JPCERT CC®

Mongolia’s Internet Health Visualized

- Risk Factors in Present and the Future

Yasutaka ISHII
Vulnerability Analyst
Early Watch Group
JPCERT Coordination Center

Background

- JPCERT/CC measures the current "healthiness" of cyber space and forecasts its future.
- **Mejiro**, an Internet risk visualization service is used in this project.
- JPCERT/CC's analysis and forecast on Mongolian cyber space will be presented today.

Agenda

1. Mejiro — the Internet risk visualization service

2. Analysis of risk trends in Mongolian cyber space

3. How to forecast and check the validity

4. Summary

1. Mejiro — the Internet risk visualization service

Mejiro

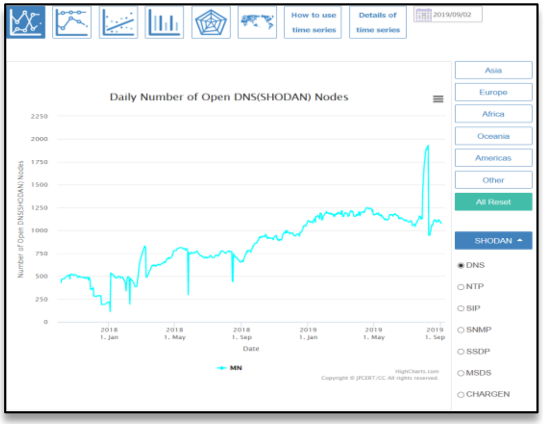
- Purpose of the service
  - Visualize risk factors that might suspend services on the Internet
  - Understand the current situation of risk factors based on the collected data

⊖ DNS (Open Resolver)
⊜ NTP
⊛ SIP
④ SNMP
⑤ SSDP
⑥ CHARGEN
⑦ microsoft-ds

- Data available
  - The number of Open UDP Server nodes scanned
  - Mejiro Index

Publicly available as a Demonstration Test

Internet Risk Visualization Service -Mejiro-
https://www.jpcert.or.jp/english/mejiro/

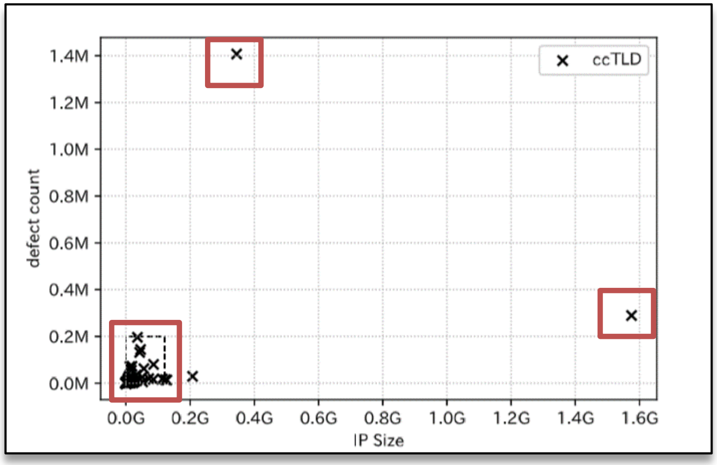


Mejiro Index

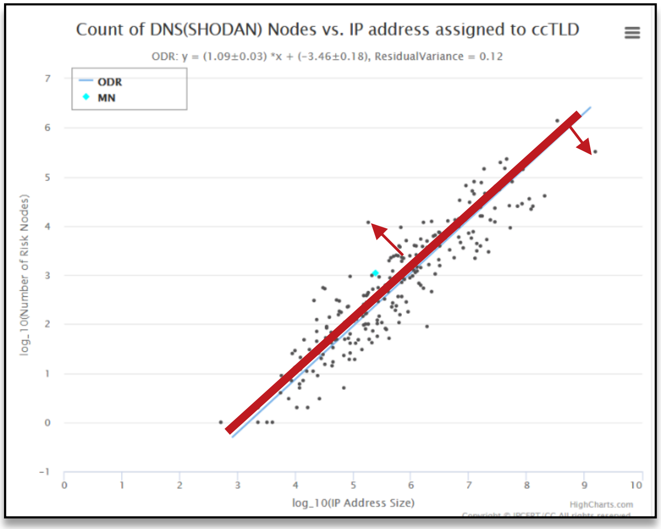
- Comparison by the absolute number of nodes cannot avoid the influence from the total number of IP addresses in each country.

- Mejiro Index for relative evaluation
  - The number of risk factors/IP addresses are logarithmically transformed.
  - Draw an ODR line and evaluate by the divergence from the standard

The lower, the better.



Logarithmically transform



Our next steps with Mejiro

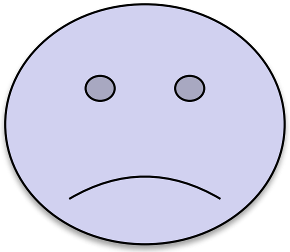
- What's next?
  - Mejiro succeeded in visualizing the **present** state.

➡ (1) Analyze the current risks and

(2) Forecast the **future**

- Purpose
  - Detect unusual activities and take actions more promptly
  - Increase the efficiency of cyber clean up activities

"How to" instructions are also required.



2. ANALYSIS OF RISK TRENDS IN MONGOLIAN CYBER SPACE

Current state of Mongolian cyber space

- What you can learn from time series data
  - Periodicity (recurring pattern seen in a shorter cycle)
  - Trend (Increase/decrease in a longer span of time)

Protocol	Avg. # of nodes (Apr – Aug)	Periodicity	Trend
DNS	1188.776	Not found	Moderate decrease
NTP	4315.63	Not found	Moderate increase
SIP	84.88	Monthly, Weekly	Flat
SNMP	982.21	Not found	Increase
SSDP	65.41	Weekly	Flat
MSDS	377.76	Not found	Flat

Selecting the target of analysis/forecasting

- Selection criteria
  - Amount of data (length/frequency of data collection)
  - Rate of unavailable data

The more data, the better!

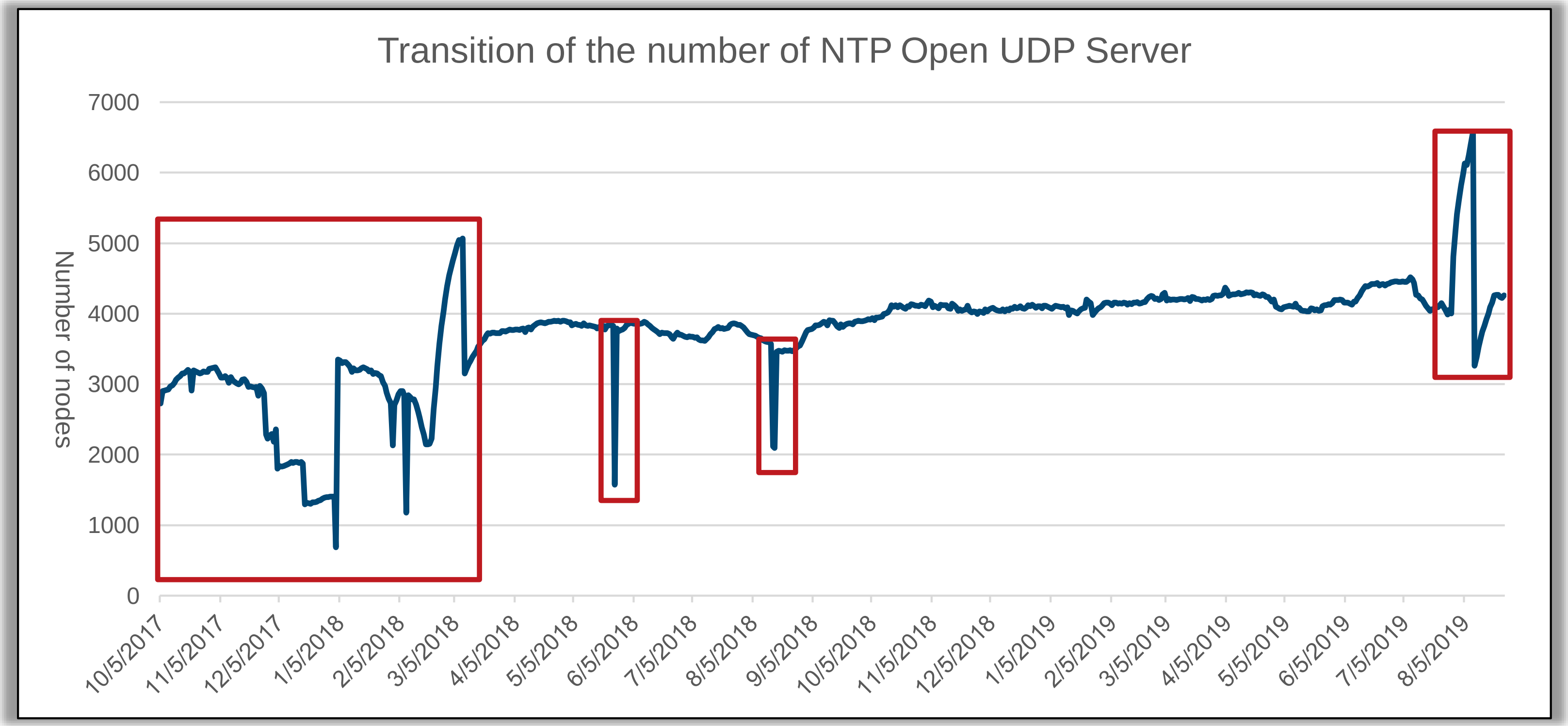
- Accumulated data (Oct 5, 2017 – Aug 20, 2019)

Data Provider	Protocol	Data collected since	Frequency	# of data
SHODAN	DNS,NTP,SIP, SNMP,SSDP,MSDS	**Oct 5, 2017**	Daily	**689**
	CHARGEN	Jan 14, 2019	Daily	166
censys	DNS	Oct 5, 2017	Daily	621
	SMB	Jan 14, 2019	Daily	215
CyberGreen	All of above	Jan 14, 2019	Weekly	33

Unusual change in Open UDP Server

- Dramatic change in the number of nodes
  - It is particularly outstanding in the first several months since the beginning of the record
  - The number skyrocketed by 50% in Jul 28-Aug 15, 2019

➞ Due to the change of Shodan’s scan method

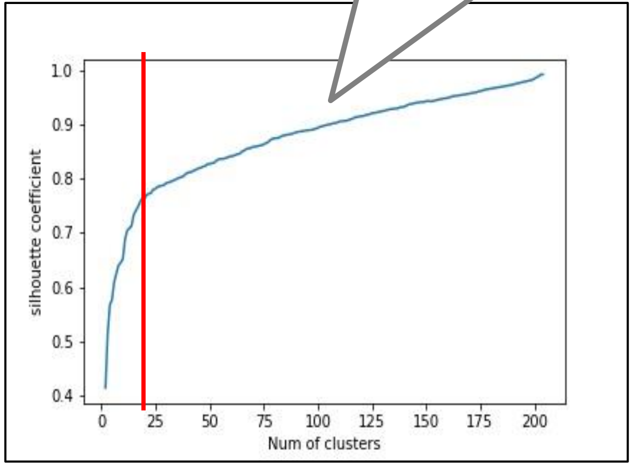


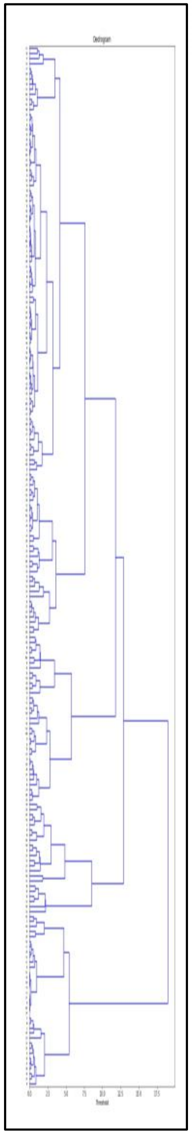
Clustering

- Clustering for comparison
  - Categorization by the normalized number of each protocol’s Open UDP Server nodes (April 2019 – July 2019)

- Evaluation of the results
  - Silhouette value: Distance between the sample and the adjacent cluster
  - 20 clusters
    Result: 0.764
    Clustering was successful to some extent.

Complete clustering is impossible because the silhouette value keeps increasing.

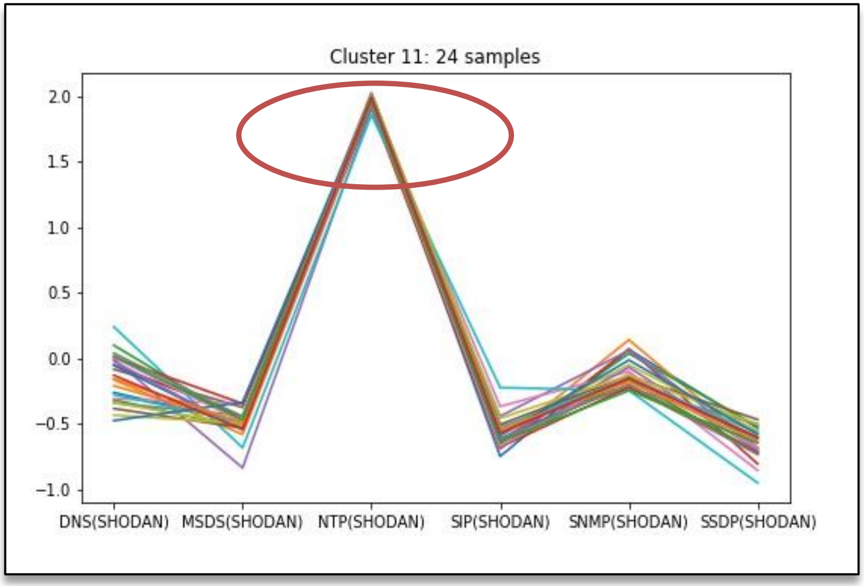


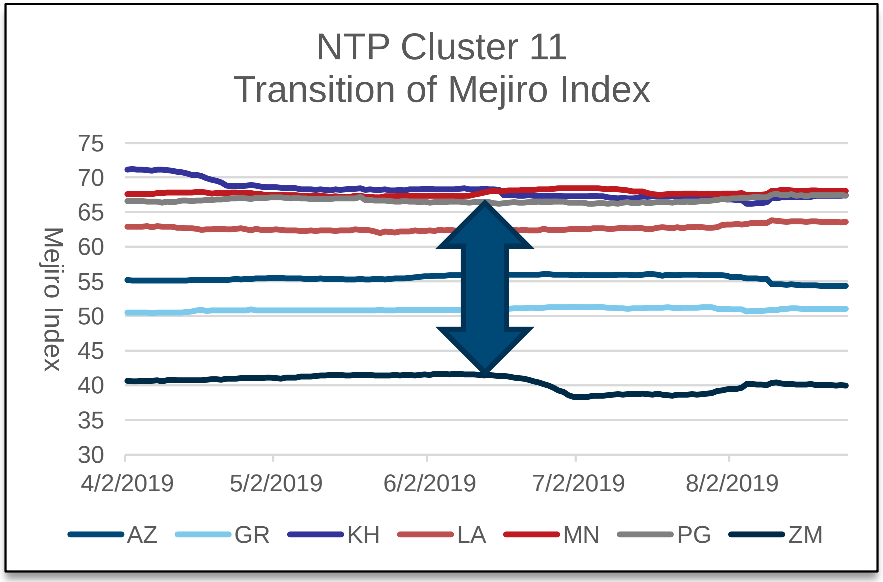


Clustering Result

- Findings
  - Mongolia is one of the 24 countries in the cluster in which only the number of NTP's Open UDP Server nodes is higher than other protocols.
  - Mejiro index differ even within a cluster.

→ Clues to further research/actions





3. HOW TO FORECAST THE TRENDS AND CHECK ITS VALIDITY

Target of forecast

- Mongolia’s Mejiro Index for each protocol

  1. DNS(SHODAN)
  2. NTP (SHODAN)
  3. SIP (SHODAN)
  4. SNMP (SHODAN)
  5. SSDP (SHODAN)
  6. MSDS (SHODAN)

- Mejiro Index requires…

  1. The number of Open UDP Server nodes for each protocol
  2. Total number of IP address
  3. ODR line parameter for each protocol

#1 (Mongolia) and #3 are the target of forecast today.

How to check the validity of the forecast

- Validation method
  - Mean absolute percentage error (MAPE)

$MAPE = \frac{100}{n} \sum_{i=0}^{n} \left| \frac{f_i - y_i}{y_i} \right|$ f = forecasted value, y = actual record

- Errors are represented in ratio.
- The smaller the number is, the more suitable the model is.
- MAPE tends to be high when the target of validation is small.

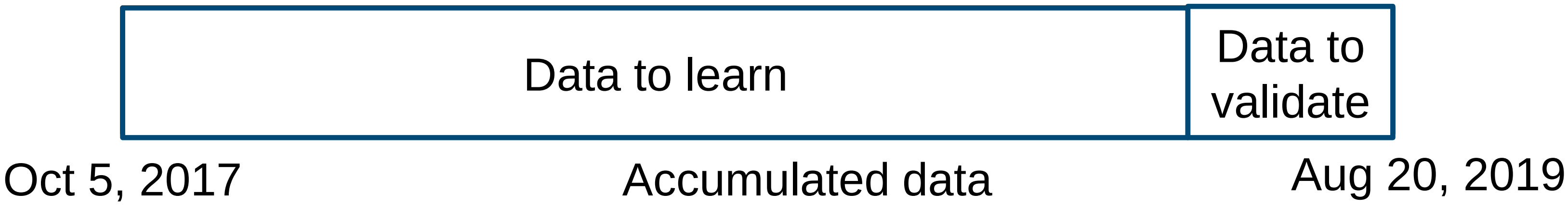
➡ MAPE shows the relative accuracy, and thus it is easy to compare different countries.

Example: Actual value = 100, Forecast = 150, MAPE=50%

How to check the validity of the forecast

- Target period of validation:
    1. June 1, 2019 – June 30, 2019
    2. July 1, 2019 – July 31, 2019
    3. August 1, 2019 – August 20, 2019

- A forecast model that shows the lowest MAPE is used.
- Data before the target periods was used for machine learning.

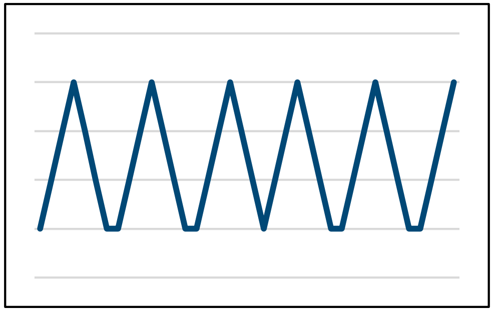


Time series data

- Time series data
  - consists of data recorded at a certain interval (i.e. POS data)
  - might change due to

1. trends
2. Periodicity
3. Seasonal/annual variation
4. Other irregulars

Stationary

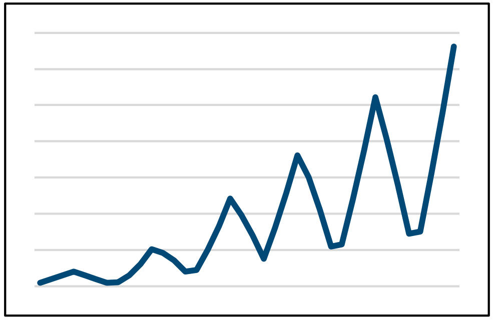


- Stationary
  - Always shows the same probability distribution

The value changes only within a certain range.

- Stationary process
- Non-stationary process

Non-stationary



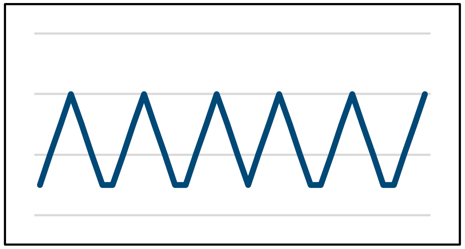
Forecasting Models

1. ARIMA (Auto Regressive Integrated Moving Average) model
   - — one of the major time series forecasting models.
   - — converts non-stationary data to stationary data by removing differences.

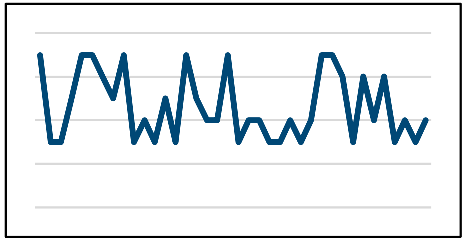
ARIMA (p,d,q) model equation:

$$y_t - y_{t-d} = c + \emptyset_1 y_{t-1} \dots \emptyset_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} \dots \emptyset_p y_{t-p}$$

- Auto regressive model
  — The value changes only within a certain range.



- Moving average model
  — The change is irregular.



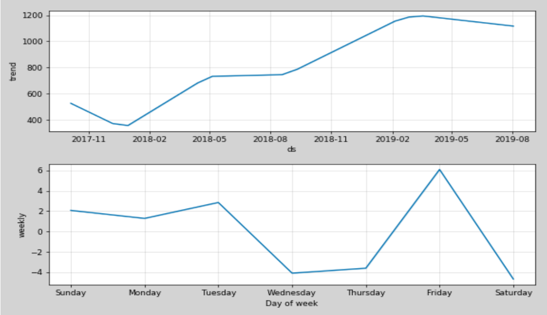
Forecasting Models

2. Prophet
   - an open source time series forecasting library by Facebook.
   - detects trends and seasonal variations automatically.
   - another type of Generalized Additive Model.

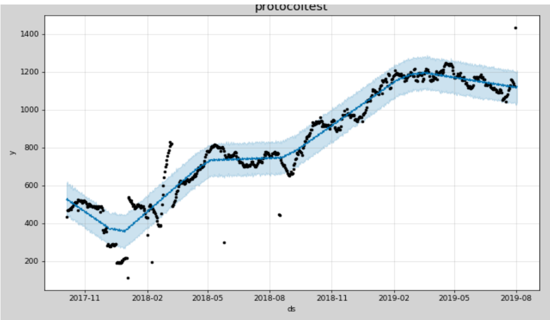
Model equation: $y(t) = g(t) + s(t) + h(t) + \epsilon_t$

y: Forecast
g: Trends
s: Seasonal changes

h : Holidays
ε : Error



   - Steps to forecast
     - ⊖ Draw a spline curve
     - ⊜ Extend the curve to the point of forecast



3. Average of latest actual record

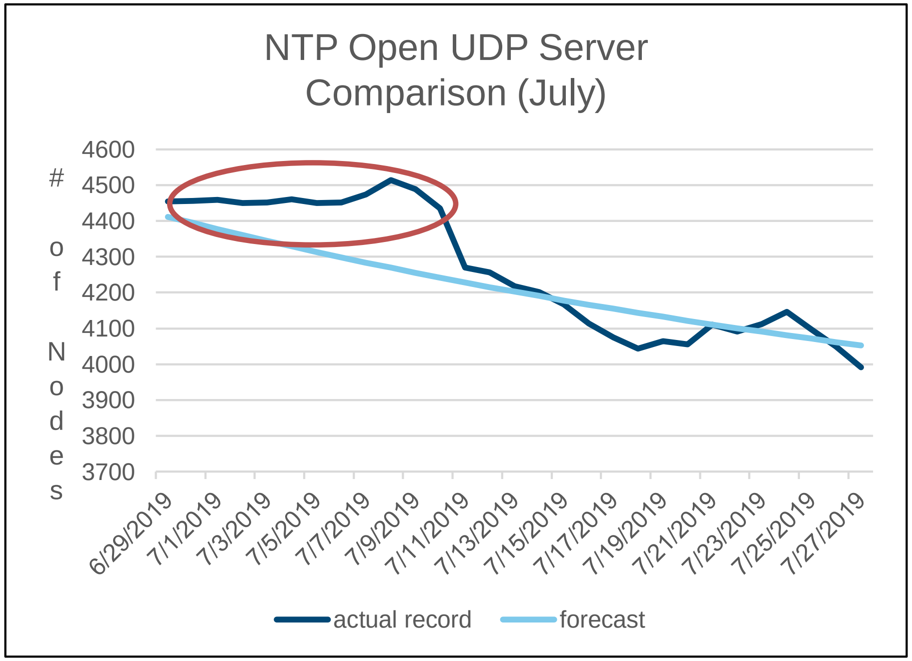
Results (Open UDP Server nodes)

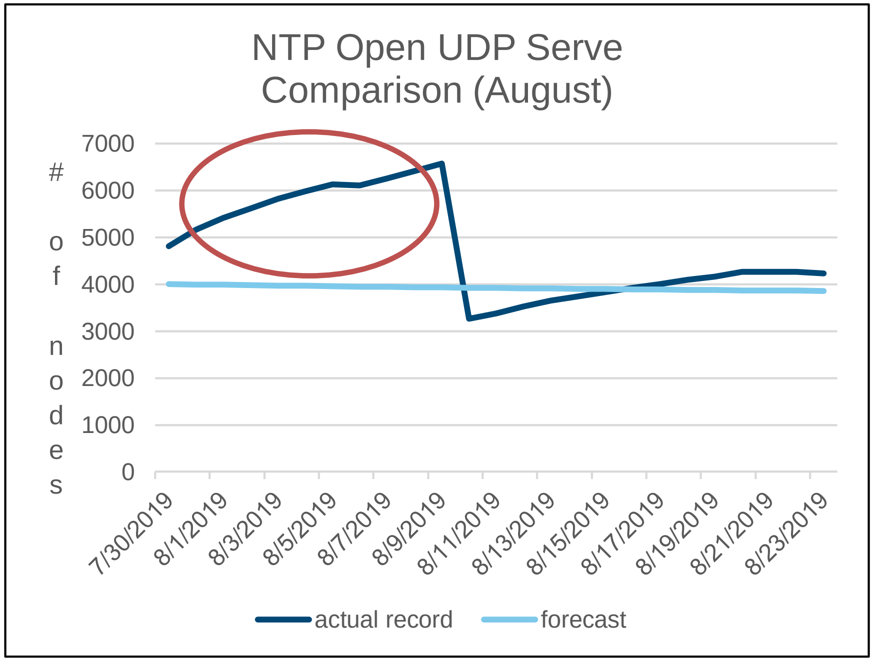
- The average MAPE of every protocol was under 15%.
- The average MAPE is high in August due to abnormal change.

Protocol	Model	Period used for Learning	Avg. MAPE (Jul-Aug)	Avg. MAPE (Jun)	Avg. MAPE (Jul)	Avg. MAPE (Aug)
DNS	Prophet	1 year	10.65%	1.86%	4.56%	(2) 18.20%
NTP	Prophet	all	10.90%	2.43%	5.82%	17.21%
SIP	Prophet	1 year	11.66%	2.12%	8.19%	17.15%
SNMP	ARIMA	1 year	10.66%	2.87%	4.89%	17.82%
SSDP	ARIMA	1 year	15.14%	(1) 2.78%	13.49%	17.18%
MSDS	Prophet	all	11.75%	3.24%	4.12%	21.20%

Comparison (Open UDP Server nodes)

- Findings
  - The forecast captured the trend of decrease.
  - The forecast was not accurate on daily level.





Results (ODR line parameter)

- Equation

  Y=(1.14(㊀)±0.04(㊁))*x+(-4.45(㊂)±0.26(④)),residualVariance=0.2(⑤)

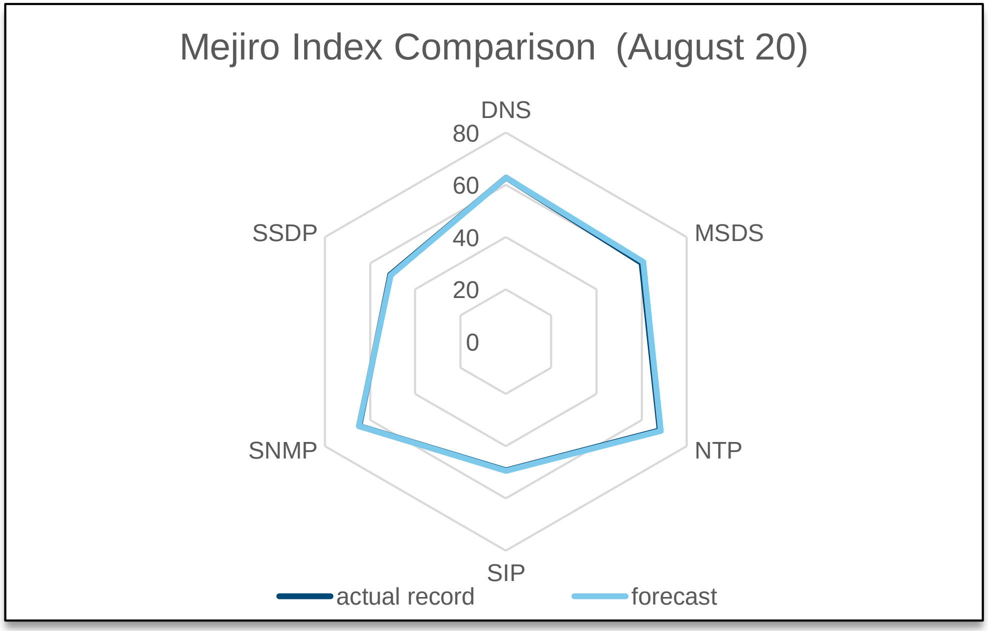
- DNS

Protocol	Model	Period used for Learning	Avg. MAPE (Jul-Aug)	Avg. MAPE (Jun)	Avg. MAPE (Jul)	Avg. MAPE (Aug)
(1)	Average	1 week	0.84%	0.88%	1.04%	0.51%
(2)	Average	1 week	0.00%	0.00%	0.00%	0.00%
(3)	ARIMA	1 year	1.89%	1.51%	1.45%	1.58%
(4)	Average	1 week	2.38%	3.52%	1.07%	4.41%
(5)	Average	1 week	4.73%	4.10%	4.41%	4.76%

Forecasting Mejiro Index

- Mejiro index was calculated using the forecast of the Open UDP Server and ODR line parameter.
  - There were very few errors in every protocol.
  - This may be because the actual number of each country’s Open UDP Server nodes was used.

➡ Further research is needed.



4. SUMMARY

Summary

- Mongolia’s cyber space
  - The analysis gave some insights for further research.
  - The analysis is not yet deep enough for us to come up with an actual measure.
  - Some parts of the past data have unusual changes due to the data collection methods.

Summary

- Forecasting
  - The number of Open UDP Server nodes
    - Forecasting the number of Open UDP Server nodes during one month was successful with the average MAPE under 15%
    - We could not yet forecast the transition of short span of time.
  - ODR line parameters
    - The forecast was successful with the average MAPE under 5%.
    - The forecasting was easy because the value did not change largely.

➡ **Forecasting is usable to detect unusual change**

Plans for the future

- Add forecasting service to Mejiro
- For the time being, Mejiro needs to…
  - start analyzing more countries.
  - use data from other sources for further analysis.

Still, it is <u>difficult for JPCERT/CC alone</u> to analyze all countries in the world.

Exchanging analysis results can benefit both countries.

We can provide Mejiro's data of your region upon request.

Thank You!

JPCERT Coordination Center

- **Email：global-cc@jpcert.or.jp**
- **Tel：+81 03-6271-8901**
- **https://www.jpcert.or.jp/**