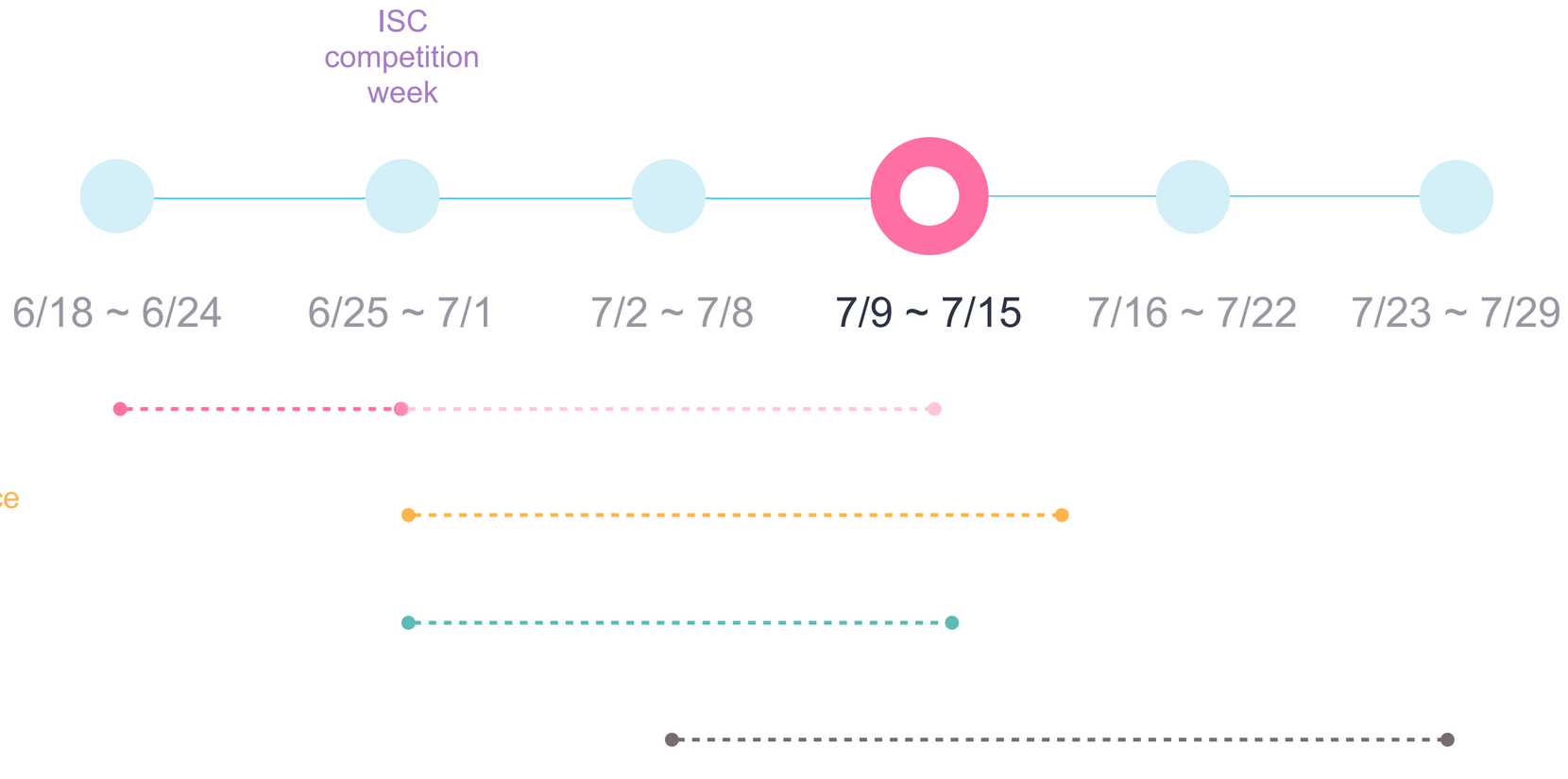
Task weekly **Report**

updated on **9** June 2018

Task plan each **Week**



Preprocess Data

Preprocess I **Grouping**

type 1

Description

Try grouping the attributes that there are categorical data

- IP address
- MAC address
- Port number
- MAC address type
- LLC type
- IP version

Preprocess I not Grouping

type 2

Description

Try to convert all of categorical data to be in vector format.

How

Encoding attribute by using library in keras.
Ref: https://machinelearningmastery.com/how-to-one-hot-encode-sequence-data-in-python/

Problem

MemError in python because of too many attributes.
Example: IP_src had more than 20,000 members that there are difference IP.
Not interest in relation with the time that packet arrival

Preprocess I **using window**

type 3

Description

find the relation of time by add new attributes

ADD - difference time between the first pair (IP_src & IP_dst) packet arrival and new packet arrival (same pair IP)

- sum of pair packets, ip source, ip destination
- sum of pair port, port source, port destination with the same packet
- add weight in the third highest ip & port that occur in each windows

Remove - IP_src, IP_dst, sport, dport

Problem may occurs

find the appropriate window size for calculate attribute

Experiment **Dataset**

Experiment with **SVM model**

- using pyKMLib that support GPU
- svmlight format (convert .csv to svmlight)

Experiment with **Deep Learning model**

- using Keras that support GPU

Summary Result

Preprocess	SVM	Deep Learning
Type 1	Training Time : 68.7851 sec Predict Time : 2969.1802 sec Accuracy : 0.5365	Training Time : 85.8437 sec Predict Time : 6.6769 sec Accuracy : 0.5105
Type 2 *reduce some attributes	Training Time : 61.8569 sec Predict Time : 3016.9244 sec Accuracy : 0.5069	Training Time : 72.9299 sec Predict Time : 4.0889 sec Accuracy : 0.5107
Type 3	waiting… preprocess data	waiting… preprocess data

*used only
[Status,Ether_or_Dot3,MAC_src,MAC_dst,Ether_type,LLC,LLC_ssap,LLC_dsap,IP_ttl,IP_version,TCP,UDP,ARP,ICMP,pLen]

Summary Tasks on this week

Preprocess Data

- [x] Preprocess set 1
- [x] Preprocess set 2
- [x] Preprocess set 3

Experiment evaluate model

- [x] Testing with Preprocess set 1
- [x] Testing with Preprocess set 2
- [ ] Testing with Preprocess set 3

Plan to do

- try training preprocess 3
- Evaluate model with MSE and confusion matrix
- Tuning model