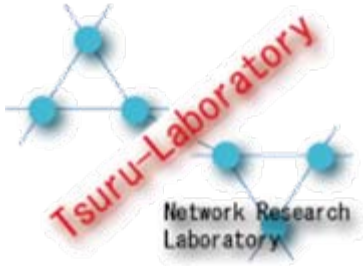


Unsupervised Ensemble Anomaly Detection through Time-Periodical Packet Sampling

Shuichi Nawata : Kyushu Institute of Technology
Masato Uchida : Kyushu Institute of Technology
Yu Gu : NEC Laboratories America
Masato Tsuru : Kyushu Institute of Technology
Yuji Oie : Kyushu Institute of Technology

Intrusion Detection Methods

- Signature detection
  - Find the traffic patterns that match the predefined set of attack signatures.

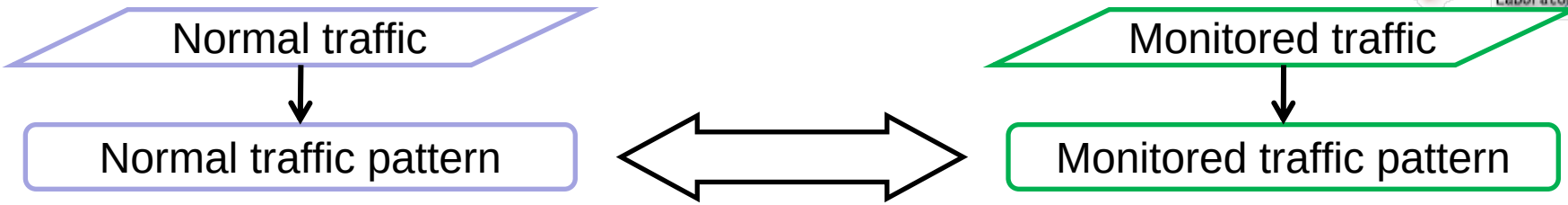
    •Efficient for known attack detection.
    •No alarm is raised for unknown attacks that are not contained in the signature database.

- Anomaly detection
  - Find the traffic patterns that deviate from the baseline model describing normal behavior of network traffic.

    •Useful for unknown attack detection.
    •Some degree of false alarm is inevitable.

The proposed method is categorized as anomaly detection.

Anomaly Detection

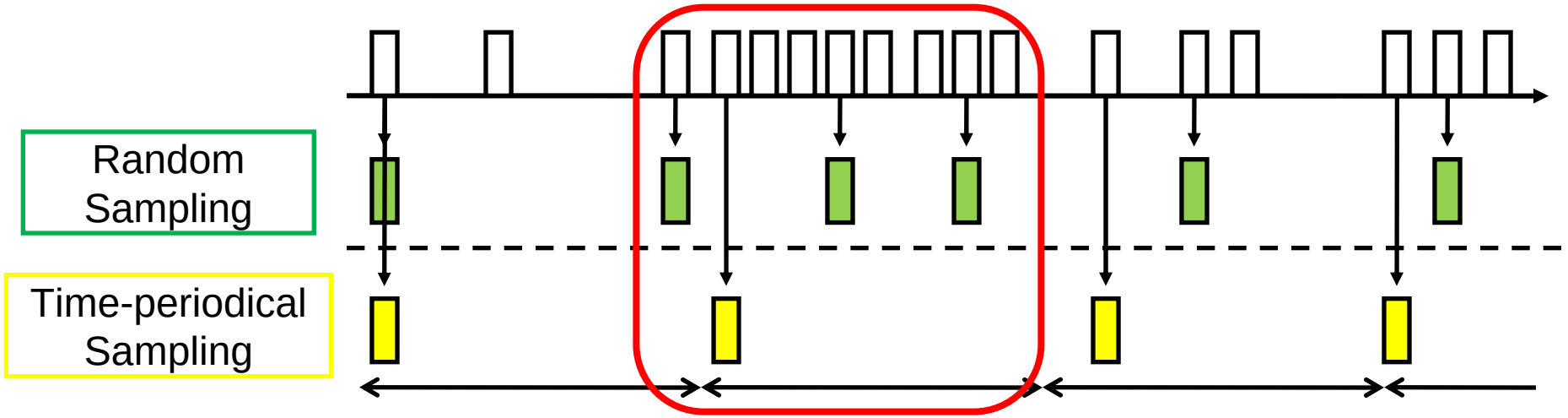


- Problem
  - Anomaly detection needs the classification of the traffic data into normal or anomalous packets to obtain normal traffic data which is used to build the normal traffic pattern model.
  - It is expensive and time-consuming because it is done by human experts.
- Our Approach
  - Take advantage of a drawback of packet sampling.
    - Packet sampling loses the original traffic characteristics, while it provides greater scalability for network measurements.

Sampled traffic data would be biased to normal packets by skipping anomalous packets.

Packet Sampling

- Target
  - Burst anomalous traffic regarding TCP SYN packets.*
  - Majority of today's significant operational threat.
- Idea
  - Apply time-periodical sampling to skip the periods in which burst anomalies occur.

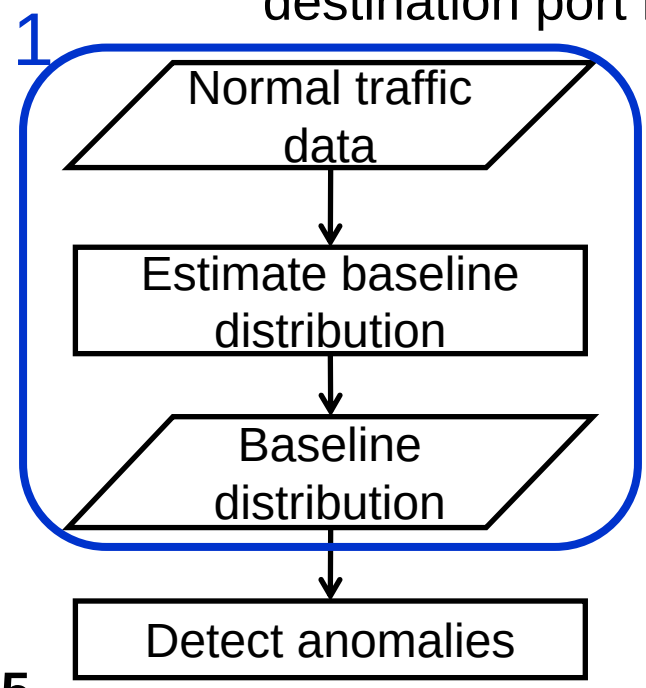


Mixing ratio of anomaly packets in the time-periodically sampled traffic will be low.

* "Worldwide Infrastructure Security Report", http://www.arbornetworks.com/report

Related Work

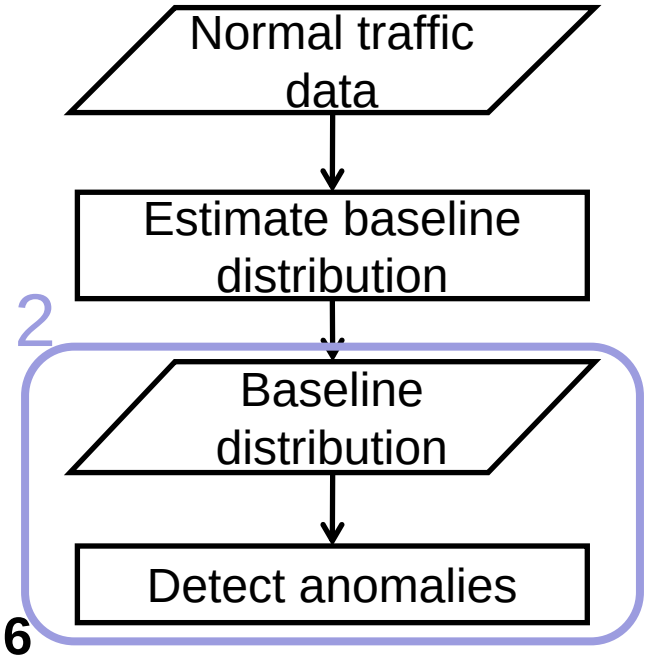
- Anomaly detection based on comparison between packet class distribution of audit traffic and baseline distribution.
  - Baseline distribution
    - describes normal traffic pattern.
  - Packet class
    - consists of 2348 classes according to the protocol and the destination port number.



1. Estimate baseline distribution from normal traffic data based on maximum entropy principle.
2. Detect anomalous traffic flow using sliding window detection approach.

Related Work

- Anomaly detection based on comparison between packet class distribution of audit traffic and baseline distribution.
  - Baseline distribution
    - describes normal traffic pattern.
  - Packet class
    - consists of 2348 classes according to the protocol and the destination port number.



1. Estimate baseline distribution from normal traffic data based on maximum entropy principle.
2. Detect anomalous traffic flow using sliding window detection approach.

Effective for detecting anomalies regarding TCP SYN packets.

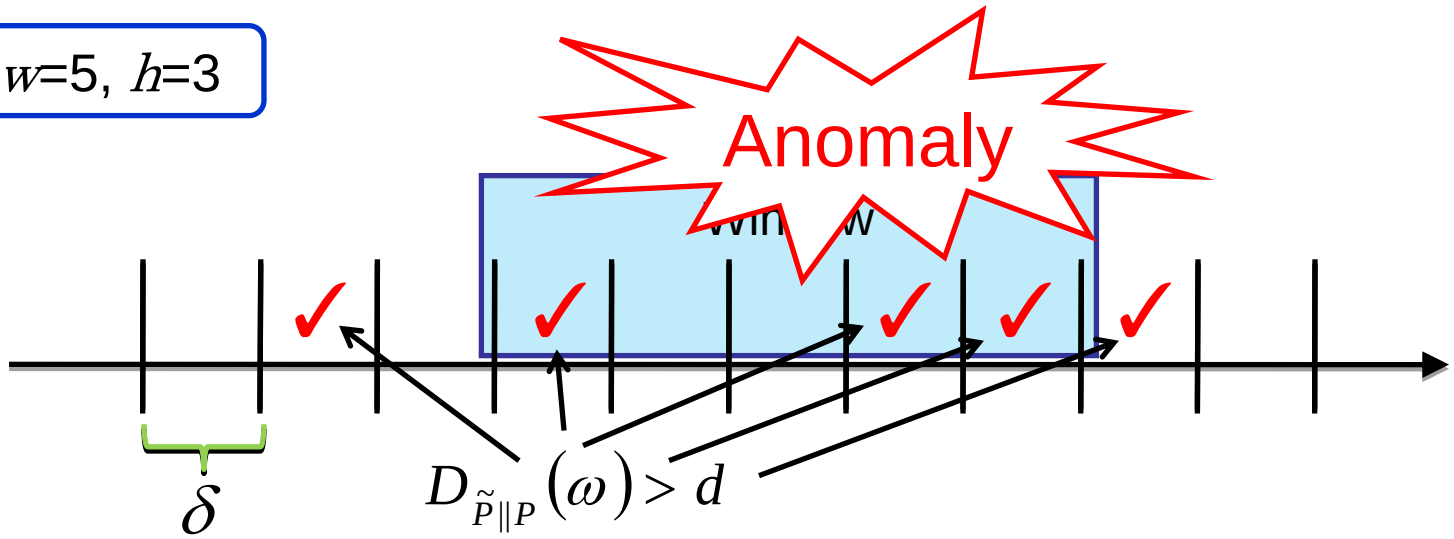
Sliding Window Detection Approach

1. The time is divided into slots with fixed length δ.
2. If for a packet class ω, $D_{\tilde{P}\|P}(\omega) > d$ holds for more than h times in a time window of w time slots, an alarm is raised.

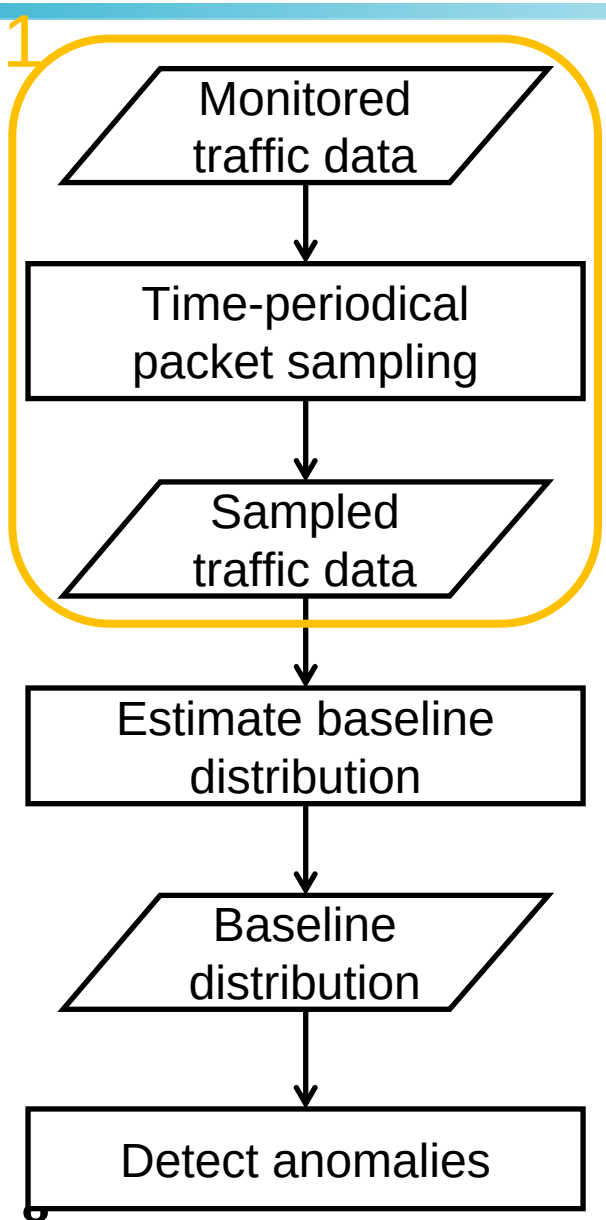
Partial relative entropy: $$D_{\tilde{P}\|P}(\omega) = \tilde{P}(\omega)\log\frac{\tilde{P}(\omega)}{P(\omega)}$$

$\tilde{P}$:Empirical distribution of monitored traffic.
P :Baseline distribution.

Case: w=5, h=3

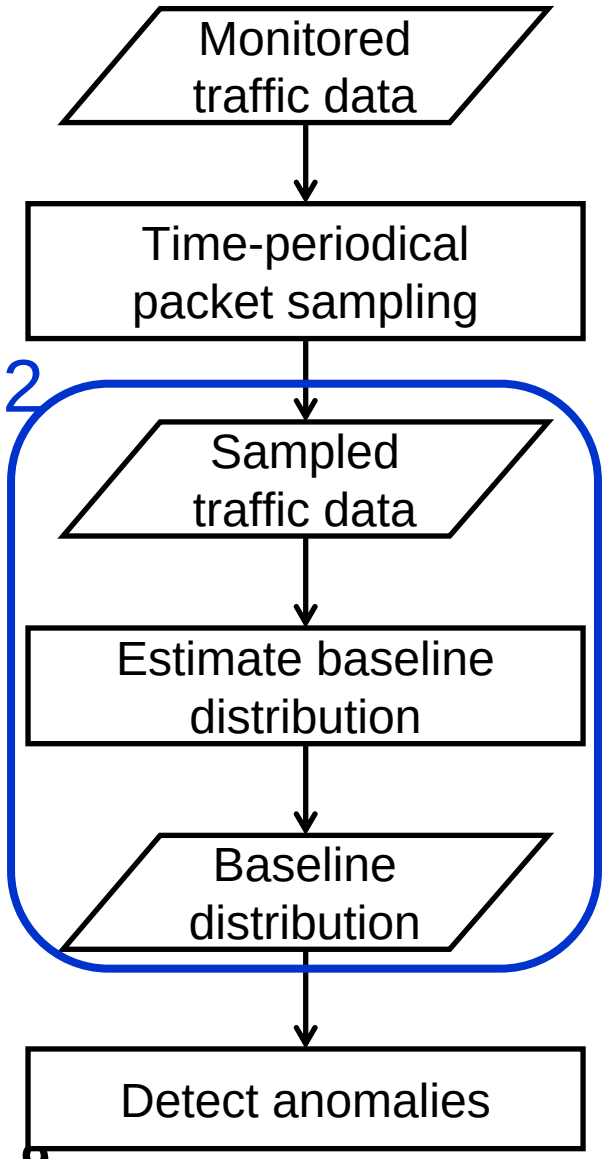


Proposed Method



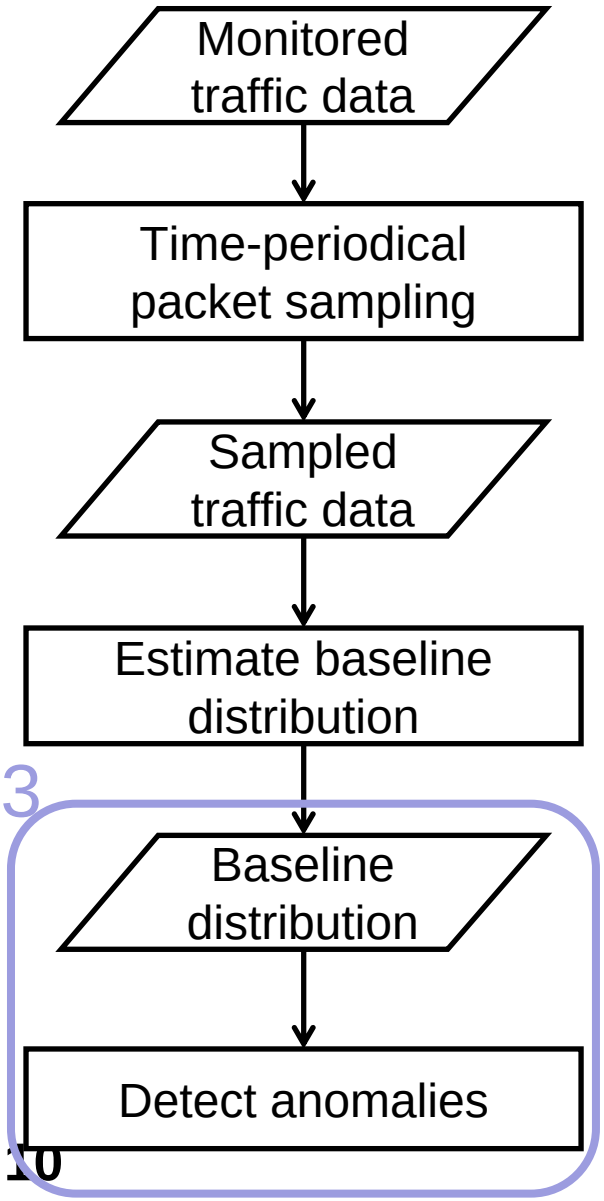
1. Extract traffic data of which mixing ratio of anomalous packets is low using time-periodical packet sampling.
   - Time-periodical sampling : A sampling intervals follow an identical exponential distribution with expectation τ.
2. Estimate baseline distribution from sampled traffic data based on maximum entropy principle.
3. Detect anomaly traffic using sliding window detection approach.
   - The anomaly detection performance of the baseline distribution varies from sample to sample.

Proposed Method



1. Extract traffic data of which mixing ratio of anomalous packets is low using time-periodical packet sampling.
   - Time-periodical sampling : A sampling intervals follow an identical exponential distribution with expectation τ.
2. Estimate baseline distribution from sampled traffic data based on maximum entropy principle.
3. Detect anomaly traffic using sliding window detection approach.
   - The anomaly detection performance of the baseline distribution varies from sample to sample.

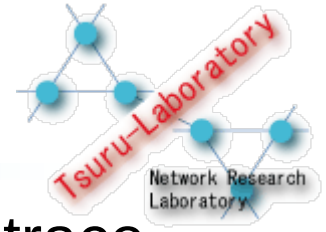
Proposed Method



1. Extract traffic data of which mixing ratio of anomalous packets is low using time-periodical packet sampling.
   - Time-periodical sampling : A sampling intervals follow an identical exponential distribution with expectation τ.
2. Estimate baseline distribution from sampled traffic data based on maximum entropy principle.
3. Detect anomaly traffic using sliding window detection approach.
   - The anomaly detection performance of the baseline distribution varies from sample to sample.

Detect anomalies using multiple baseline distributions

Evaluation with Real Traffic Data



- Evaluation of the proposed method using the traffic trace measured at the UMass Internet gateway router.
  - July 16 to July 22, 2004
  - 9:30 a.m. to 10:30 a.m.
- Detection of anomaly traffic using baseline distributions estimated from 4 kinds of traffic data.
  - Normal traffic data (extracted from original traffic data)
  - Original traffic data (before sampling)
  - Time-periodically sampled traffic data : 10 sets

    (The sampling intervals follow an independent and identical exponential distribution with expectation τ [sec]. τ =0.1, 0.01, 0.001.)
  - Randomly sampled traffic data : 10 sets

    (Sampling probability p = 0.001, 0.01, 0.1)

Effect of Time-Periodical Sampling

- Mixing ratios of anomaly packets [%].
  - Average value taken over 10 sets of sampled traffic data.

Date	Original traffic	Time-periodically sampled traffic (average)			Randomly sampled traffic (average)		
		0.1s	0.01s	0.001s	0.001	0.01	0.1
July 16(Fri)	8.48	4.20	4.32	4.67	8.43	8.47	8.48
July 17(Sat)	8.71	7.49	7.59	7.86	8.46	8.71	8.71
July 18(Sun)	18.18	14.68	14.80	15.35	18.31	18.19	18.18
July 19(Mon)	11.02	6.13	6.19	6.57	11.07	11.02	11.02
July 20(Tues)	8.36	3.50	3.50	3.79	8.37	8.36	8.36
July 21(Wed)	6.62	3.17	3.20	3.39	6.63	6.61	6.62
July 22(Thurs)	2.97	1.36	1.35	1.47	3.00	2.97	2.97

Mixing ratios of time-periodically sampled data is lower.

Effect of Time-Periodical Sampling

- Mixing ratios of anomaly packets [%].
  - Average value taken over 10 sets of sampled traffic data.

Date	Original traffic	Time-periodically sampled traffic (average)			Randomly sampled traffic (average)		
		0.1s	0.01s	0.001s	0.001	0.01	0.1
July 16(Fri)	8.48	4.20	4.32	4.67	8.43	8.47	8.48
July 17(Sat)	8.71	7.49	7.59	7.86	8.46	8.71	8.71
July 18(Sun)	18.18	14.68	14.80	15.35	18.31	18.19	18.18
July 19(Mon)	11.02	6.13	6.19	6.57	11.07	11.02	11.02
July 20(Tues)	8.36	3.50	3.50	3.79	8.37	8.36	8.36
July 21(Wed)	6.62	3.17	3.20	3.39	6.63	6.61	6.62
July 22(Thurs)	2.97	1.36	1.35	1.47	3.00	2.97	2.97

The time-periodical packet sampling is useful for extracting normal packets.

Anomaly Detection Performance

- Anomaly detection based on baseline distributions trained using the traffic data measured on July 20 (Tue).
  - The sum of FP and FN for the time-periodically sampled
  - traffic data is smaller than one for original traffic data and randomly sampled traffic data on all days.

Date	Normal traffic	Original traffic	Time-periodically sampled traffic			Randomly sampled traffic		
			0.1s	0.01s	0.001s	0.001	0.01	0.1
July 16(Fri)	1/0	1/5	~~2/0~~	1/0	1/0	1/5	1/5	1/5
July 17(Sat)	0/1	0/5	2/1	2/1	1/1	1/5	0/5	0/5
July 18(Sun)	0/0	0/2	0/0	0/0	0/0	0/2	0/2	0/2
July 19(Mon)	1/0	1/3	1/1	1/0	1/0	1/5	1/4	1/3
July 21(Wed)	0/0	0/9	0/2	0/2	0/2	0/9	0/9	0/9
July 22(Thurs)	0/1	0/4	0/3	0/3	0/3	0/4	0/4	0/4

The time-periodical packet sampling is useful for anomaly detection.

Anomaly Detection Performance

- Anomaly detection based on baseline distributions trained using the traffic data measured on July 20 (Tue).
  - The sum of FP and FN for the time-periodically sampled
  - traffic data is smaller than one for original traffic data and randomly sampled traffic data on all days.

Date	Normal traffic	Original traffic	Time-periodically sampled traffic			Randomly sampled traffic		
			0.1s	0.01s	0.001s	0.001	0.01	0.1
July 16(Fri)	1/0	1/5	2/0	1/0	1/0	1/5	1/5	1/5
July 17(Sat)	0/1	0/5	2/1	2/1	1/1	1/5	0/5	0/5
July 18(Sun)	0/0	0/2	0/0	0/0	0/0	0/2	0/2	0/2
July 19(Mon)	1/0	1/3	1/1	1/0	1/0	1/5	1/4	1/3
July 21(Wed)	0/0	0/9	0/2	0/2	0/2	0/9	0/9	0/9
July 22(Thurs)	0/1	0/4	0/3	0/3	0/3	0/4	0/4	0/4

The time-periodical packet sampling is useful for anomaly detection.

Individual Performance

- Anomaly detection performance based on individual baseline distribution trained using 10 sets of time-periodically sampled traffic data.
  - Traffic data measured on July 20.
  - Target traffic data measured on July 17.
  - Sampling intervals follow an identical exponential distribution with expectation 0.1.

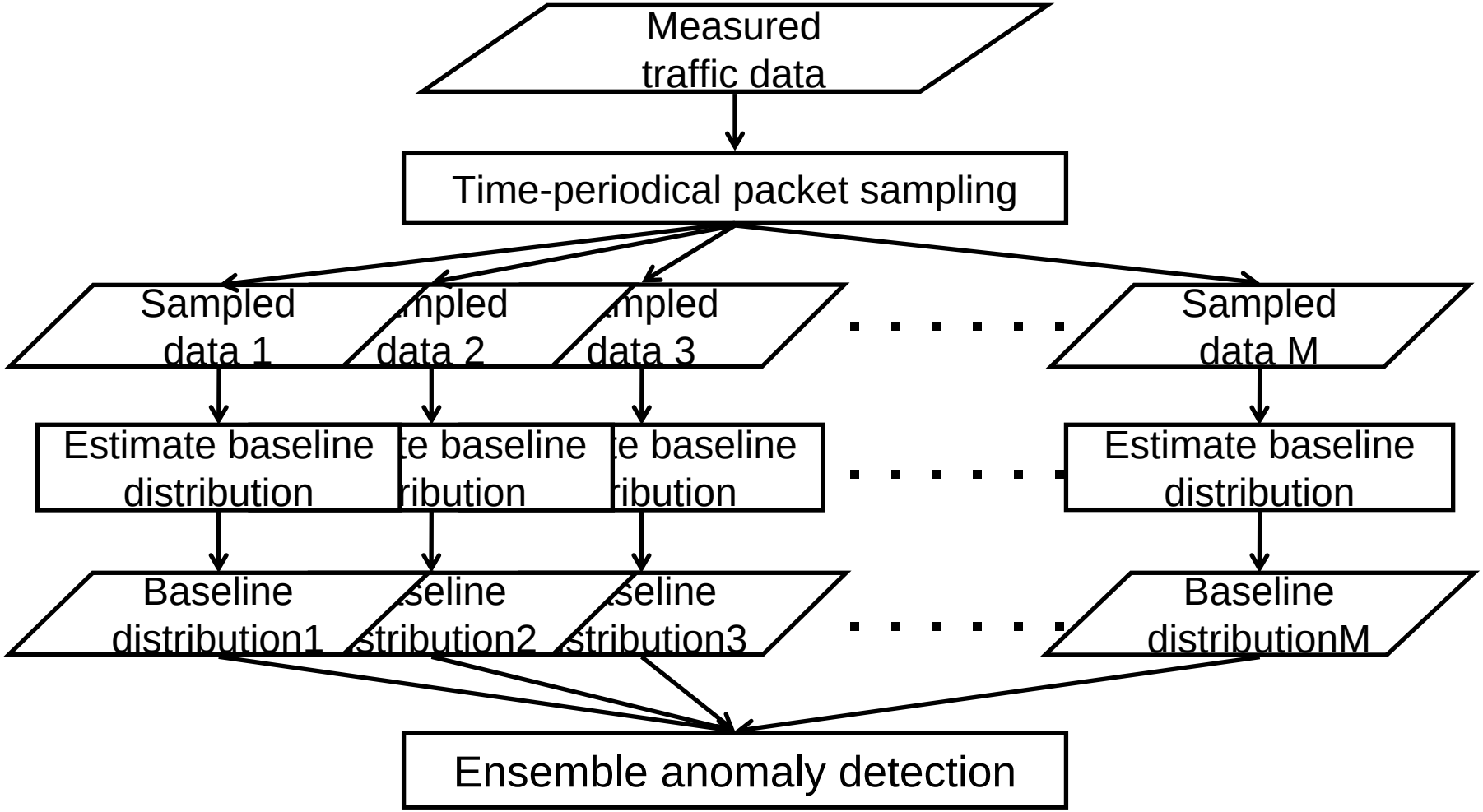
Trials	1	2	3	4	5	6	7	8	9	10
False Negative	1	1	1	1	1	1	1	1	1	1
False Positive	1	0	2	0	2	1	0	2	0	1

Anomaly detection performance varies from sample to sample.

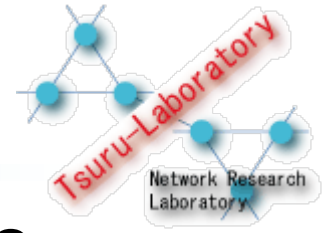
Integration of multiple baseline distributions in order to reduce the variation of the detection accuracy.

Ensemble Anomaly Detection

- Outline



Integration of Baseline Distributions



- Compare the average of 10 partial relative entropy with the threshold d.

$$\frac{1}{10}\sum_{i=1}^{10} D_{\tilde{P}\|P_i}(\omega) > d$$

$\underset{\sim}{P}_i$: i-th baseline distribution
$\tilde{P}$: The empirical distribution
$D_{\tilde{P}\|P_i}(\omega)$: The partial relative entropy

Ensemble Detection Performance

- The results for the individual baseline distributions (worst result) and the integrated baseline distribution trained using 10 sets of time-periodically sampled traffic data.
  - Traffic data measured on July 20.

Date	Normal traffic	Original traffic	Individual baseline distributions (worst result)			Integrated baseline distribution		
			0.1s	0.01s	0.001s	0.1s	0.01s	0.001s
July 16	1/0	1/5	2/0	~~1/0~~	~~1/0~~	**1**/0	1/0	1/0
July 17	0/1	0/5	2/1	2/1	1/1	**1**/1	**1**/1	**0**/1
July 18	0/0	0/2	0/0	0/0	0/0	0/0	0/0	0/0
July 19	1/0	1/3	1/1	~~1/0~~	~~1/0~~	1/**0**	1/0	1/0
July 21	0/0	0/9	0/2	0/2	0/2	0/2	0/2	0/2
July 22	0/1	0/4	0/3	~~0/3~~	~~0/3~~	0/**2**	0/3	0/3

Integration of baseline distributions is useful for improving the detection accuracy

Ensemble Detection Performance

- The results for the individual baseline distributions (worst result) and the integrated baseline distribution trained using 10 sets of time-periodically sampled traffic data.
  - Traffic data measured on July 20.

Date	Normal traffic	Original traffic	Individual baseline distributions (worst result)			Integrated baseline distribution		
			0.1s	0.01s	0.001s	0.1s	0.01s	0.001s
July 16	1/0	~~1/5~~	~~2/0~~	~~1/0~~	~~1/0~~	**1**/0	1/0	1/0
July 17	0/1	~~0/5~~	~~2/1~~	~~2/1~~	~~1/1~~	**1**/1	**1**/1	**0**/1
July 18	0/0	0/2	0/0	0/0	0/0	0/0	0/0	0/0
July 19	1/0	~~1/3~~	~~1/1~~	~~1/0~~	~~1/0~~	1/**0**	1/0	1/0
July 21	0/0	0/9	0/2	0/2	0/2	0/2	0/2	0/2
July 22	0/1	~~0/4~~	~~0/3~~	~~0/3~~	~~0/3~~	0/**2**	0/3	0/3

Integration of baseline distributions is useful for improving the detection accuracy

Conclusion

- Ensemble anomaly detection method that does not require human effort.
  - Employing packet sampling for a different purpose from which it was intended.
    - Time-periodical packet sampling can extract normal packets from original traffic which may include burst anomalous traffic.
    - Probabilistic variation of sampled data affects the detection performance.
  - Integrating multiple baseline distributions.
    - Ensemble method can improve overall performance in anomaly detection.

Proposed method is able to reduce the cost of human effort without sacrificing detection accuracy for detecting anomalies.