Jie He, Yue-xiang Yang, Yong Qiao, Wen-ping Deng, 2015. Fine-grained P2P traffic classification by simply counting flows. *Frontiers of Information Technology & Electronic Engineering*, **16**(5):391-403. [doi:10.1631/FITEE.1400267]

Fine-grained P2P traffic classification by simply counting flows

Key words: Traffic classification, Peer-to-peer (P2P), Fine-grained, Host-based

Contact: Jie He
E-mail: hejie@nudt.edu.cn

ORCID: http://orcid.org/0000-0003-2244-7594

Motivation

- The increasing P2P traffic leads to many issues with bandwidth, security, and management.
- For better management and control of P2P traffic, it is vital to classify P2P traffic accurately.
- A novel approach for accurate P2P traffic classification at a fine-grained level has been presented in this work.

Classification methodology

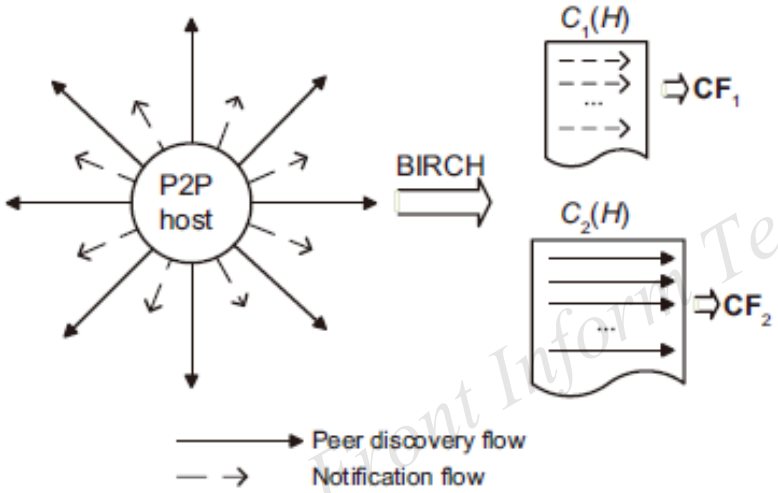

Fig. 2 Flow clustering for P2P host

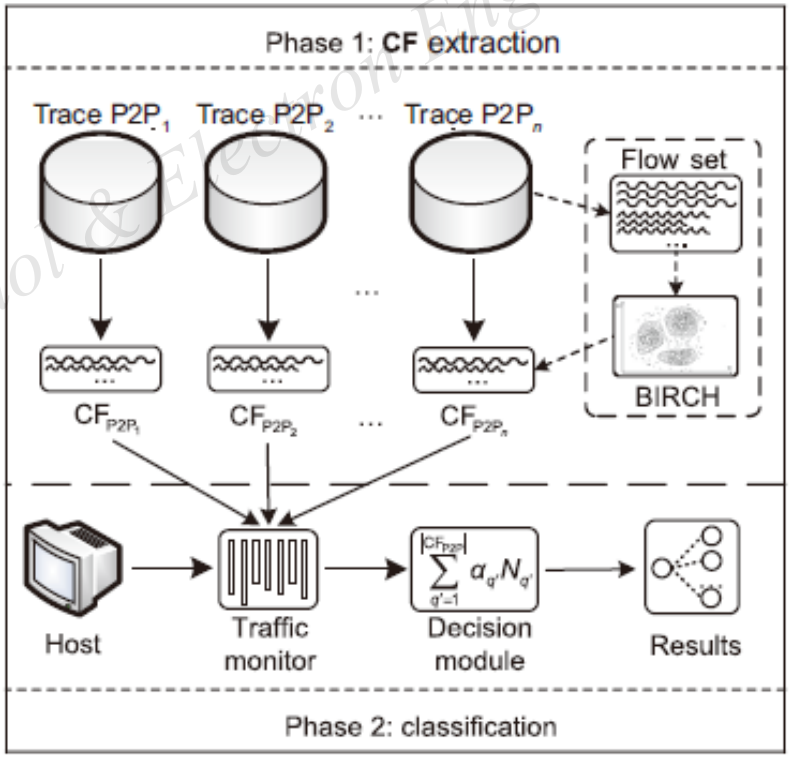

Fig. 3 Workflow of the classification mechanism

Experimental results (trace datasets)

Table 1 Description of training traces

Trace	Duration (d)	Number of packets ($\times 10^6$)	Sum of flows ($\times 10^6$)	Sum of successful UDP flows ($\times 10^3$)	Sum of successful TCP flows ($\times 10^3$)
BC	7	610.73	5.21	2873.00	44.20
BT	7	1397.95	1.67	250.27	116.98
EM	7	1398.33	0.35	105.96	114.82
VG	7	631.39	0.23	55.37	42.24
TD	7	629.36	2.33	144.93	157.51

BC: BitComet; BT: BitTorrent; EM: eMule, VG: Vagaa; TD: Thunder

Table 2 Description of real-world traces

App.	Sum of hosts	Active time (h)	Sum of flows ($\times 10^6$)	Sum of successful UDP flows ($\times 10^3$)	Sum of successful TCP flows ($\times 10^3$)
BC	8	48	1.47	666.75	12.48
BT	2	14	0.12	19.24	6.06
EM	21	84	0.34	84.77	105.23
VG	13	104	0.25	43.34	37.15
TD	56	840	11.14	1424.24	1114.24
BG	215	1720	32.85	9173.65	15552.74

App.: application. BC: BitComet; BT: BitTorrent; EM: eMule; VG: Vagaa; TD: Thunder; BG: background

Experimental results

Table 6 Confusion matrix of classification results

	T.W	BC	BT	EM	VG	TD	BG
BC	48	**100%**	0	0	0	12.5%	0
BT	14	0	**100%**	0	0	0	0
EM	84	0	0	**100%**	5.95%	0	0
VG	104	0	0	11.54%	**99.04%**	0	0.96%
TD	840	0	0	0	0	**100%**	0
FPR	–	0	0	10.35%	5.67%	11.11%	–

T.W: total number of time windows. BC: BitComet; BT: BitTorrent; EM: eMule; VG: Vagaa; TD: Thunder; BG: background. FPR: false positive rate. Labels on the rows (except the last) represent the ground truth, and labels on the columns (except the second) represent the classification results. Bold values represent the true positive rate (TPR) of the corresponding application

Conclusions

- We classify P2P applications by simply counting some special flows. No additional complicated information except several generic properties of flows is needed.
- This approach can work well with hosts in a complex network context, while most existing host-based approaches cannot deal with this situation.
- The experimental results show that we can classify P2P file-sharing applications with a TPR higher than 97.22% and a FPR lower than 2.78%.