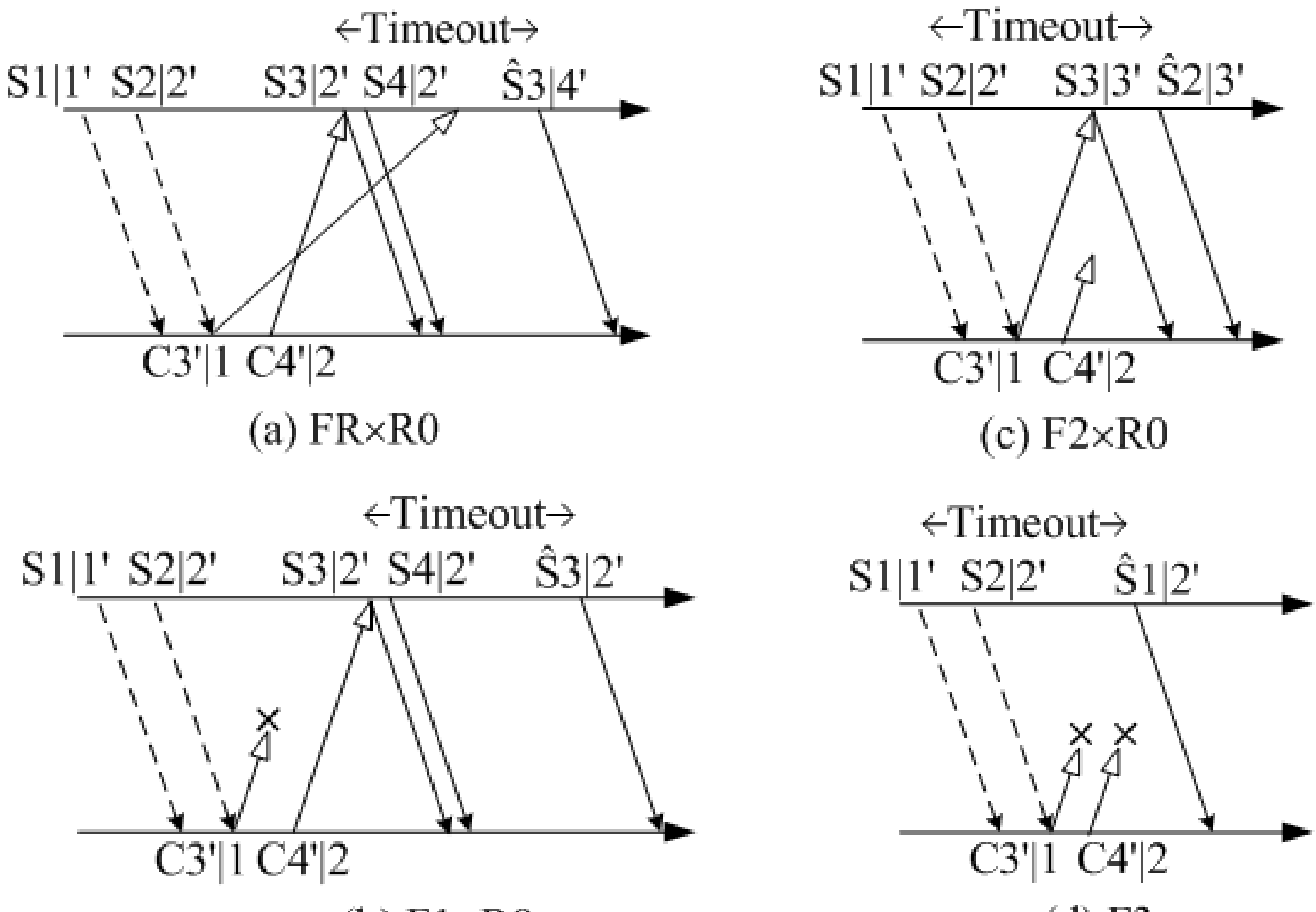


- Each probe packet elicits a response packet.
  - Adv. Window = 2 and acknowledge only 1 packet.

Bootstrapping and continuous monitoring



Loss and reordering measurement via response diversity



(a) FR×R0

(c) F2×R0

(b) F1×R0

(d) F3

18 possible path events

	R0	RR	R1	R2	R3
F0	√	√	√	√	√
FR	√	√	√	√	√
F1	√	√	√	√	√
F2	√	–	√	–	–
F3	–	–	–	–	–

Path events	1st response packets	2nd response packets	3rd response packets
1. F0 × R0	$S3\|3'$	$S4\|4'$	–
2. F0 × RR	$S4\|4'$	$S3\|3'$	–
3. F0 × R1	$S4\|4'$	$\widehat{S}3\|4'$	–
4. F0 × R2	$S3\|3'$	$\widehat{S}3\|4'$	–
5. F0 × R3	$\widehat{S}3\|4'$	–	–
6. FR × R0	$S3\|2'$	$S4\|2'$	$\widehat{S}3\|4'$
7. FR × RR	$S4\|2'$	$S3\|2'$	$\widehat{S}3\|4'$
8. FR × R1	$S4\|2'$	$\widehat{S}3\|4'$	–
9. FR × R2	$S3\|2'$	$\widehat{S}3\|4'$	–
10. FR × R3	$\widehat{S}3\|4'$	–	–
11. F1 × R0	$S3\|2'$	$S4\|2'$	$\widehat{S}3\|2'$
12. F1 × RR	$S4\|2'$	$S3\|2'$	$\widehat{S}3\|2'$
13. F1 × R1	$S4\|2'$	$\widehat{S}3\|2'$	–
14. F1 × R2	$S3\|2'$	$\widehat{S}3\|2'$	–
15. F1 × R3	$\widehat{S}3\|2'$	–	–
16. F2 × R0	$S3\|3'$	$\widehat{S}2\|3'$	–
17. F2 × R1	$\widehat{S}2\|3'$	–	–
18. F3	$\widehat{S}1\|2'$	–	–

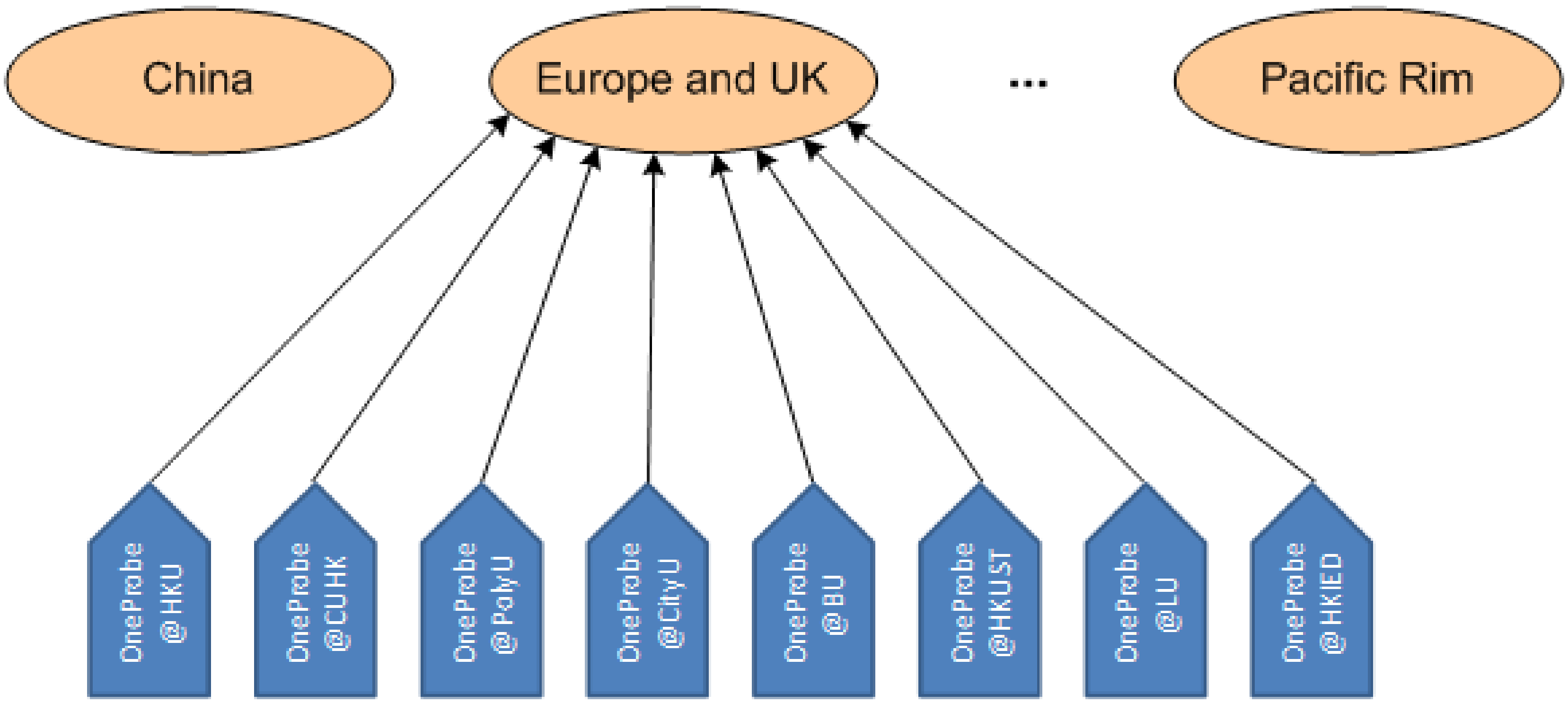
Path event distinguishability

- All 18 cases can be distinguished except for
  - A1. F1×R2 and F1×R3
  - A2. F1×RR and F1×R1
  - A3. F0×R3 and FR×R3
- Resolving the ambiguities
  - A1 and A2: use RTT.
  - A3: use TCP timestamping.

Other issues

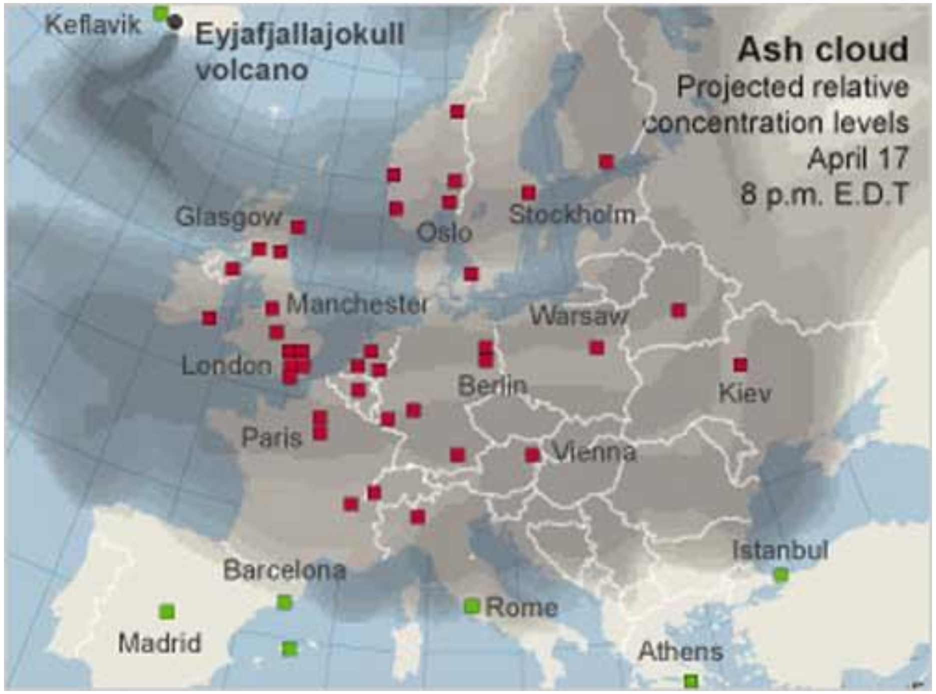
- Use TCP ACKs to improve performance
- When to start a new probe round when timeout is involved
- Use concurrent TCP connections to increase sampling rate.

Collaborative path monitoring (Harnet)

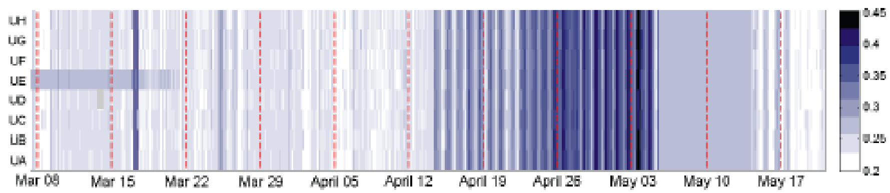


2. Ash Cloud or Deep-Sea Current?

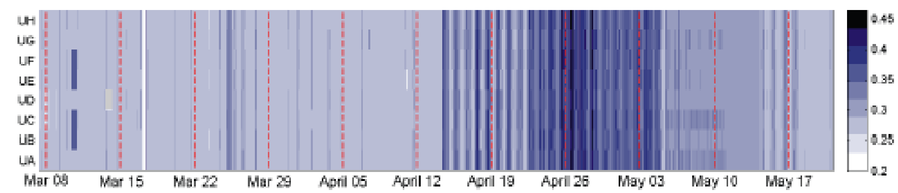
Eyjafjallajöekull volcano eruption



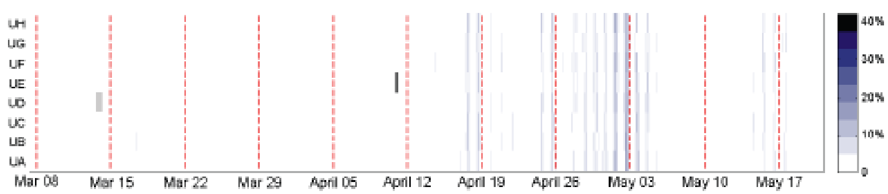
Path-quality degradation for NOK (Finland) and ENG (in UK)



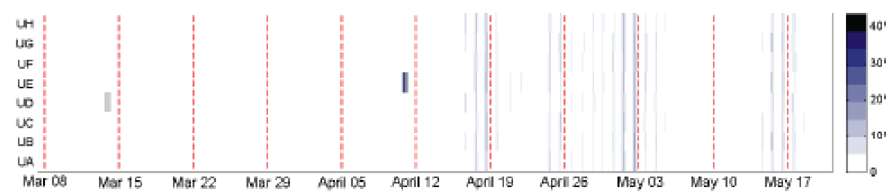
(a) The ENG path's RTT.



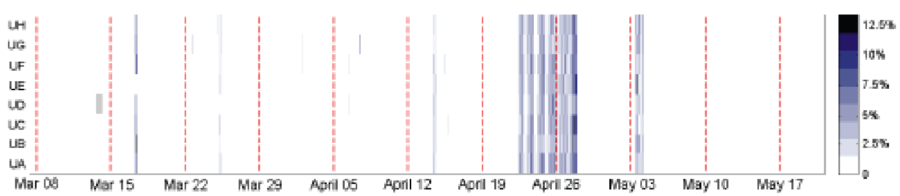
(b) The NOK path's RTT.



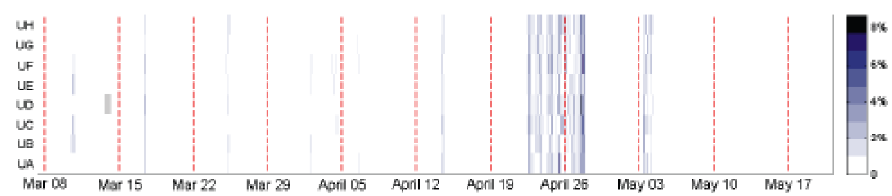
(c) The ENG path's forward-path loss.



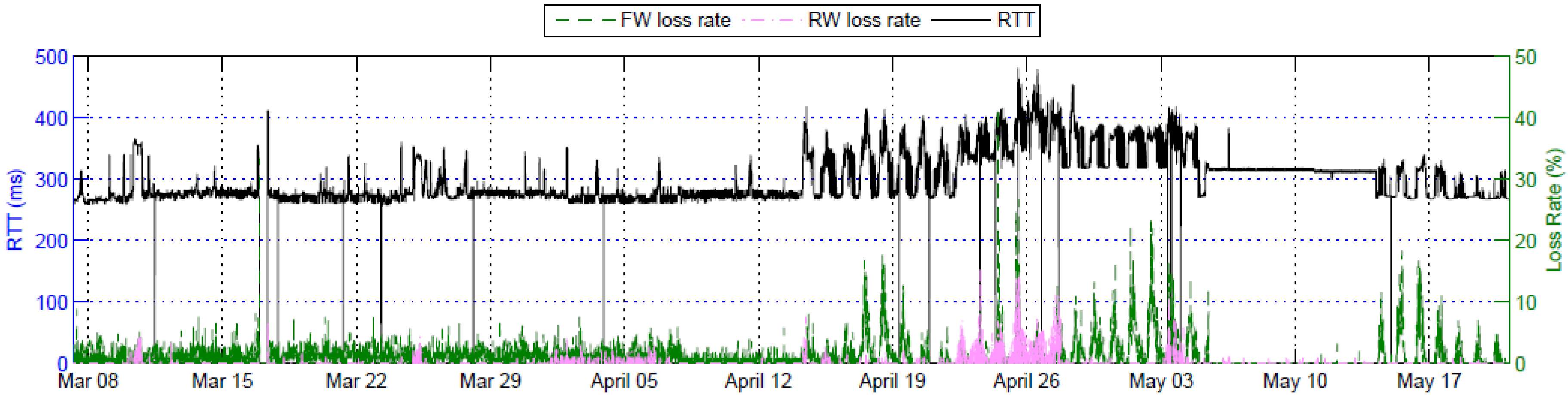
(d) The NOK path's forward-path loss.



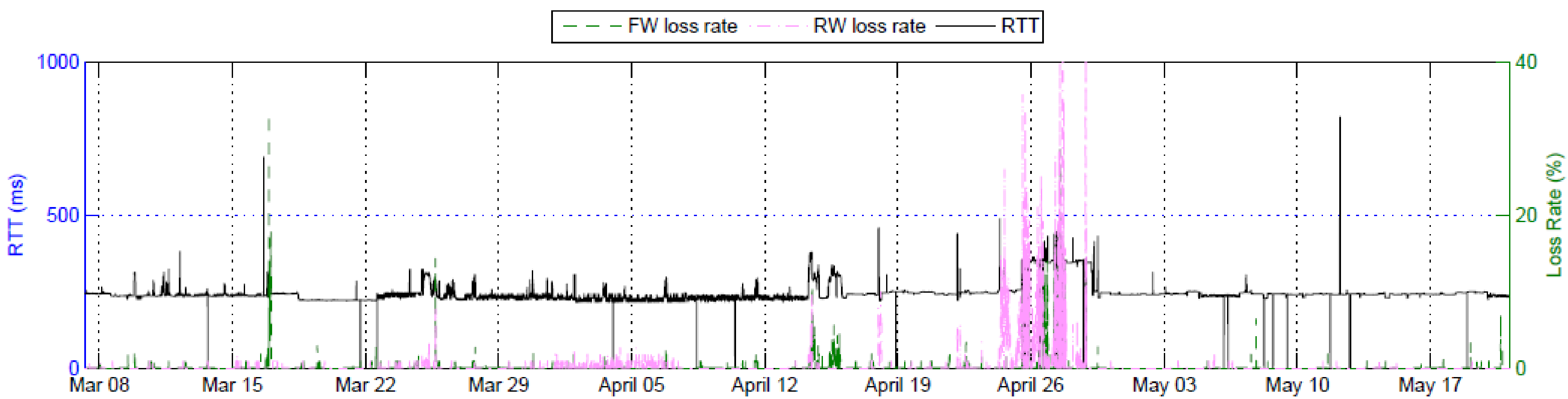
(e) The ENG path's reverse-path loss.



(f) The NOK path's reverse-path loss.



(a) UB→NOK's RTT and loss rates.



(b) UB→BBC's RTT and loss rates.

Network congestion caused by the volcano ashes?

- The surges on packet loss and RTT occurred on 14 April 2009.
- But
  - The onsets of the path congestion and air traffic disruption do not entirely match.
  - Some of the peak loss rate and RTT occurred on weekends.
  - Path congestion can still be observed at the end of the measurement period.

A traceroute analysis

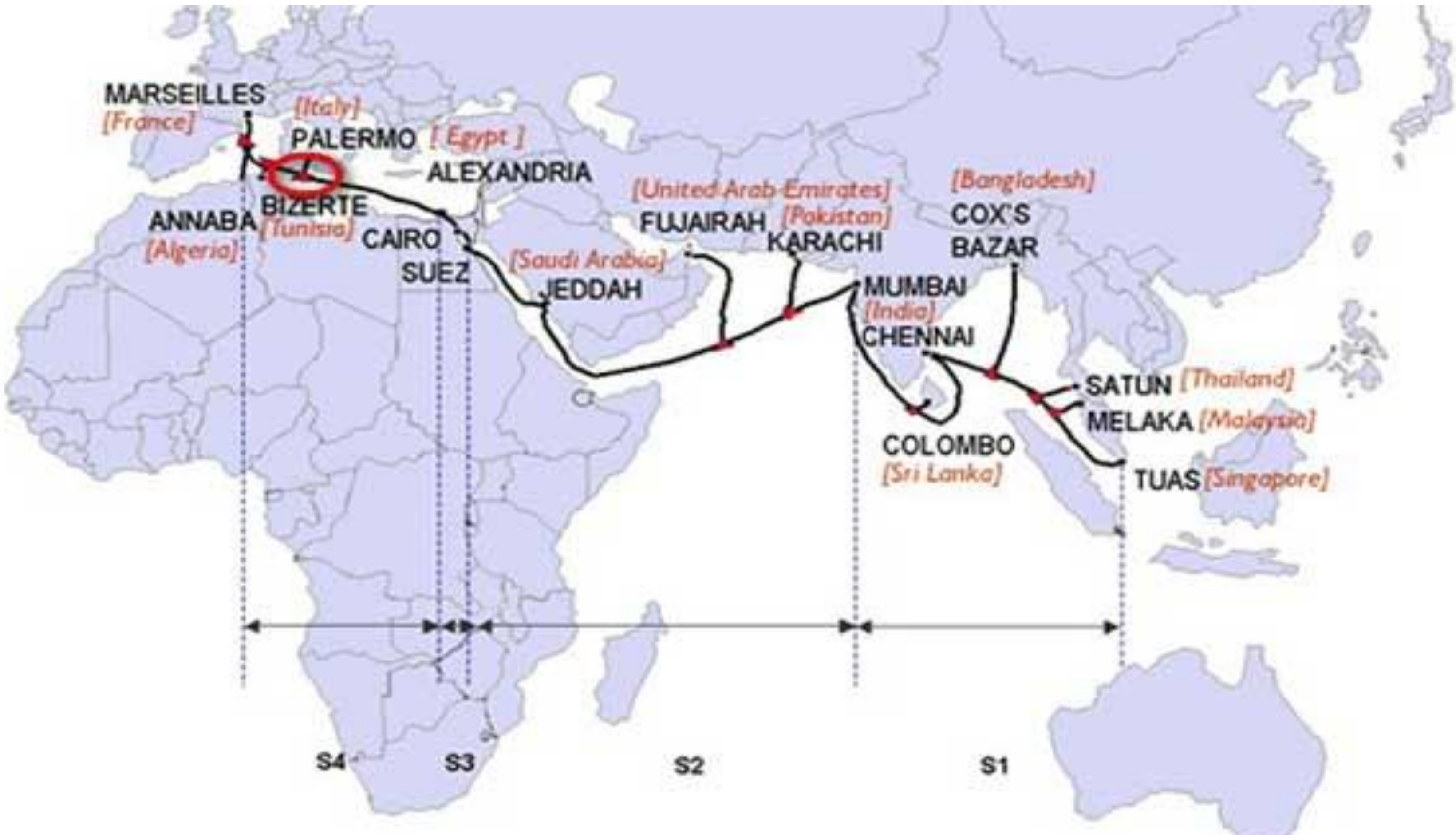
- Comparing the 64 paths from HK to destinations in Europe (8 HK nodes x 8 web servers)
- Only the ENG and NOK paths experienced the congestion.
- The only AS that was common to the ENG and NOK paths was the FLAG network which, however, did not appear in other paths
  - Except for the BBC paths before 16 Apr 2010 07:39:00 GMT.

3. An impact analysis of a submarine cable fault

A SEA-ME-WE 4 cable fault

- The South East Asia-Middle East-Western Europe 4 (SEA-ME-WE 4) submarine cable
- The SEA-ME-WE 4 cable encountered a shunt fault on the segment between Alexandria and Marseille on 14 April 2010.
- The repair was started on 25 April 2010, and it took four days to complete.
- During the repair, the service for the westbound traffic to Europe was not available.

The SEA-ME-WE 4 cable



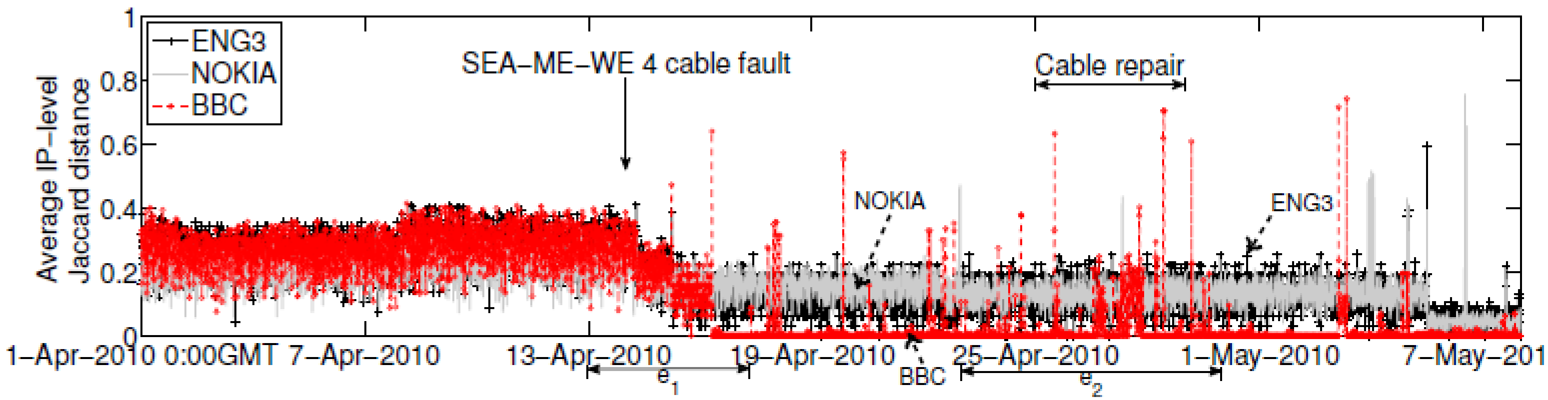
Measurement metrics

- Path-quality metrics:
  - RTT
  - Unidirectional packet loss rates
  - Unidirectional packet reordering rates
  - Loss-pair analysis (correlation between packet loss and RTT)
- Route changes:
  - IP-level Jaccard distance for IP route changes

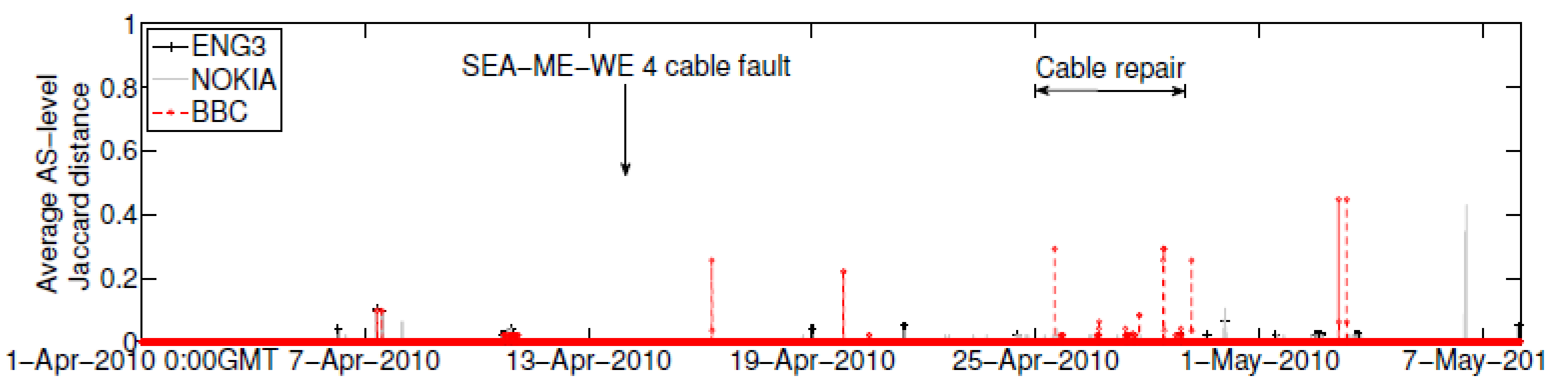
$$J_\delta(R_{i-1}, R_i) = 1 - \frac{|R_{i-1} \cap R_i|}{|R_{i-1} \cup R_i|}$$

  - AS-level Jaccard distance for AS path changes

Route changes for ENG3, NOKIA, and BBC

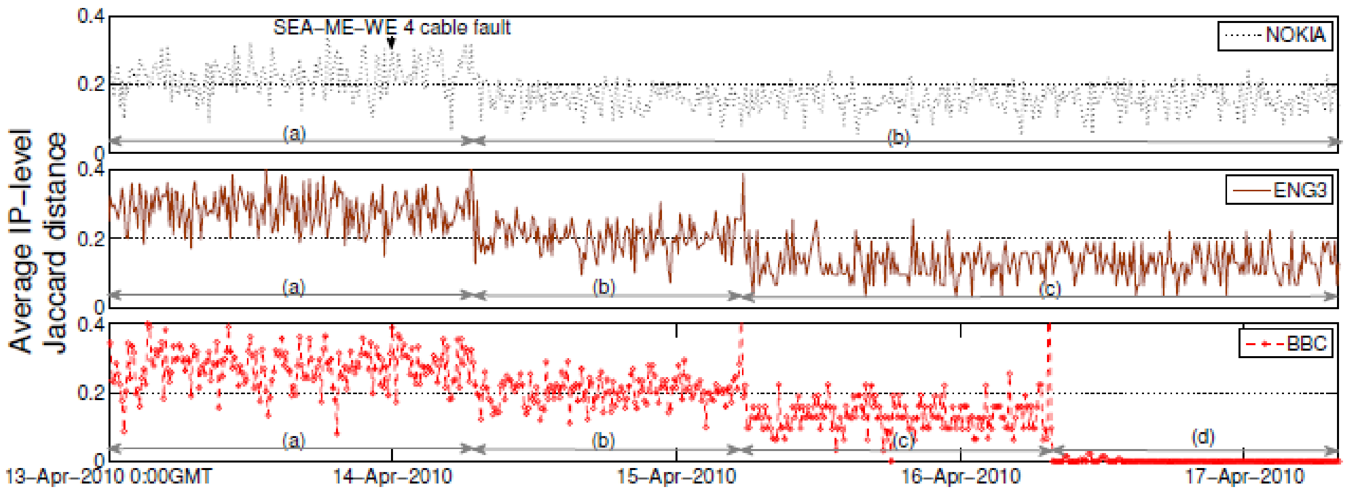


(a) IP-level Jaccard distances.

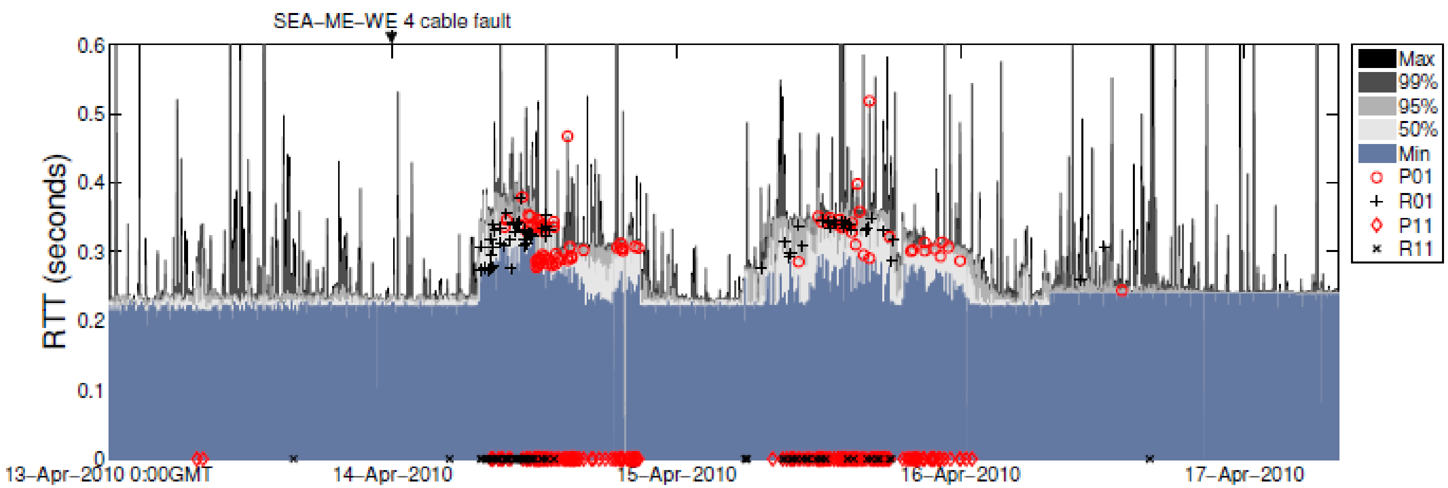


(b) AS-level Jaccard distances.

Zooming into episode e_1

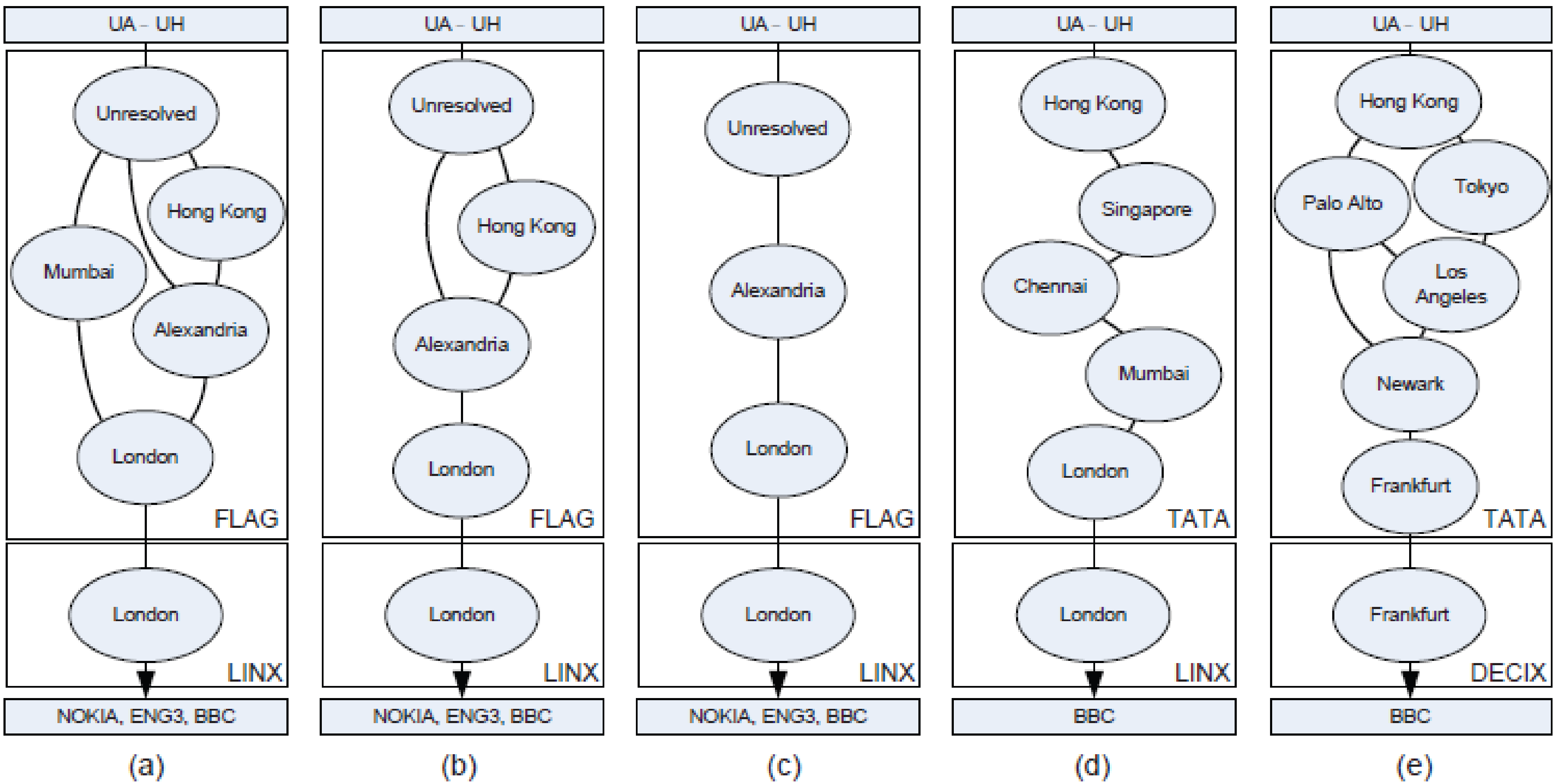


(a) Time series for the average IP-level Jaccard distance during e_1.



(b) RTT time series for the path between UB and BBC during e_1.

Traceroutes for phases (a)-(d)



Phases (a) and (b)

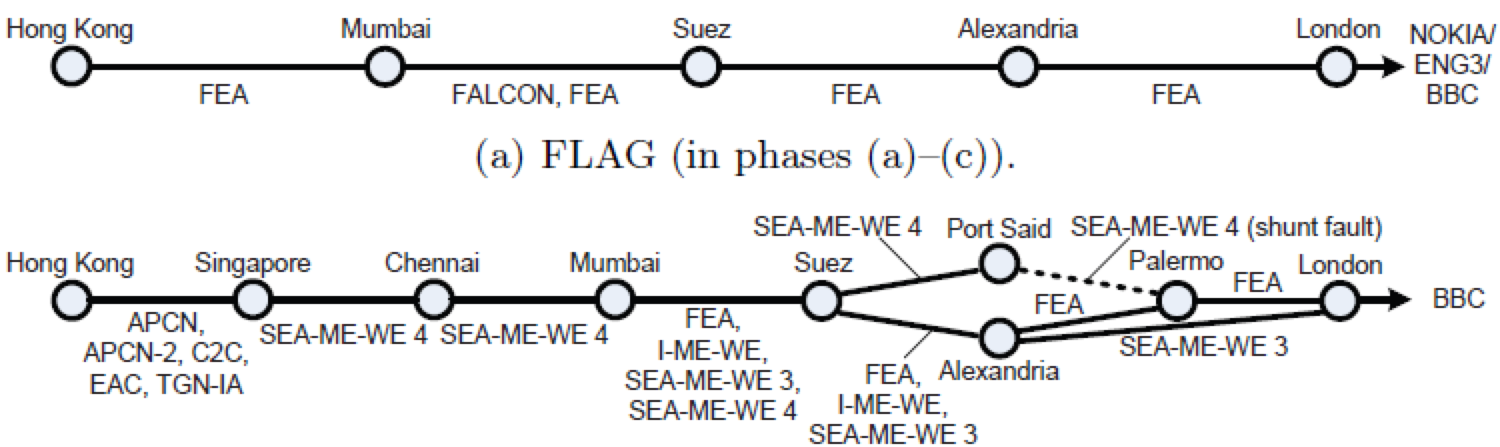
- In phase (a), three subpaths inside the FLAG network
- Upon the onset of phase (b) on 14 April 7:00 GMT (the same day of the reported cable fault), the IP-level Jaccard distance started declining
  - A result of the missing subpath via Mumbai in (b)
- The path UB⇋BBC (similarly for other BBC, NOKIA, and ENG3 paths) suffered from significant congestion in phase (b).
  - RTT inflation
  - Packet losses
- The path performance improved in the second half of the phase which corresponds to the non-working hours in the UK.

Phases (c) and (d)

- On the onset of phase (c), there is a further reduction in the IP-level Jaccard distance for the ENG3 and BBC paths.
  - Only the subpath via Alexandria and London was retained in FLAG
- More prolonged RTT inflation and packet losses in phase (c)
- At the beginning of phase (d), the upstream provider changed from FLAG to TATA (AS6453) only for the BBC paths
  - The IP-level Jaccard distance dropped to almost zero
  - A significant improvement of the BBC path's performance
  - The NOKIA and ENG3 paths still suffered from severe congestion in this phase

Submarine cables available to FLAG and TATA

- FLAG does not use the SEA-ME-WE 4 cable for Hong Kong → NOKIA, ENG3, and BBC
- FLAG uses FEA for Hong Kong → NOKIA, ENG3, and BBC
- TATA uses different cables between Mumbai and London.



(a) FLAG (in phases (a)–(c)).

(b) TATA (in phase (d)).

A plausible explanation for the network congestion

- The congestion in the FLAG network in phase (b) is taking on rerouted traffic from the SEA-ME-WE 4 cable after the cable fault.
  - FLAG uses FEA for all three websites.
  - TATA could have used FEA to reach BBC when the SEA-ME-WE 4 segment in the Mediterranean region was not available.
  - The BBC path performance improved after switching to TATA which could have used other cables, e.g., SEA-ME-WE 3.
- Another factor is the reduced path diversity from phase (a) to phase (c).

Conclusions and future work

- Both path-quality measurement and route tracing are important for diagnosing network problems.
  - Route tracing alone may not capture the path-quality degradation as a result of secondary effects
- Route changes could be responses to path-quality degradation.
  - But an ad hoc rerouting may cause congestion to other network paths.
  - However, some operators may not respond to the quality degradation
- We will report our impact analysis of other submarine cable faults, such as a SEACOM cable fault in Africa in July 2010.

Thanks