All Features Generated by FeaturesGen.py (Cells with green background are features used by Cambridge. Cells with box checked are features I plan to use and are likely to be changed in the future :))

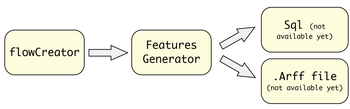
☐	Protocol (TCP/UDP)	int
☐	Client ip	
☑	Client port	int
☐	Server ip	
☑	Server port	int
☐	min IAT	float
☐	max IAT	float
☑	mean IAT	float
☐	median IAT	float
☑	var IAT	float
☐	min Ether size	int
☐	max Ether size	int
☐	mean Ether size	int
☐	median Ether size	int
☑	var Ether size	float
☐	min IP size	int
☐	max IP size	int
☐	mean IP size	int
☑	median IP size	int
☐	var IP size	float
☐	min TCP size	int
☐	max TCP size	int
☐	mean TCP size	int
☐	median TCP size	int
☐	var TCP size	float
☐	min UDP size	int
☐	max UDP size	int
☐	mean UDP size	int
☐	median UDP size	int
☐	var UDP size	float
☑	First 4 pkts direction	string
☑	First 4 pkts size #1	int
☑	First 4 pkts size #2	int
☑	First 4 pkts size #3	int
☑	First 4 pkts size #4	int
☐	min Ether size a b	int
☐	max Ether size a b	int
☐	mean Ether size a b	int
☐	median Ether size a b	int
☐	var Ether size a b	float
☐	min Ether size b a	int
☐	max Ether size b a	int
☐	mean Ether size b a	int
☐	median Ether size b a	int
☐	var Ether size b a	float
☐	min IP size a b	int
☐	max IP size a b	int
☐	mean IP size a b	int
☐	median IP size a b	int
☐	var IP size a b	float
☐	min IP size b a	int
☐	max IP size b a	int
☐	mean IP size b a	int
☐	median IP size b a	int
☐	var IP size b a	float
☑	min TCP size a b	int
☐	max TCP size a b	int
☑	mean TCP size a b	int
☐	median TCP size a b	int
☐	var TCP size a b	float
☑	min TCP size b a	int
☐	max TCP size b a	int
☑	mean TCP size b a	int
☐	median TCP size b a	int
☐	var TCP size b a	float
☐	min UDP size a b	int
☐	max UDP size a b	int
☐	mean UDP size a b	int
☐	median UDP size a b	int
☐	var UDP size a b	float
☐	min UDP size b a	int
☐	max UDP size b a	int
☐	mean UDP size b a	int
☐	median UDP size b a	int
☐	var UDP size b a	float
☐	mean IAT a b	int
☐	var IAT a b	float
☐	mean IAT b a	int
☐	var IAT b a	float
☐	total pkts a b	int
☐	total pkts b a	int
☐	ack pkts sent a b	int
☐	ack pkts sent b a	int
☐	pure acks sent a b	int
☐	pure acks sent b a	int
☐	sack pkts sent a b	int
☐	sack pkts sent b a	int
☐	dsack pkts sent a b	int
☐	dsack pkts sent b a	int
☐	max sack blks a b	int
☐	max sack blks b a	int
☐	unique bytes sent a b	int
☐	unique bytes sent b a	int
☑	actual data pkts a b	int
☑	actual data pkts b a	int
☐	rexmt data pkts a b	int
☐	rexmt data pkts b a	int
☐	zwnd probe pkts a b	int
☐	zwnd probe pkts b a	int
☐	zwnd probe bytes a b	int
☐	zwnd probe bytes b a	int
☐	outoforder pkts a b	int
☐	outoforder pkts b a	int
☑	pushed data pkts a b	int
☑	pushed data pkts b a	int
☐	SYN/FIN pkts sent a b	int
☐	SYN/FIN pkts sent b a	int
☐	req sack a b	int
☐	req sack b a	int
☐	sacks sent a b	int
☐	sacks sent b a	int
☐	urgent data pkts a b	int
☐	urgent data pkts b a	int
☐	mss request a b	int
☐	mss request b a	int
☐	max segm size a b	int
☐	max segm size b a	int
☐	min segm size a b	int
☐	min segm size b a	int
☐	avg segm size a b	int
☐	avg segm size b a	int
☐	max win adv a b	int
☐	max win adv b a	int
☐	min win adv a b	int
☐	min win adv b a	int
☐	zero win adv a b	int
☐	zero win adv b a	int
☐	avg win adv a b	int
☐	avg win adv b a	int
☑	initial window bytes a b	int
☑	initial window bytes b a	int
☐	initial window pkts a b	int
☐	initial window pkts b a	int
☐	ttl stream length a b	int
☐	ttl stream length b a	int
☐	missed data a b	int
☐	missed data b a	int
☐	truncated data a b	int
☐	truncated data b a	int
☐	truncated packets a b	int
☐	truncated packets b a	int
☐	data xmit time: a b	float
☐	data xmit time: b a	float
☐	idletime max a b	float
☐	idletime max b a	float
☐	throughput: a b	int
☐	throughput: b a	int
☑	RTT samples a b	int
☑	RTT samples b a	int
☐	RTT min: a b	float
☐	RTT min: b a	float
☐	RTT max a b	float
☐	RTT max b a	float
☐	RTT avg a b	float
☐	RTT avg b a	float
☐	RTT stdev a b	float
☐	RTT stdev b a	float
☐	RTT from 3WHS a b	float
☐	RTT from 3WHS b a	float
☐	RTT full sz smpls a b	int
☐	RTT full sz smpls b a	int
☐	RTT full sz min a b	float
☐	RTT full sz min b a	float
☐	RTT full sz max a b	float
☐	RTT full sz max b a	float
☐	RTT full sz avg a b	float
☐	RTT full sz avg b a	float
☐	RTT full sz stdev a b	float
☐	RTT full sz stdev b a	float
☐	post-loss acks a b	int
☐	post-loss acks b a	int
☐	segs cum acked a b	int
☐	segs cum acked b a	int
☐	duplicate acks a b	int
☐	duplicate acks b a	int
☐	max # retrans a b	int
☐	max # retrans b a	int
☐	min retr time a b	float
☐	min retr time b a	float
☐	max retr time a b	float
☐	max retr time b a	float
☐	avg retr time a b	float
☐	avg retr time b a	float
☐	sdv retr time a b	float
☐	sdv retr time b a	float

PREPARATION:

The following utilities need to be installed on your machine:

- python 2 (with pylibpcap module),
- perl,
- tcpdump,
- tcptrace.

You also need to set your path to the directory as it contains the necessary utility scripts (like tcpdemux) which aid the attribute generator script, which will make the utility listed above accessible system wide.



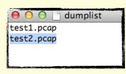
USAGE:

1) flowCreator

(thanks to BRAZIL computer lab's work from University of Cambridge)

Firstly you need to create a list which contains the tcpdump files. This is then passed to the flowCreator to reassemble the flows and outputs 'filelist' and out directory containing the reassembled flows.

```
perl flowCreator dumplist
```



2) Features Generator

The filelist created by flowCreator is passed to the Features Generator to calculate the features. This outputs Features_all that contains all the features listed on the left side.

```
python FeaturesGen.py filelist
```

3) SQL

4) Arff file creater:

I' these two parts later .