Detecting P2P bots by mining the regional periodicity

基于区域周期性挖掘的P2P僵尸主机检测方法

Citation: Yong Qiao, Yue-xiang Yang, Jie He, Chuan Tang, Ying-zhi Zeng. 2013. Detecting P2P bots by mining the regional periodicity. *Journal of Zhejiang University-Science C (Computers & Electronics)*, 14(9):682-700. [doi:10.1631/jzus.C1300053]

- In this paper, we have proposed a novel detection model called *DMRP* to detect P2P botnets. Unlike the existing methods using similarity of network behaviors, or classifying techniques based on machine learning methods, our detection model is based on the nature of P2P botnets, sending requests periodically for updating their peer lists or receiving the commands from botmasters in the C&C phase. Therefore, we identify P2P botnet traffic by analyzing the specific periodicity within traffic datasets.
- During the process to mine for hidden periodicity, we introduce a novel concept of *regional periodic pattern mining* in time series, which is novel to P2P botnet detection, and present three algorithms to solve this problem. The *brute-force algorithm* is proposed for only theoretical analysis; the *apriori-like algorithm* and *RBI algorithm* are very efficient in practical applications with a low computational cost .
- Experimental evaluation based on public datasets shows that our proposed algorithms are suitable for mining the hidden regional periodicity in time series and that the DMRP model can detect P2P botnet traffic easily and correctly.
- However, as our P2P botnet detection model is built on the basic nature of P2P protocols in the C&C phase, it is not able to detect P2P botnets in the attack phase or other types of botnets with different communication protocols, such as IRC/HTTP.

P2P bot detection model

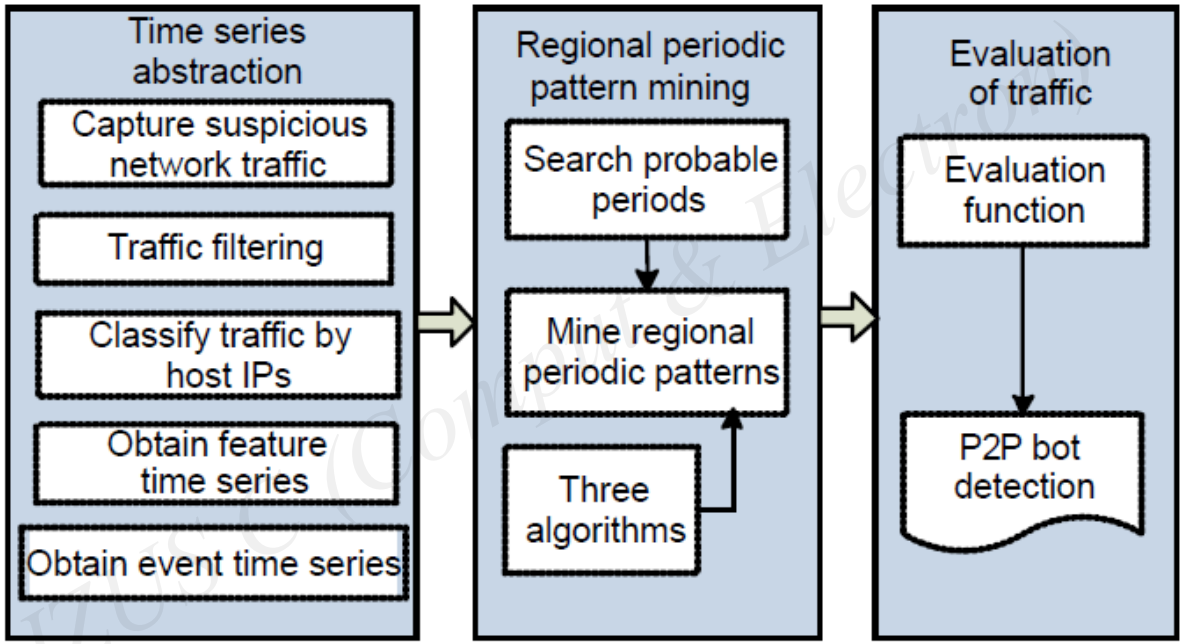

Fig. 1 P2P bot detection model: DMRP (detection by mining regional periodicity)

Extracting candidate periods by autocorrelation analysis

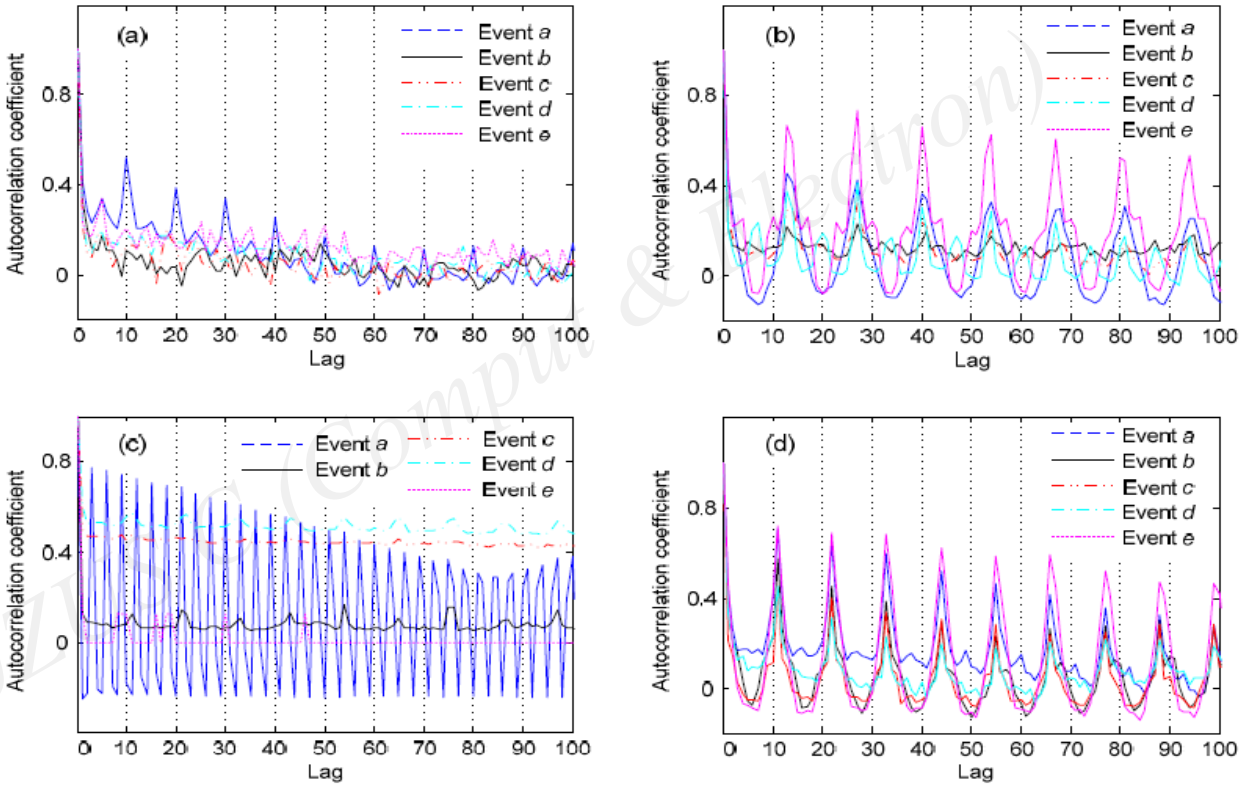

Fig. 4 Autocorrelation analysis of four sub-datasets of the first dataset (only the part of lag≤100 is shown)
(a) 172.16.0.11; (b) 172.16.0.12; (c) 172.16.2.2; (d) 172.16.2.11

Regional periodic patterns mining

Table 4 Results of regional periodic pattern mining

IP	L_length	Number of patterns	
		Brute-force (apriori-like)	RBI
172.16.0.11 (p=10)	1	2	2
172.16.0.11 (p=4)	1	0	0
172.16.0.12	1	46	46
	2	84	60
	3	84	69
	4	43	37
	5	8	7
172.16.2.2	1	83	83
	2	48	3
	3	6	1
172.16.2.11	1	29	29
	2	48	45
	3	47	46
	4	30	30
	5	12	12
	6	2	2
172.26.75.116	1	31	31
	2	46	45
	3	30	30
	4	8	8
	5	1	1

Verification of mining results

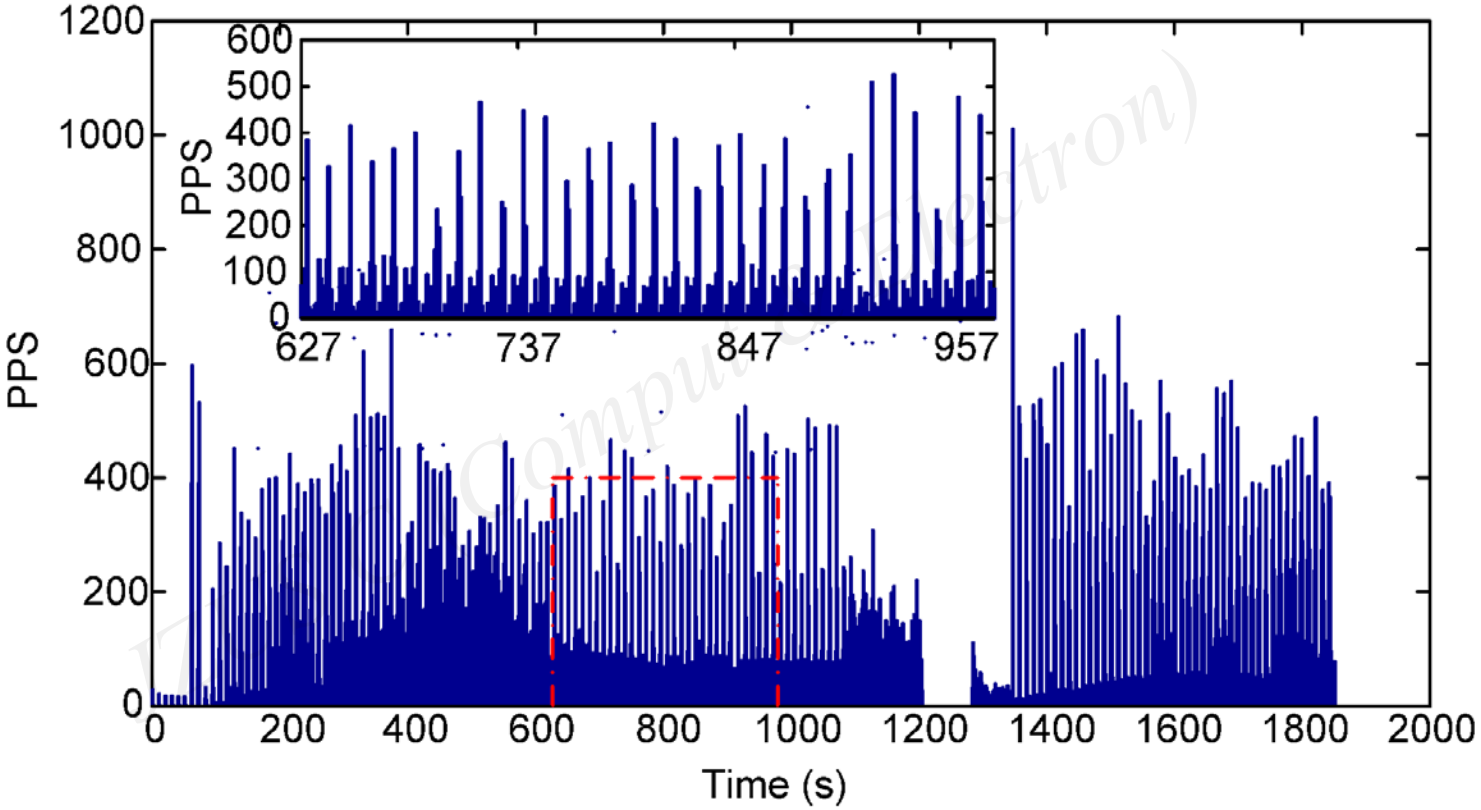

Fig. 6 The time series of the number of packets per second (PPS) for IP 172.16.2.11