TABLE 1: Turnaround Time of INSTPRO on each attack case (in seconds).

Attack Case	Data Collection	Data Preprocessing	Clue Extraction	Clue Correlation	Scenario Recovery	Total
K1	300	561.45	13.54	828.32	21.24	1724.56
K2	300	550.72	13.36	841.50	18.58	1724.17
K3	300	543.02	13.37	823.70	18.58	1698.68
K4	300	591.15	14.12	657.81	11.35	1574.43
K5	600	974.20	23.50	1712.06	10.06	3319.82
K6	300	330.34	8.11	311.56	9.14	959.15
K7	300	560.08	13.24	597.85	3.83	1474.99
K8	600	1010.53	22.43	1897.83	12.41	3543.19
K9	600	1068.75	25.97	3049.35	34.12	4778.18
K10	300	560.56	14.07	837.51	10.44	1722.57
K11	300	557.19	13.12	806.08	14.62	1691.00
K12	600	1055.78	26.20	2440.05	22.86	4144.89
K13	300	333.97	8.51	326.90	10.93	980.31
K14	300	327.31	8.02	310.35	7.73	953.40
K15	600	1042.54	24.29	1857.44	6.52	3530.80
L1	500	847.10	25.55	2292.88	22.42	3687.97
L2	143	299.59	7.59	300.01	0.87	751.06
L3	27	47.82	1.78	33.03	2.32	111.95
L4	136	301.19	7.61	276.90	1.52	723.24
L5	134	309.37	8.09	360.60	12.05	824.12
O1	141	197.78	3.58	371.72	1.47	715.55
O2	500	887.48	20.33	737.08	1.24	2146.13
O3	500	850.22	16.71	2116.03	3.88	3485.84
O4	500	945.92	19.27	2948.37	7.04	4420.60

1 TURNAROUND TIME PERFORMANCE OF INSTPRO ON EACH ATTACK CASE

To understand the turnaround time performance of INSTPRO, we measure times of data collection, data preprocessing, clue extraction, clue correlation and scenario recovery for the attack cases $K1 \sim K15$ and $L1 \sim L5$. Table 1 shows the turnaround time performance of INSTPRO. In the table, times of data collection for the attack cases $K1 \sim K15$ and $L1$ are fixed at specific values such as 300, 500 and 600 to intercept the initial rounds of the attacks while times of data collection for the remain attack cases are practical execution times. INSTPRO takes 47.82s to 1068.75s for data preprocessing, 1.78s to 26.20s for clue extraction, 33.03s to 3049.35s for clue correlation, 0.87s to 34.12s for scenario recovery, and 111.95s to 4778.18s for the total. We observe that (1) the clue extraction takes less time due to the highly efficient clue pattern matching algorithm; (2) the clue correlation takes up most of the time due to large amount of string matching among principled clues (i.e., instruction sequences); (3) the scenario recovery requires less time because of the small CPG sizes after clue correlation. In summary, the turnaround time performance of INSTPRO can be further improved by leveraging parallel and distributed computing techniques in its implementation.