# “Fault Detection, Diagnosis, and Prediction for IP-based Industrial Control Networks”

## - PhD Thesis Defense -

**Young J. Won**

**Supervisor: Prof. James W. Hong**

**Nov. 2009**

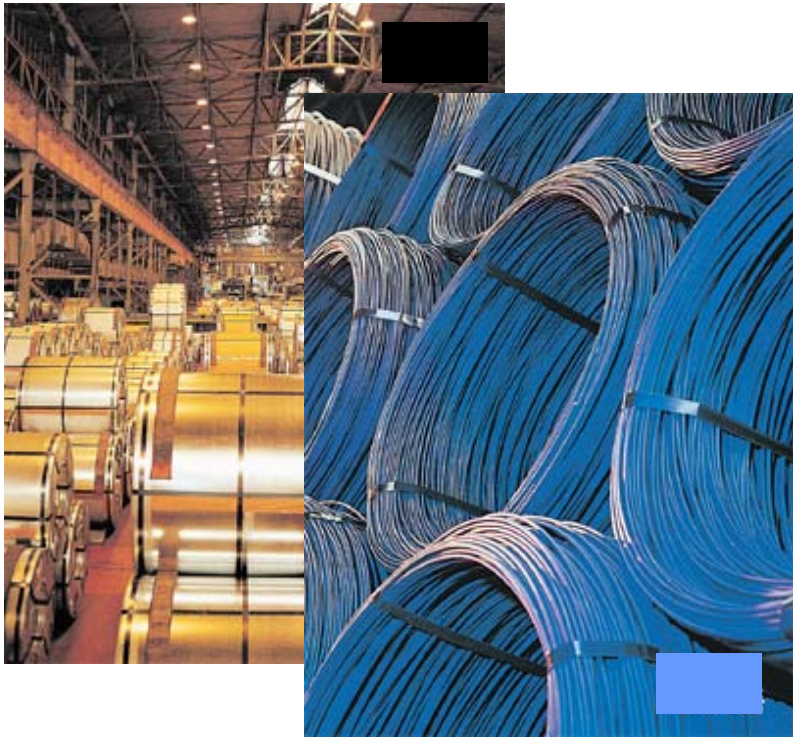
**Dept. of Computer Science and Engineering**
**POSTECH, Korea**
**yjwon@postech.ac.kr**

# Table of Contents

# What is an ICN?

## ❖ Industrial Control Networks (ICN)

- Supports robust communications between controlling and controlled devices in a manufacturing environment
  - Factory and Process Automation
- Mission-critical networks that cannot tolerate faults



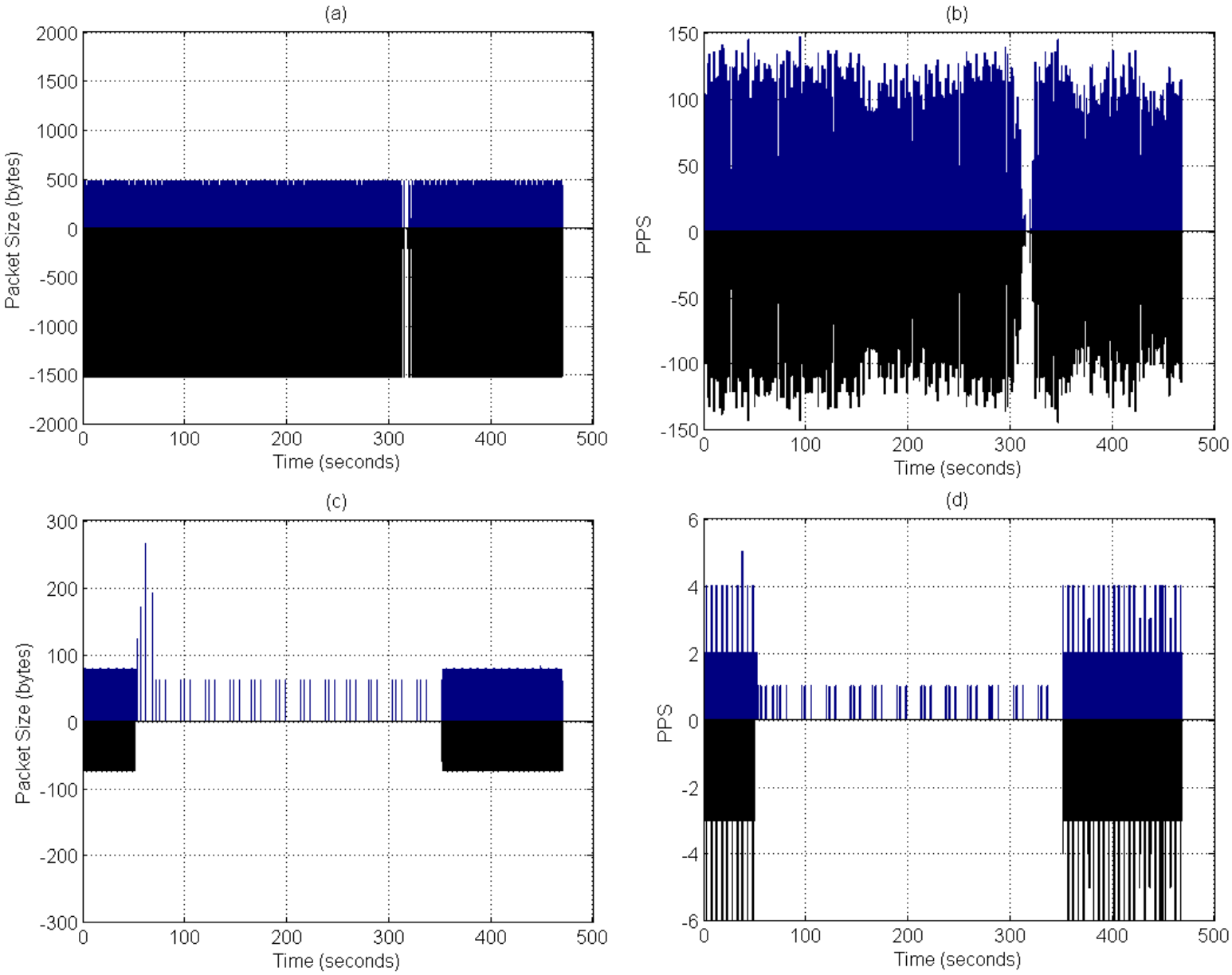
# Motivation

- **ICNs are moving from proprietary to IP-based**
  - Lowering cost (CAPEX & OPEX)
  - To avoid vendor lock-in
- **A growing need for monitoring, detection and prediction of faults in IP-based ICN**
  - Traditional faults in proprietary ICNs as well as IP-centric faults are occurring
  - Continuous, reliable and efficient operations of ICNs are critical in today's competitive world
- **Previous IP diagnosis techniques are not applicable**

# Problems

- Traditional IP fault metrics must be interpreted in different ways for ICN environment
- Accommodating ICN-specific operational processes with a single unified approach
- Consequences of network malfunctions are severe

# Traffic View – Network Instability



# Hypothesis



*The unique demands of ICNs require new fault monitoring and diagnosis techniques.*

# Research Questions (1/2)

- What is unique about ICNs?
- Are the existing fault detection and analysis techniques in conventional IP networks sufficient for diagnosing ICN specific faults?
- What features should be investigated in order to identify or represent ICN faults?

# Research Questions (2/2)

- What methods should be developed for accurate, scalable, and adaptive fault diagnosis systems for ICN?
- How can we predict fault occurrences to prevent possible failures of the network?
- What are the next research steps towards more advanced fault diagnosis?

# Research Methodology

POSTECH 2020

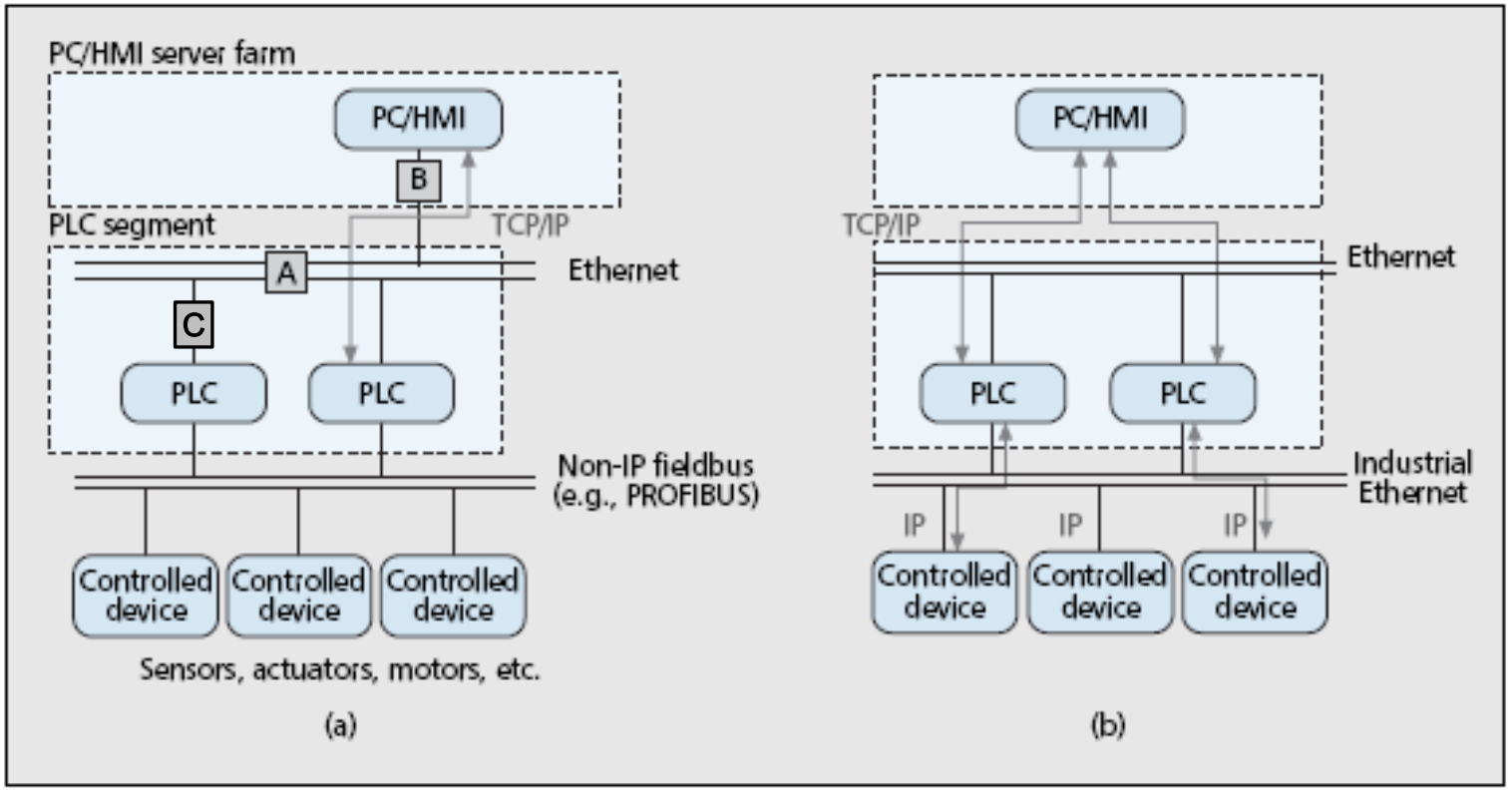
- **The research methodology consists of studying the following key ICN elements:**
  - **ICN traffic characteristics and fault features**
  - **ICN fault diagnosis and prediction methods**
  - **ICN fault diagnosis system architecture**
  - **Deployment experiences for validation of the proposed methods**
  - **A reference architecture for handling faults in a complete network management cycle: Monitoring, Analysis, Prediction, and Decision**

# Related Work

- ICN traffic characteristics
- Network design, modeling, and requirements
- ICN QoS requirements
- Industrial Ethernet applicability
- Alarm management
  - IETF, TMF, 3GPP, etc.
- Diagnostics

- **Empirical Traffic Analysis: ICN**
- **Fault Diagnostics & System Architecture**
- **Prediction & Adaptive Decision**

# Simplified View of an ICN




**Acronyms:**

| | |
|---|---|
| **PC** | **Process Computer** |
| **PLC** | **Programmable (Logic) Controller** |
| **HMI** | **Human Machine Interface** |

# Traffic Summary

| Data Set | Date | Duration | Byte | Packets | Flows | TCP | Utilization |
|---|---|---|---|---|---|---|---|
| Segment-A | 2006-09-29 | 170 hrs | 63.5 GB | 542 M | 48 K | 98 % | 1 % |
| Segment-B | 2007-02-27 | 10 hrs | 74 GB | 122 M | 25 K | 99 % | 19 % |
| Segment-C | 2006-05-11 | 5 mins | 22 MB | 84 K | 48 K | 99 % | 0.57 % |

- **Monitoring points**
  - A: Top of PLC segments (a group of PLC networks at the edge)
  - B: ICN backbone
  - C: End PLC host
- **Macroview of traffic characteristics**
  - No sudden traffic burst has been observed
    - Avg. utilization is still very low
  - A small number of identical sessions with fixed amount of hosts (PLC and PC) are continuously observed

# ICN Traffic Characteristics

POSTECH 2020

## ❖ Key observations

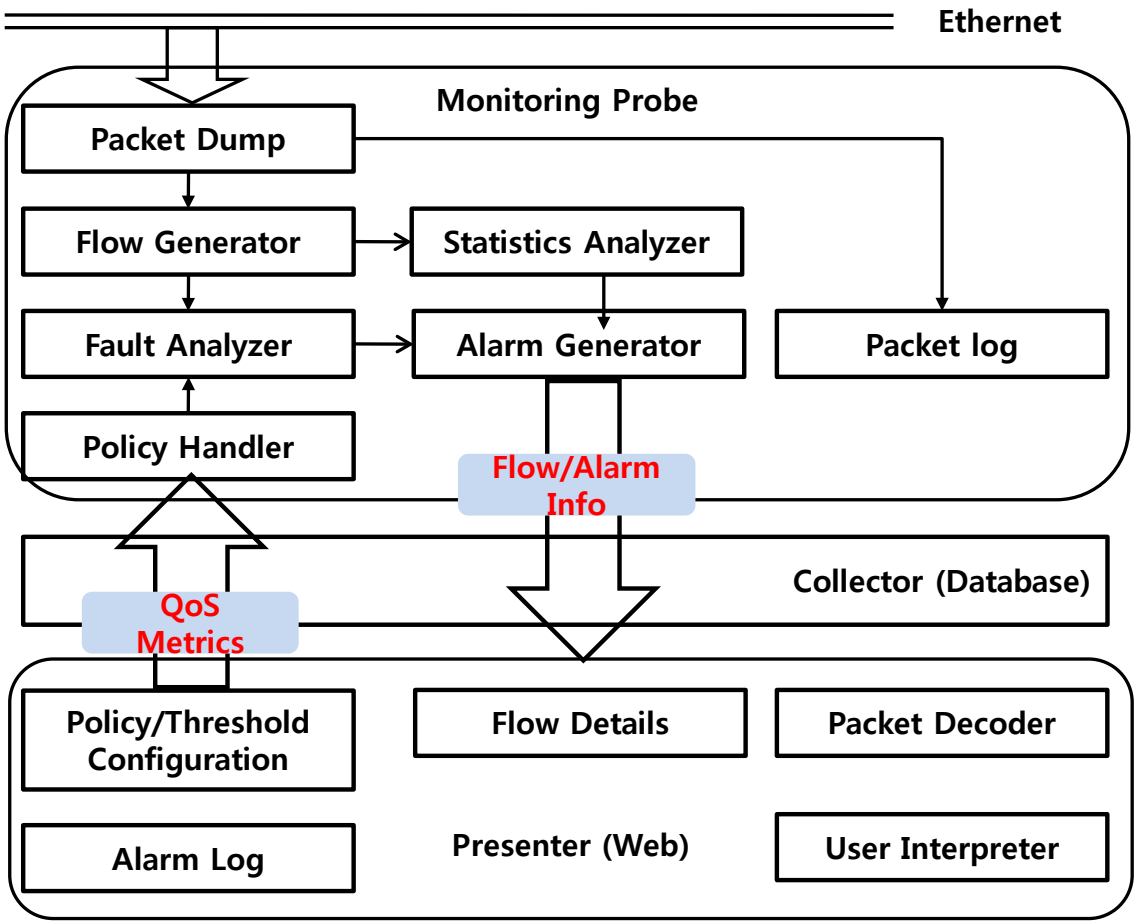
- Low-yield and steady bandwidth usage regardless of monitoring periods
- Periodic traffic cycle in terms of packet arrival sequence and inter-arrival time
- Traffic symmetry
- Occurrence of small signaling packets
- Session length distribution patterns
- High packet reordering ratios at the backbone

## ❖ Consequences

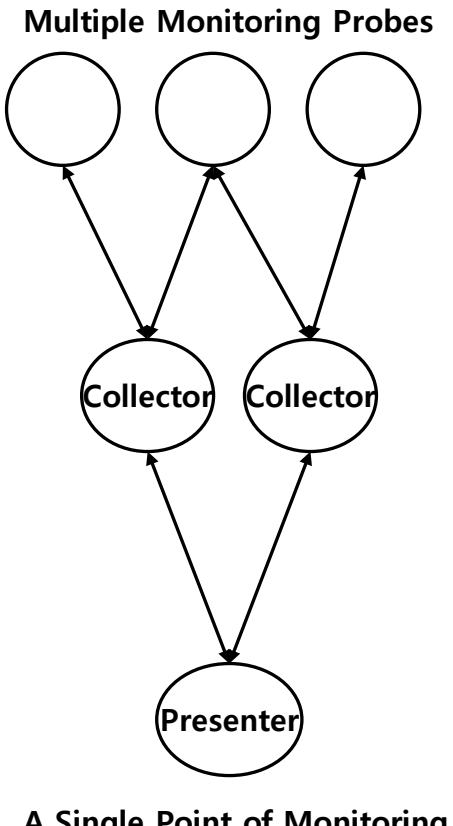
- A first work to measure and analyze control IP networks from the network perspective
- Presentation of unique characteristics of control IP networks
- Provide a precise network snapshot

- Empirical Traffic Analysis: ICN
- **Fault Diagnostics & System Architecture**
- Prediction & Adaptive Decision

# Architecture

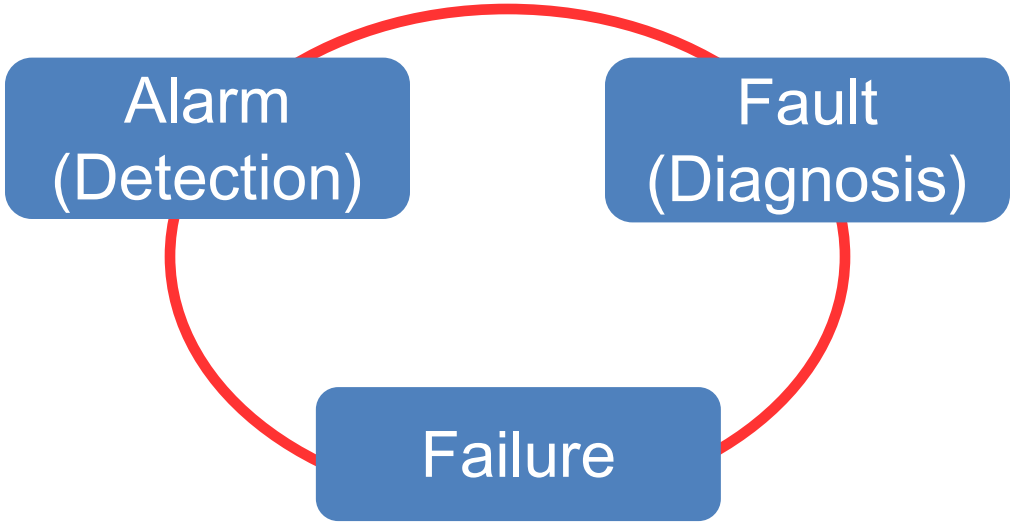



(a) System Architecture




(b) Abstraction

# Alarm, Fault, and Failure




## ❖ Alarm

- **Precautionary system notification of communication fault**
- **Invoked by early symptoms of communication difficulties**

## ❖ Fault

- **Phenomenon of unstable communication conditions**

## ❖ Failure

- **A complete stoppage of the manufacturing process**
- **Realization of a malfunctioning process**

# Relying on User Perception

| Fault Causes | Phenomenon |
|---|---|
| Ethernet duplex mismatch | •Frame collision<br>•Unauthorized IP access<br>•Irregular communication termination<br>•Communication delay and loss<br>•Unexpected shutdown<br>•Unordered packet sequence<br>•Unordered message sequence<br>•Message corruption<br>•Broadcast packet flooding<br>•Low throughput<br>•Duplicate packet arrival |
| PLC programming bugs | |
| Device driver bugs | |
| Link Corruption | |
| Damage to Interface | |
| Harsh Condition to H/W<br>(Dust in optical device, current short due to metal, high/low temperature, moisture) | |
| Duplicate addresses | |
| Protocol unawareness | |
| Unavailable bandwidth | |
| Electrical noise | |
| Power outage | |
| Mis-configuration | |
| Worms | |

# ICN QoS Metrics

POSTECH 2020

| ICN QoS Metrics | OSI Layer |
|---|---|
| Collision frame | Physical |
| CRC error frame | |
| Dropped frame | |
| Jumbo frame occurrence | Data Link |
| Runts frame occurrence | |
| IP checksum error | Network |
| Fragment packet | |
| Packet overflooding | |
| Packet inter-arrival time variation | |
| Packets per second variation | |
| Packet size variation | |
| Throughput variation | |
| TCP/UDP checksum error | Transport |
| TCP window size drop | |
| TCP packet sequence violation | |
| TCP retransmission packet occurrence | |
| Unsupported protocol packet occurrence | Transport or Application |
| Rule sequence violation | Application |
| Rule inter-arrival cycle violation | |
| Policy cycle violation | |

# Early Symptoms for ICN Faults

POSTECH 2020

| Index | Network metrics | Alarm conditions |
|---|---|---|
| 1 | Collision frames | First appearance, or threshold-based |
| 2 | Jumbo (≥ 1514 bytes) frames | First appearance |
| 3 | Runts (≤ 64 bytes) frames | First appearance |
| 4 | CRC error frames | First appearance |
| 5 | IP/TCP checksum errors | First appearance |
| 6 | Fragment packets | Threshold-based |
| 7 | Retransmission packets | First appearance, or threshold-based |
| 8 | Packet interarrival time (ms) | Increase to the previous value |
| 9 | Throughput (b/s) | Decrease, drop to 0, or pattern analysis over monitoring period |
| 10 | Packets per second (or packet burst) | Increase, decrease, drop to 0, or pattern analysis over monitoring period |
| 11 | Min/max/diff packet size (bytes) | Change in difference of max and min sizes over monitoring period |
| 12 | Min/max/diff TCP window size | Drop to 0, change in difference of max and min sizes over monitoring period |
| 13 | Out-of-order sequence packets | First appearance |
| 14 | Broadcast packets | Threshold-based |
| 15 | Unsupported protocol packets | Threshold-based |

# Beyond Primitive Symptoms

POSTECH 2020

## ❖ Application-layer QoS requirements must be met

- Packet reordering to Message sequence violation
- Packet delay, loss to Cycle (timing) violation
- Packet delay to Message delay
- Packet loss to Message loss
- Malformed packet to Message corruption

**Mapping these relations using a "Policy-based approach"**

## ❖ Requires Deep Packet Inspection as well

- Packet inspection capability +
  Network-level QoS condition check +
  Message condition check
- Signature matching technique
  - Signature is a portion of payload data that is static and distinguishable for applications (strings or hex)

# Policy Details

POSTECH 2020

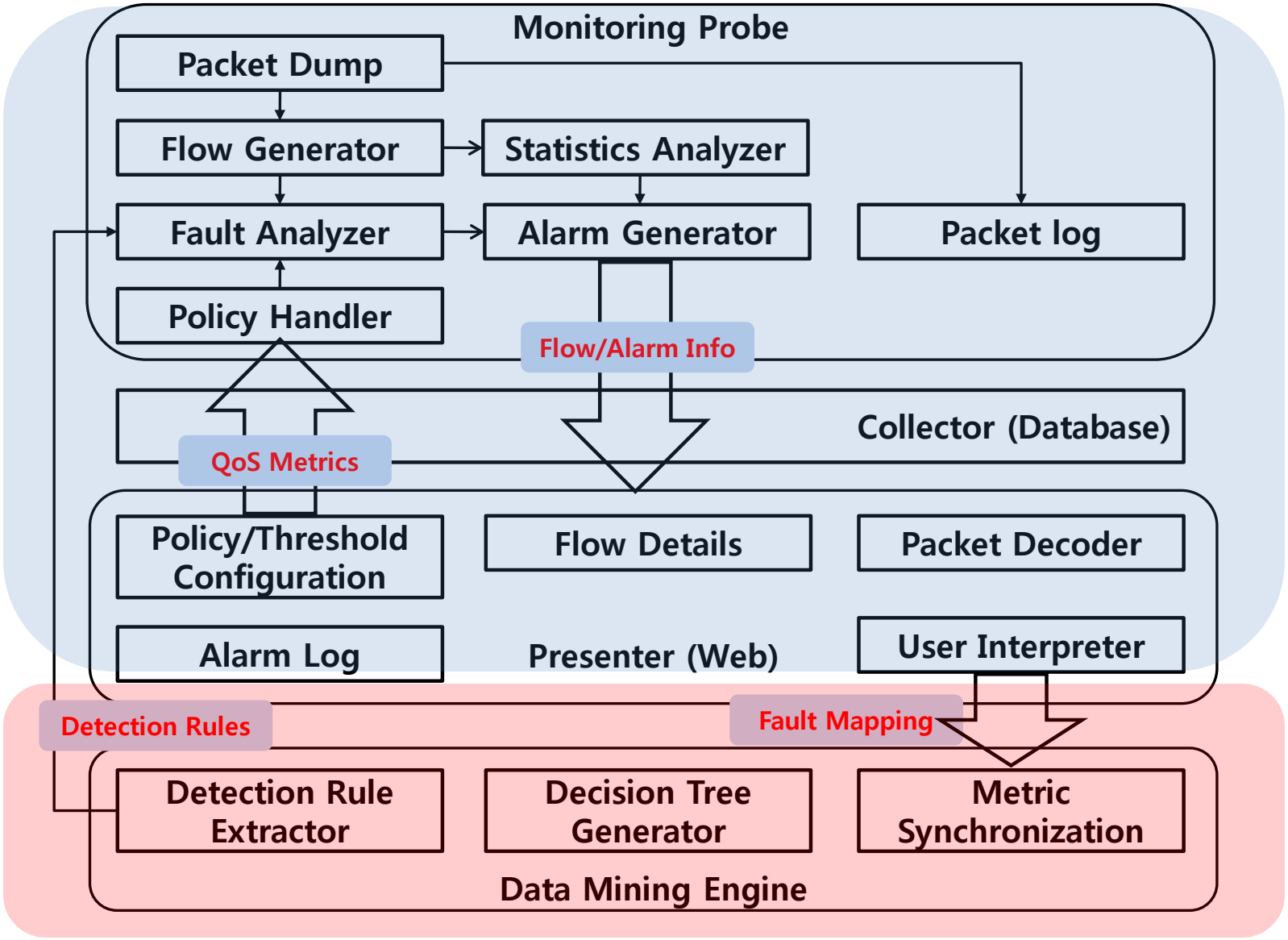
| Attributes | Description | Type |
|---|---|---|
| Policy name | Fault Case | String |
| Policy cycle | Maximum PLC cycle time (ms) for message exchange. It must not exceed the total sum of next sequence arrival time in each rule. | Integer |
| Source/Destination MAC address | Ethernet hardware addresses (hex) of two communication devices | String |
| Source/Destination IP address | IP addresses of two communication devices | String |
| Source/Destination Port | Port number | Integer |
| Signature | A pattern of hexadecimal digits or specific strings that are preset in the packet's payload. It refers to a portion of PLC message or protocol (e.g., 0x6000 at offset 0) | Hex or String |
| Offset | Position of signature in packet's payload | Integer |
| Word size | Byte size of signature | Integer |
| Rule sequence index | Arrival order among the rules | Integer |
| Rule inter-arrival time | Expected inter-arrival time till the same rule occurs again | Integer |
| Rule next sequence arrival cycle | Expected inter-arrival time between the current rule and upcoming rule | Integer |

# Policy Example

```
<policy-table>
    <policy name="PLC_reply_error">
        <policy-list cycletime="2000">
            <rule
                src_mac="08007023420e" dst_mac="aa0004003250"
                src_ip="130.30.141.53" dst_ip="130.30.10.41"
                src_port="1026" dst_port="8453"
                signature="6000" offset="0" word_size="2"
                rule_sequence="1"
                rule_intertime="1500"
                rule_nexttime="500"
            />
            <rule
                src_mac="08007023420e" dst_mac="aa0004003250"
                src_ip="130.30.141.53" dst_ip="130.30.10.41"
                src_port="1026" dst_port="8453"
                signature="e000" offset="0" word_size="2"
                rule_sequence="2"
                rule_intertime="1000"
                rule_nexttime="500"
            />
        </policy-list>
</policy-table>
```

- Empirical Traffic Analysis: ICN
- Fault Diagnostics & System Architecture
- **Prediction & Adaptive Decision**

# Architecture – Extension of Decision Engine



# Prediction Model

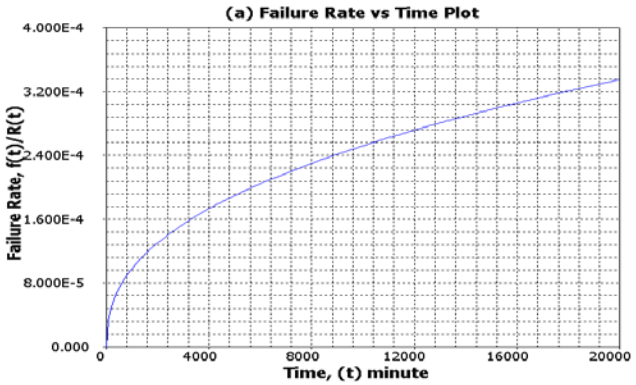
- **Measurement of alarm occurrence ratios against various threshold values for each fault cases**
  - Modeling against real-world alarm occurrence ratios in POSCO
  - Projection of alarm occurrences
  - Two-parameter Weibull Distribution Model:

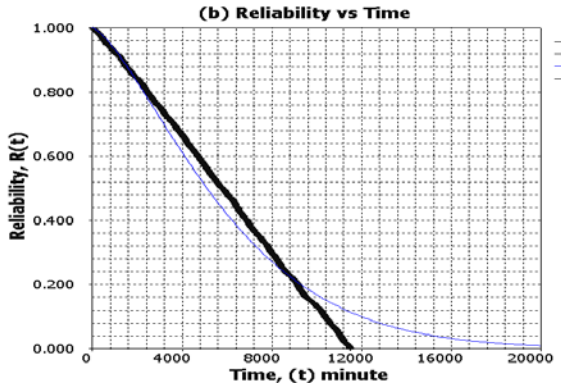
$$f(t) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} e^{-(\frac{t}{\eta})^{\beta}}$$

  - *Reliability (t) = 1 – F(t)*
  - *Failure (t) = f(t)/R(t)*
- **Optimized threshold values for the following fault cases: Window (1), Out of sequence (50), Retransmission (100)**
  - Most frequently occurring faults in the testing traces
  - Manipulating alarm occurrence ratios with various threshold values

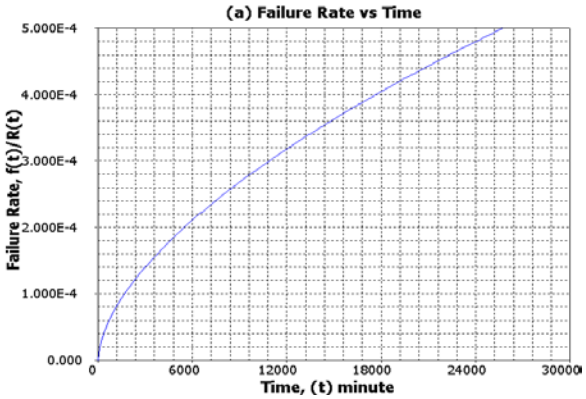
| Alarm category | Threshold value | Segment A | | Segment B | |
|---|---|---|---|---|---|
| | | β | η | β | η |
| Window size error 1 | 1 | 1.4107 | 6627.8192 | 2.1505 | 372.6533 |
| Out of sequence 10 | 10 | 1.5938 | 6941.4623 | 2.9220 | 313.0396 |
| | 50 | 1.6234 | 6957.8985 | 2.9220 | 313.0396 |
| | 100 | 1.6244 | 6959.1457 | 2.9220 | 313.0396 |
| Retransmission 100 | 100 | 1.1852 | 6228.7397 | 2.9836 | 319.2062 |
| | 500 | 3.1364 | 276.5284 | 255.3608 | 264.4731 |
| | 1000 | 3.1567 | 277.3797 | 255.3608 | 264.4731 |

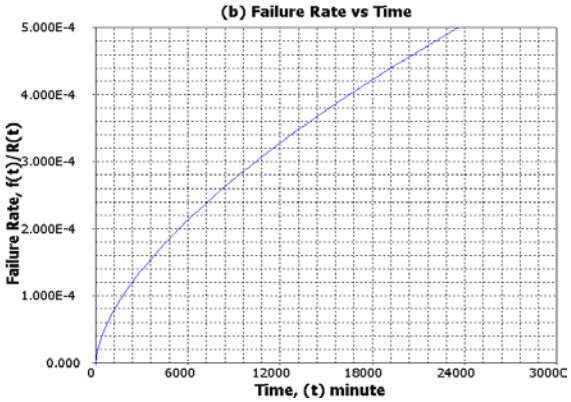
# Evaluation (1/2)



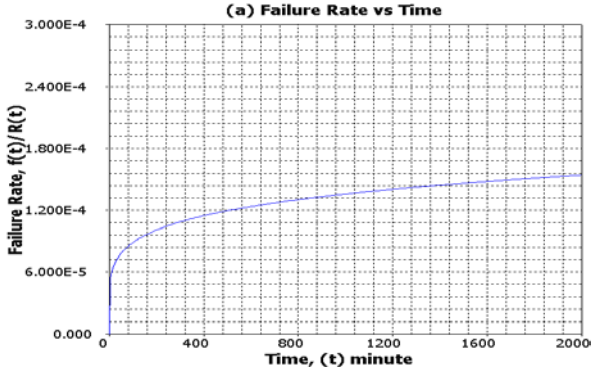



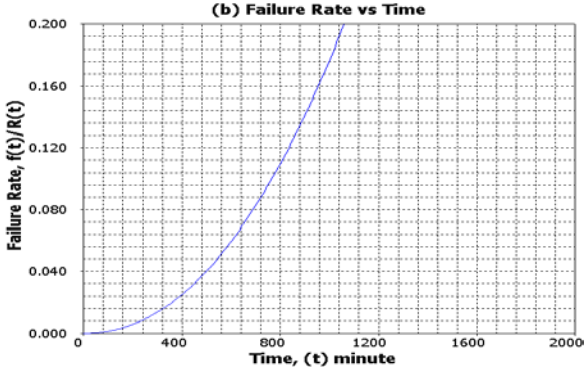

**Segment A,**
**window size error:**
**Failure (alarm) probability and reliability fit the empirical complementary CDF**




**Segment A,**
**out-of-sequence:**
**Failure (alarm) probability for threshold 10-(a) and 50-(b)**




**Segment A,**
**Retransmission:**
**Failure (alarm) probability for threshold 100–(a) and 500–(b)**

# Evaluation (2/2)

POSTECH 2020

## ❖ Fault Prediction

- Diagnosis system currently in operation
- Based on log data (2007) and offline pattern analysis
- Reliability engineering approach
  - Probability of fault occurrence
  - Projected value of fault reoccurrence over time-series (~∞)
  - Weibull distribution

## ❖ Accuracy analysis

- These faults have been observed in a week long set of testing traces from Segment A

| Segment A | Window error | Out of sequence | Retransmission |
|---|---|---|---|
| Minutes | Probability of alarm (%) | Probability of alarm (%) | Probability of alarm (%) |
| 1440 (1 day) | 10.95 | 7.45 | 100 |
| 2880 (2 days) | 26.54 | 21.24 | 100 |
| 4320 (3 days) | 42.11 | 36.95 | 100 |
| 5760 (4 days) | 55.97 | 52.09 | 100 |
| 7200 (5 days) | 67.49 | 65.25 | 100 |
| 8640 (6 days) | 76.62 | 75.85 | 100 |
| 10080 (7 days) | 83.58 | 83.88 | 100 |

# Adaptive Decision Engine

POSTECH 2020

## ❖ An excellent case for Machine Learning (ML)

- Continuous, cyclic, no frequent switching between steady-state and burst states ➔ No Surprises

## ❖ Our approach includes

- Defining up to 30 classifiers
- Using the current diagnosis system as a fault detector
- Feedback control rules and decision engine
- Discrete time updater and accuracy analysis

## ❖ Limitations

- The size of data set is a problem from the perspective of ML researchers
  - SVM is more suitable, but not applicable in this case

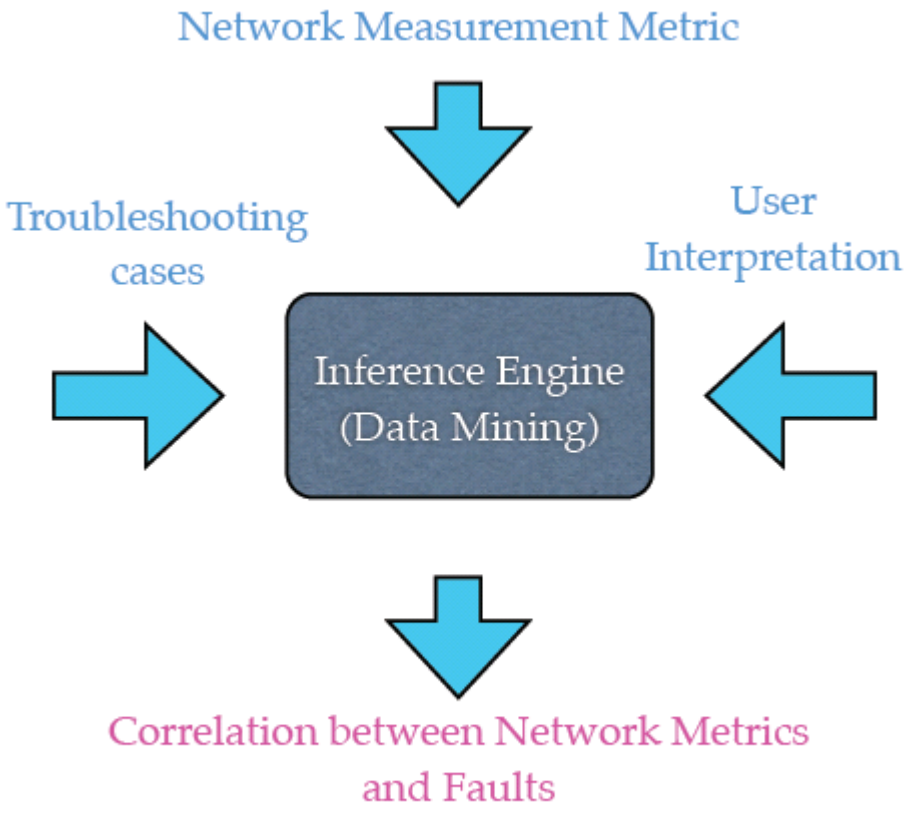
# Methodology for Causal Analysis



- **Data mining technology: Inference engine for fault detection**

- **Inputs**
  - Network metrics
  - Troubleshooting cases
  - User interpretation (yes/no)

- **Output**
  - Correlation between network metrics and actual fault
  - Fault detection rules
  - Determined fault flows

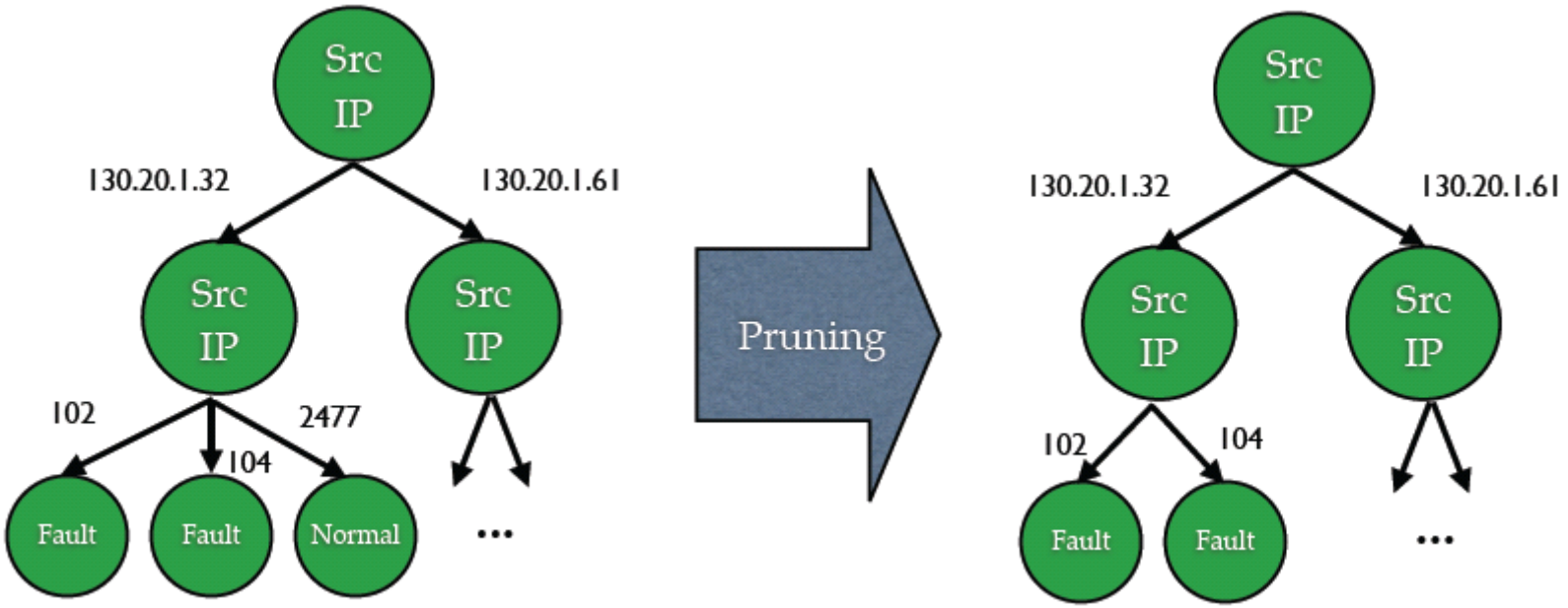


# Decision Tree Algorithm



## ❖ Binary classification: Fault vs. Normal

- **Based on C4.5 algorithm**
  - **Selectively uses up to 30 classifiers**
- **Modifications are optimized for handling unbalanced classification (skewed data set)**



# Validation - Setup

## ❖ Data collection

- Full packet traces from the frequently troubled ICN in POSCO
- 5-day long uninterrupted data, 75 Gbytes

## ❖ Data processing

- Aggregate traffic flow
- Extract features from flow
- Manual analysis
- Filter training/testing data
- 3,058 fault flows vs. 60,572 normal flows

| Index | Source IP | Destination IP | Source Port | Destination Port | … | Class |
|---|---|---|---|---|---|---|
| 1 | 10.1.1.76 | 224.0.6.127 | 1043 | 8044 | … | normal |
| 2 | 10.1.1.87 | 244.0.6.17 | 1043 | 8044 | … | normal |
| 3 | 10.1.1.88 | 10.255.255.255 | 138 | 138 | … | normal |
| … | … | … | … | …. | … | …. |
| 7 | 130.20.21.213 | 130.20.21.255 | 138 | 138 | … | fault |
| 8 | 130.20.21.213 | 130.20.21.255 | 138 | 138 | … | fault |
| … | … | … | … | …. | … | …. |
| 63630 | 130.21.21.157 | 130.1.21.255 | 138 | 138 | … | fault |

# Validation Against Ground Truth

POSTECH 2020

- **Reducing false negatives (FN) is critical**
  - Several empirical studies to reduce FNs
  - False positive (FP), True positive (TP)
- **Accuracy analysis of fault decisions**
  - Combinations of feature set and algorithms

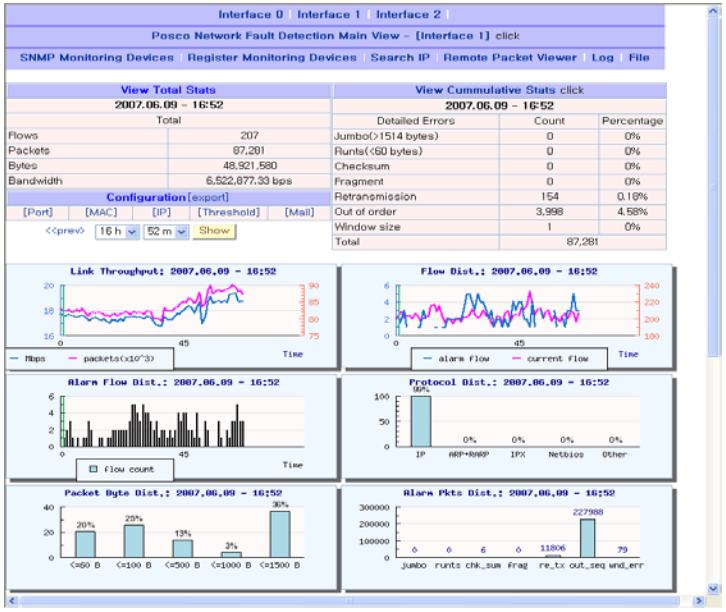
- $precision = \frac{TP}{TP + FP}$
- $recall = \frac{TP}{TP + FN}$
- $F1 = \frac{2 \cdot recall \cdot precision}{recall + precision}$

| Experiment | TP Rate | FP Rate | Precision | Recall | F-measure | Class | # of FNs | # of FPs |
|---|---|---|---|---|---|---|---|---|
| Standard C4.5 | 0.974 | 0 | 1 | 0.974 | 0.987 | abnormal | 128 | 0 |
| | 1 | 0.026 | 0.989 | 1 | 0.994 | normal | | |
| IP pair as a feature | 0.974 | 0 | 1 | 0.992 | 0.995 | abnormal | 40 | 3 |
| | 1 | 0.008 | 0.997 | 1 | 0.998 | normal | | |
| IP pair as a feature without pruning | 0.992 | 0 | 0.999 | 0.99 | 0.995 | abnormal | 39 | 7 |
| | 1 | 0.008 | 0.997 | 1 | 0.998 | normal | | |
| Multi-stage decision tree | 0.995 | 0 | 1 | 0.995 | 0.997 | abnormal | 22 | 2 |
| | 1 | 0.005 | 0.998 | 1 | 0.999 | normal | | |

# Deployment Experience

- **System deployment at POSCO, Pohang**
  - #2 Hot Rolled Plate (HRP) plant since 2007
  - #1 and #3 HRP plants since 2008
  - Deployed multiple systems at other plants

- **Different deployed application scenarios**
  - PLC message abnormality recognition - 2007.10
  - Packet flooding recognition - 2008.4
  - Additional diagnosis cases
    - Restricted due to confidentiality constraints



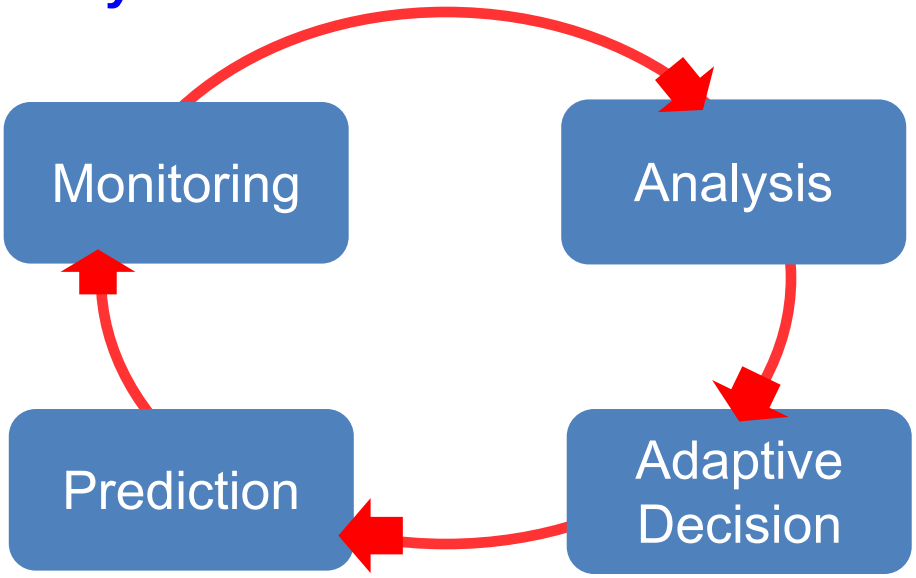

# Concluding Remarks

# Summary

- Differentiate ICN from traditional IP networks
- Analyze IP-based ICN operations to develop fault diagnosis and prediction methods
- Additional work to meet the need of real-world use case led to adaptive decision engine
- Methods validated against the real-world ICN data

# Contributions

- **Derived features for new fault diagnosis mechanism by distinguishing IP-based ICN traffic from traditional IP-based traffic**
- **New novel fault diagnosis, prediction, and adaptive decision methods are described and validated with real-world ICN data**
- **Achieving a complete cycle of management**
  - **A unique reference study for network monitoring and measurement**
- **Moving beyond analysis**



# Future Work

- **Enhance the system to cope multi-gigabit environments in ICN**
  - Adapting specialized network interface cards (programmable interface with NPUs)
- **Prediction model for actual network failures**
  - Matching the alarms with real network outage cases
- **Sophisticated correlation models**
  - Passively reacting to the fault case after the situation awareness
  - Relying on ‘Ontological Representation’ among fault, failure, and counter-actions
  - Automated policy extraction strategy
- **Autonomic management of ICNs**
  - A step towards self-* or using autonomic architecture like FOCALE


# 바쁘신 시간 내주셔서 감사합니다

# Publications (1/3)

POSTECH 2020

## ❖ International Journal/Magazine Papers (4)

1. **Young J. Won**, Mi-Jung Choi, James W. Hong, Chan-Kyu Hwang, and Jae-Hyoung Yoo, "Measurement of Download and Play and Streaming IPTV Traffic," IEEE Communications Magazine, Vol. 46, No. 10, October, 2008, pp. 154-161. **(SCI)**
2. **Young J. Won**, Mi-Jung Choi, James W. Hong, Myung-Sup Kim, Hwa Won Hwang, Jun Hyub Lee, and Sung-Gyoo Lee, "Fault Detection and Diagnosis in IP-based Mission Critical Industrial Process Control Networks," IEEE Communications Magazine, Vol. 46, No. 5, May 2008, pp. 172-180. **(SCI)**
3. Myung-Sup Kim, **Young J. Won**, and James W. Hong, "Characteristic Analysis of Internet Traffic from the Perspective of Flows," Elsevier Computer Communications, Vol. 29, Issue 10, June 19 2006, pp. 1639-1652. **(SCIE)**
4. Myung-Sup Kim, **Young J. Won**, and James Won-Ki Hong, "Application-Level Traffic Monitoring and Analysis on IP Networks," ETRI Journal, Vol.27, No.1, Feb. 2005, pp.22-42. **(SCI)**

## ❖ International Conference/Workshop Papers (14)

5. Jae Yoon Chung, Byungchul Park, **Young J. Won**, John Strassner, and James W. Hong, "Traffic Classification Based on Flow Similarity", 8th IEEE International Workshop on IP Operations & Management, Accepted to appear, Oct. 29-30, 2009.
6. Seong-Cheol Hong, Jin Kim, Byungchul Park, **Young J. Won**, and James W. Hong, "Traffic Growth Analysis over Three Years in Enterprise Networks", 15th Asia-Pacific Conference on Communications, Accepted to appear, Oct. 8-10, 2009.
7. Suman Pandey, **Young J. Won**, Hong-Taek, Ju, and James. W. Hong, "Dimensioning of IPTV VoD Service in Heterogeneous Broadband Access Networks", 12th Asia-Pacific Network Operations and Management Symposium, Accepted to appear, Sept. 23-25, 2009.
8. Sung-Su Kim, **Young J. Won**, Mi-Jung Choi, James W. Hong, and John Strassner, "Towards Management of the Future Internet," IFIP/IEEE Workshop on Management of the Future Internet (conjunction with IM 2009), New York, USA, June 5, 2009.

# Publications (2/3)


9. Byunchul Park, **Young J. Won**, Hwanjo Yu, James W. Hong, Hong-Sun Noh, and Jang Jin Lee, “Fault Detection in IP-based Process Control Networks using Data Mining,” 11th IFIP/IEEE Integrated Network Management, New York, USA, June 1-5, 2009.
10. Byung-Chul Park, **Young J. Won**, Mi-Jung Choi, Myung-Sup Kim, and James W. Hong, “Empirical Analysis of Application-level Traffic classification using Supervised Machine Learning,” 11th Asia-Pacific Network Operations and Management Symposium (APNOMS 2008), **LNCS** 5297, Beijing, China, October 2008, pp. 474-477.
11. **Young J. Won**, Byung-Chul Park, Mi-Jung Choi, James W. Hong, Hee-Won Lee, Chan-Kyu Hwang, Jae-Hyoung Yoo, “End-User IPTV Traffic Measurement of Residential Broadband Access Networks,” 6th IEEE International Workshop on End-to-End Monitoring Techniques and Services (E2EMON), Salvador, Brazil, April 7, 2008, pp. 95-100.
12. Byung-Chul Park, **Young J. Won**, Myung-Sup Kim, and James Won-Ki Hong, “Towards Automated Application Signature Generation for Traffic Identification,” 11th IFIP/IEEE Network Operations and Management Symposium (NOMS), Salvador, Brazil, April 7-11, 2008, pp. 160-167.
13. **Young J. Won**, Mi-Jung Choi, Myung-Sup Kim, Hong-Sun Noh, Jun Hyub Lee, Hwa Won Hwang, and James W. Hong, “Measurement Analysis of IP-Based Process Control Networks,” 10th Asia-Pacific Network Operations and Management Symposium (APNOMS), **LNCS** 4773, Sapporo, Hokkaido, Japan, Oct. 10-12, 2007, pp. 385-394.
14. **Young J. Won**, Mi-Jung Choi, Jang Jin Lee, Jun Hyub Lee, Hwa Won Hwang, and James W. Hong, “Detecting Network Faults on Industrial Process Control IP Networks,” 7th IEEE International Workshop on IP Operations & Management (IPOM), **LNCS** 4786, San Jose, CA, USA, Oct. 31-Nov. 2, 2007, pp. 184-187.
15. **Young J. Won**, B.C. Park, S.C. Hong, K.B. Jung, H.T. Ju, James W. Hong, “Measurement Analysis of Mobile Data Networks,” 8th Passive and Active Measurement Conference (PAM), **LNCS** 4427, Louvain-la-neuve, Belgium, April 5-6, 2007, pp. 223-227.
16. **Young J. Won**, Byung-Chul Park, Hong-Taek Ju, Myung-Sup Kim, and James W. Hong, “A Hybrid Approach for Accurate Application Traffic Identification,” 4th IEEE International Workshop on End-to-End Monitoring Techniques and Services (E2EMON), Vancouver, Canada, April 3, 2006, pp. 1-8.
17. Seung-Hwa Chung, **Young J. Won**, Deepali Agrawal, Seong-Cheol Hong, and James Won-Ki Hong, “Detection and Analysis of Packet Loss on Underutilized Enterprise Network Links,” 3rd IEEE International Workshop on End-to-End Monitoring Techniques and Services (E2EMON), Nice, France, May 15, 2005, pp. 164-176.
18. Myung-Sup Kim, **Young J. Won**, Hyung-Jo Lee, James W. Hong, and Raouf Boutaba, “Flow-based Characteristic Analysis of Internet Application Traffic,” 2nd IEEE International Workshop on End-to-End Monitoring Techniques and Services (E2EMON), San Diego, California, USA, October 3, 2004, pp. 62-67.


## ❖ Domestic Journal Papers (3)

## ❖ Domestic Conference Papers (9)

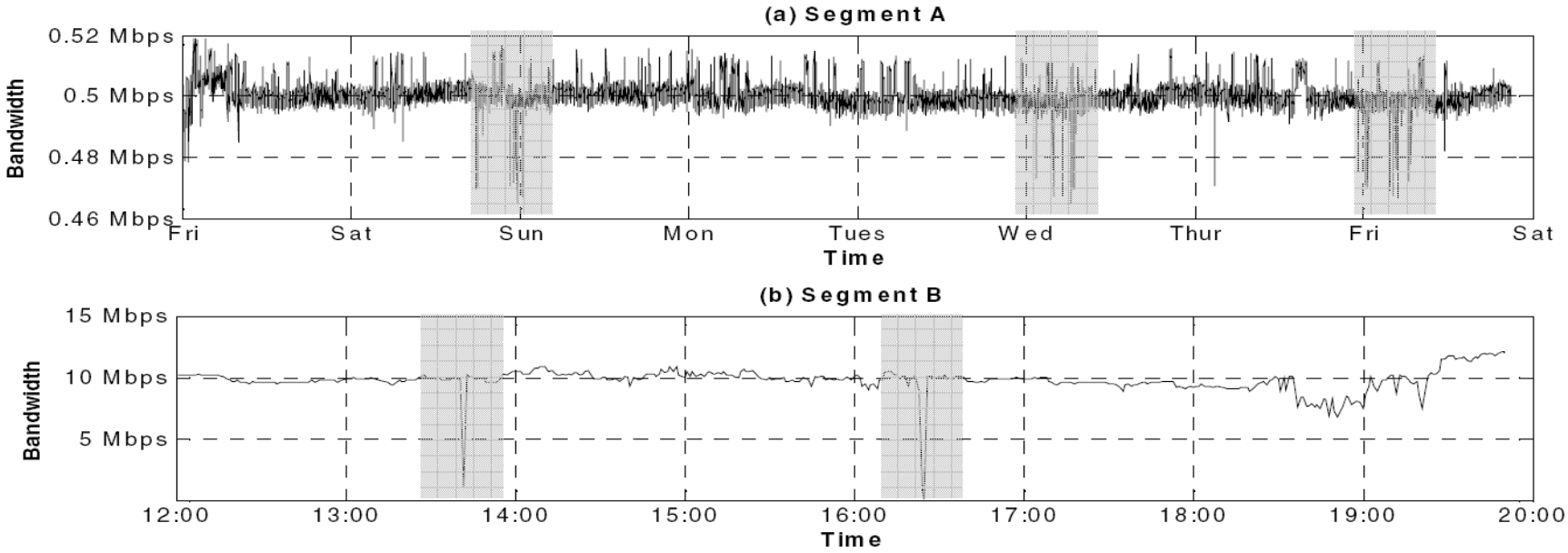
# Publications (3/3)

## ❖ Patents (4) – Filed in Korea

- **Registration Date: 2007-04-16, Registration #: 10-0710047, "Apparatus for Traffic Identification on Internet Protocol Network Environment"**
- **Registration #: 10-2008-0135020, "Method and System for Detecting Error in Process Control Network", FILED in Dec. 2008.**
- **Registration #: 10-2008-0133944, "Method and Apparatus for Predicting Error in Process Control Network", FILED in Dec. 2008.**
- **Registration #: 10-2009-0046093, "Signature Generation Apparatus for Network Behavior of Applications, Collection Server, Detection System for Network Behavior, and Signature Generation Method for Network Behavior", FILED in May 2009.**

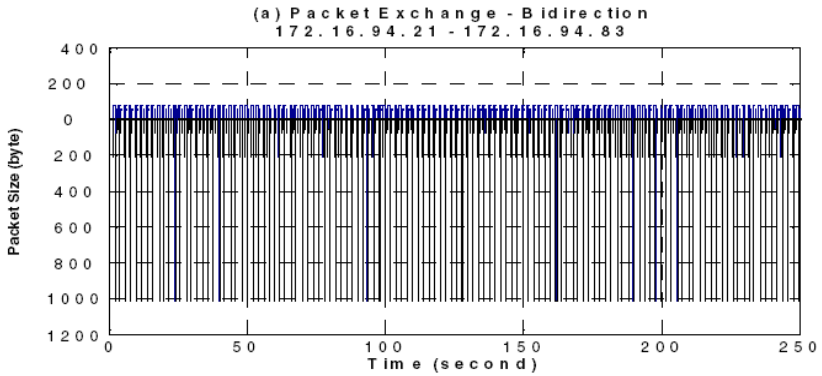
# APPENDIX A.

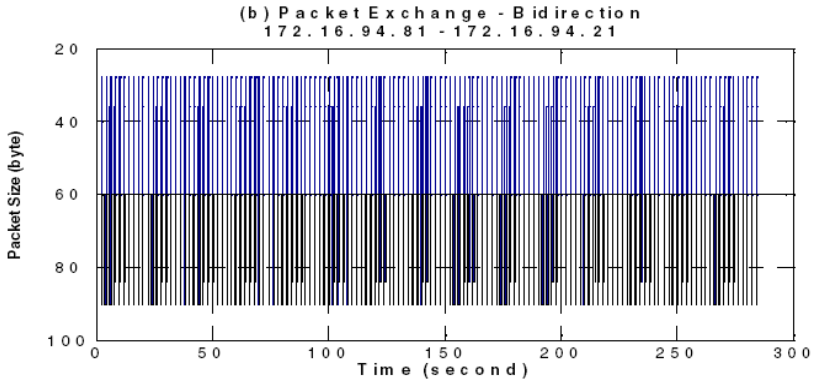
# Bandwidth Usage

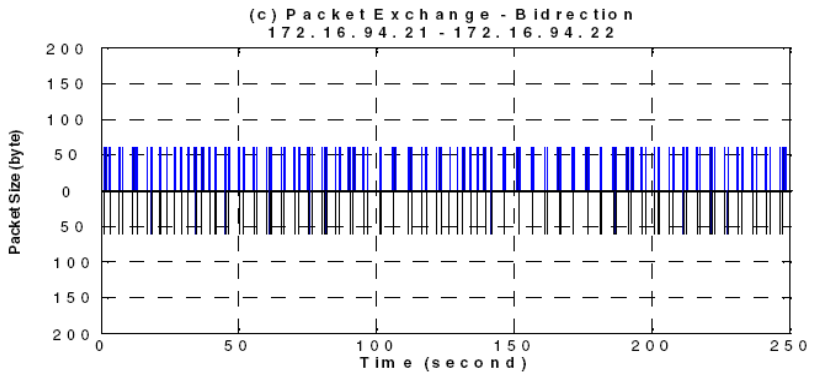


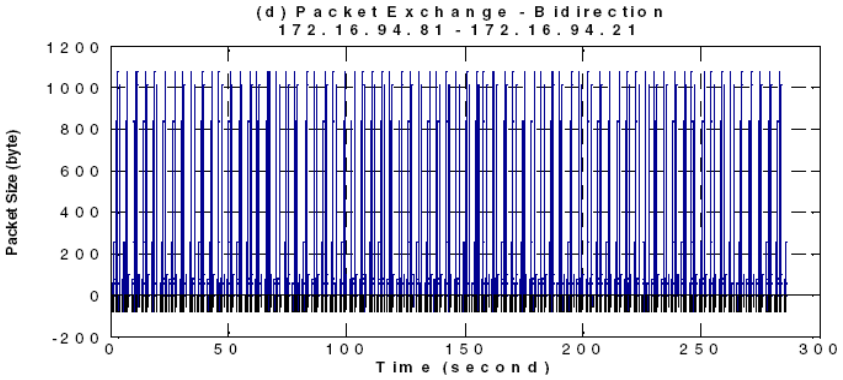

- **Bandwidth usage is not bound to time-of-day effect**
  - **Steady & Predictable, a precise prediction of traffic growth may be possible**
  - **Its volume is strictly proportional to the number of devices in the network**
  - **Fixed pattern in every session occupying the network?**

# Traffic Cycle – Arrival Sequence (1/2)



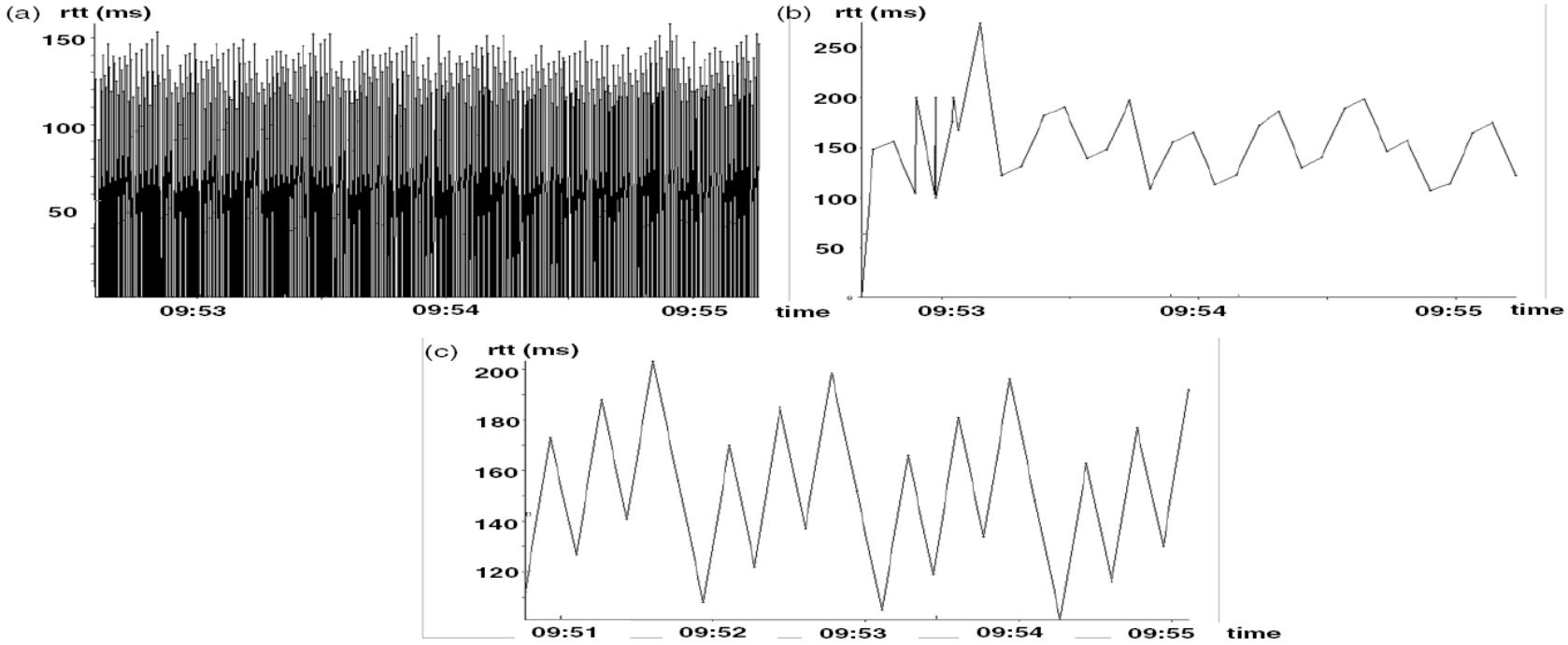




- **Bidirectional packet arrival sequence for PLC-PC sessions**
  - A unique and regular cycle of dense and sparse regions
- **Packet inter-arrival time**
  - It ranges from 120 ms to 1.5 sec.
  - Hundreds of millisecond or above are tolerable

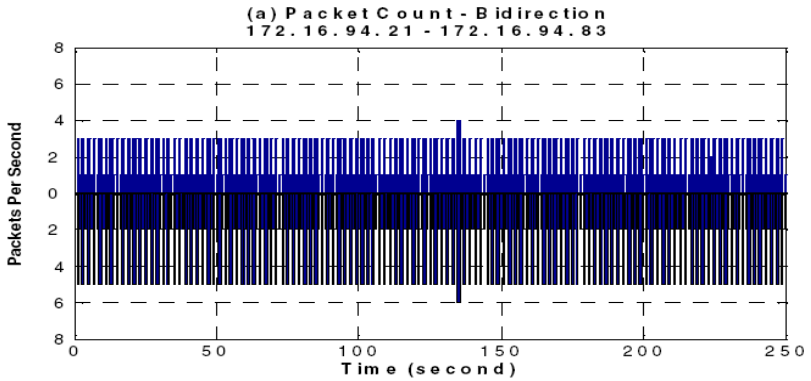
# Traffic Cycle - RTT (2/2)

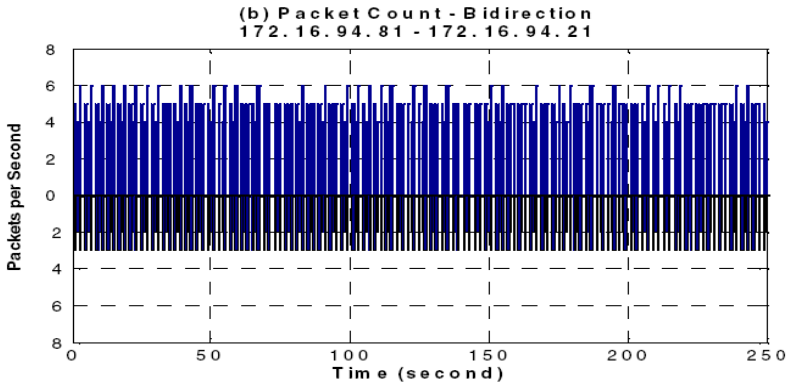


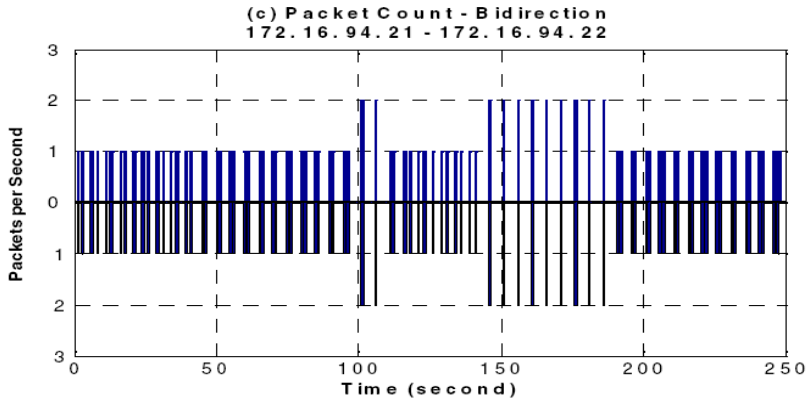

- **A clear periodic pattern exists in RTT measurement**
  - **The measured RTT ranges from 150 to 250 msec**
  - **The recurring shape of the graphs shows again fixed packet arrival sequence**

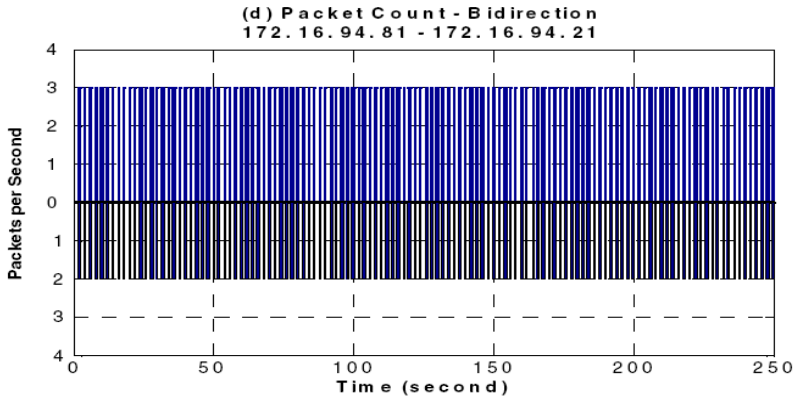
- **It is important to recognize such pattern information beforehand**
  - **A knowledge base for determining 'irregularity' of communication patterns**
  - **Removing ambiguity from definition of anomaly in control IP networks**

# Traffic Symmetry



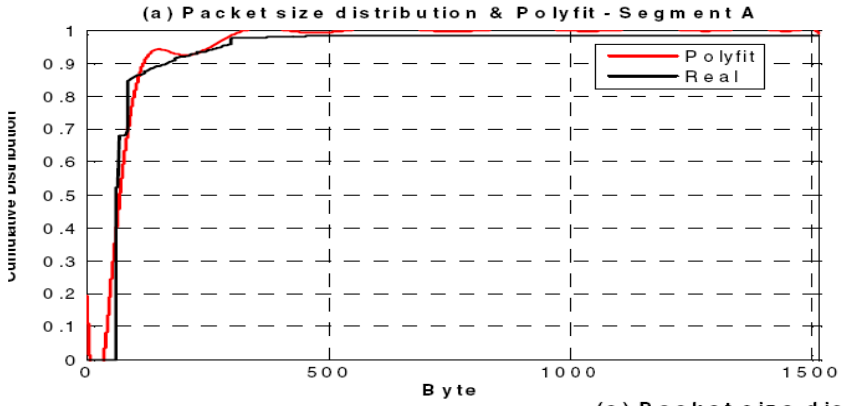



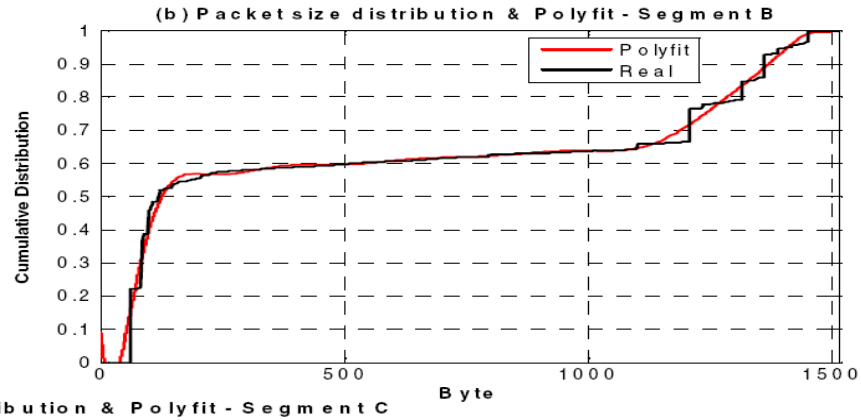



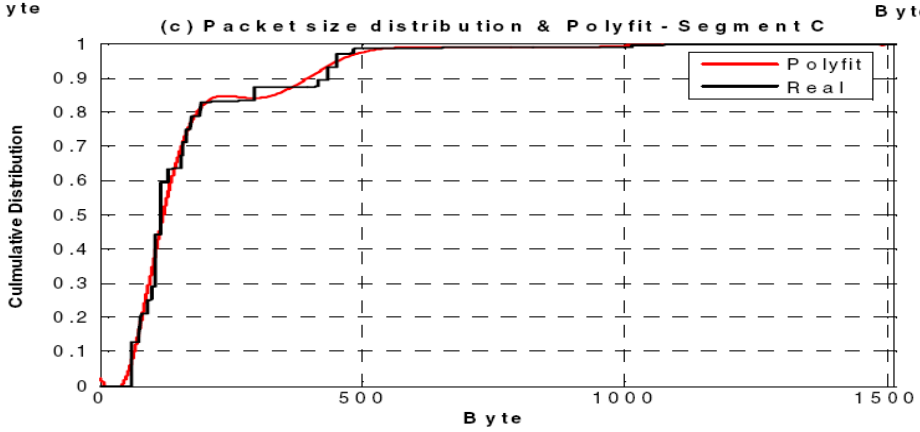




- **It follows a similar trail of HTTP behavior**
  - **But, the request object (or service) and the corresponding reply here are very much fixed in size and repetitive**
  - **Packet counts in the upper & lower planes are almost identical in all cases**
  - **Less than 10 packets per second ratio**

# Packet Size Distribution



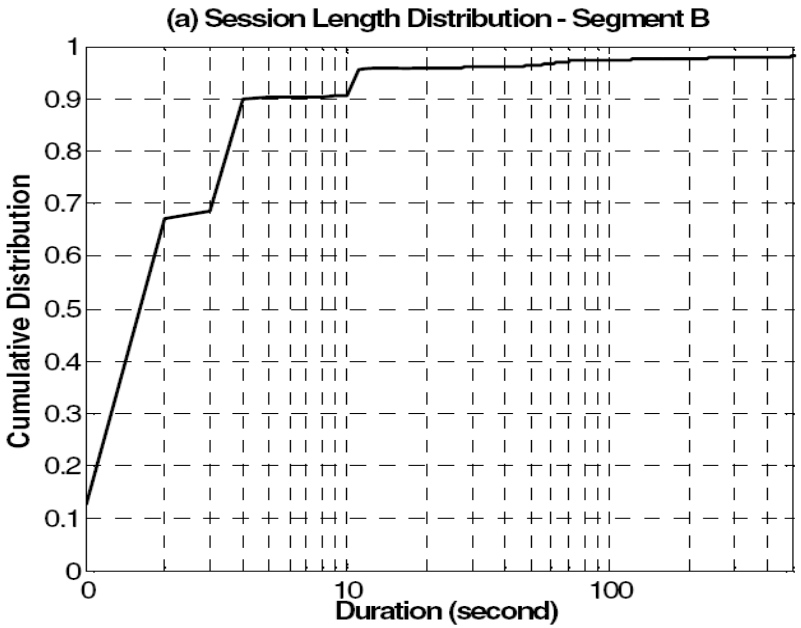



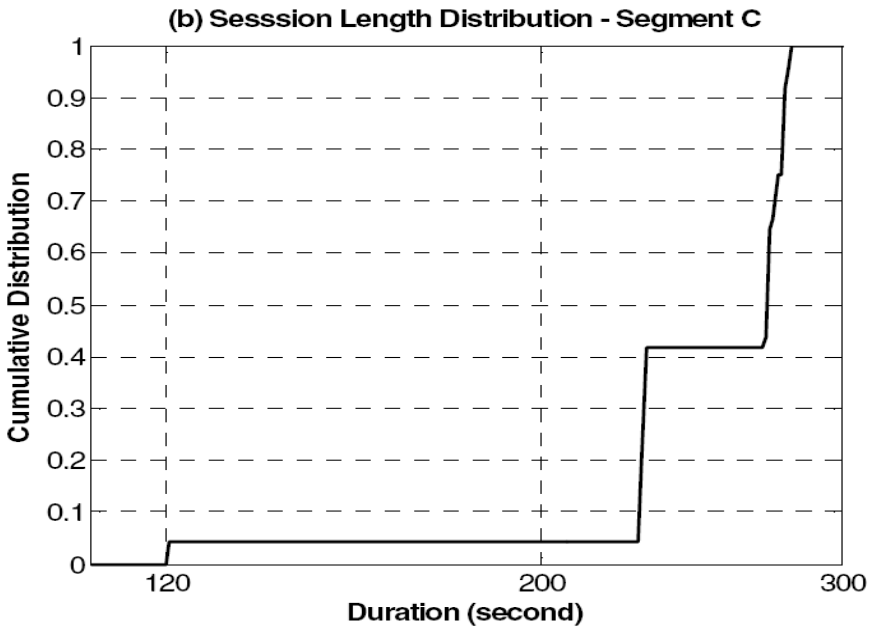




## ❖ Packet types

- Signaling for PLC, HMI display, and management purposes
- 90% of packets range from 60 to 80 byte and are for signaling purpose
- Packet size distribution may vary between backbone and edge: (a), (c) vs. (b)
  - HMI-related packets involves in graphical representation of status

# Session Length Distribution

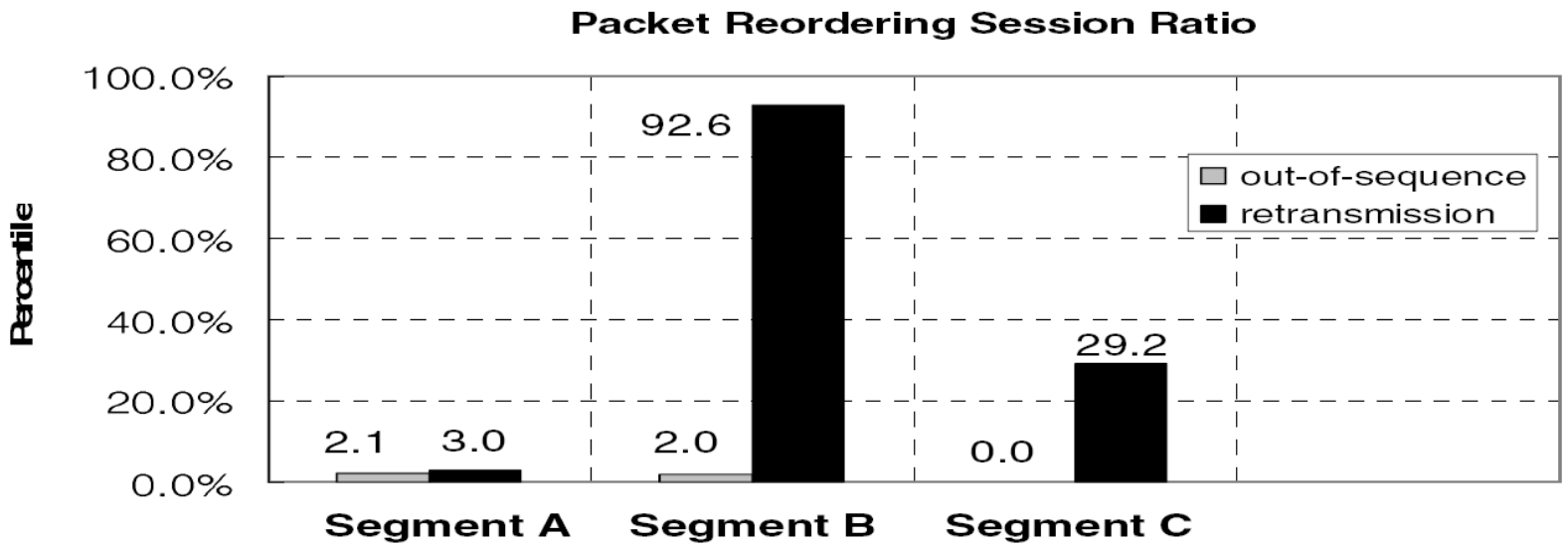



- **Two concrete patterns**
  - **(a) Short session length at the backbone, PC-related**
    - **Mixture of PC-PLC and PLC-PLC sessions**
    - **90 % of flows terminates less than 10 sec.**
  - **(b) Long session length at the end host, PLC-related**
    - **PLC-PLC sessions within the local segment tends to stay longer.**

- **Continuous and repetitive tasks based on action-trigger behavior**
  - **Lighter commands + Corresponding action trigger at the bottom**

# Packet Reordering



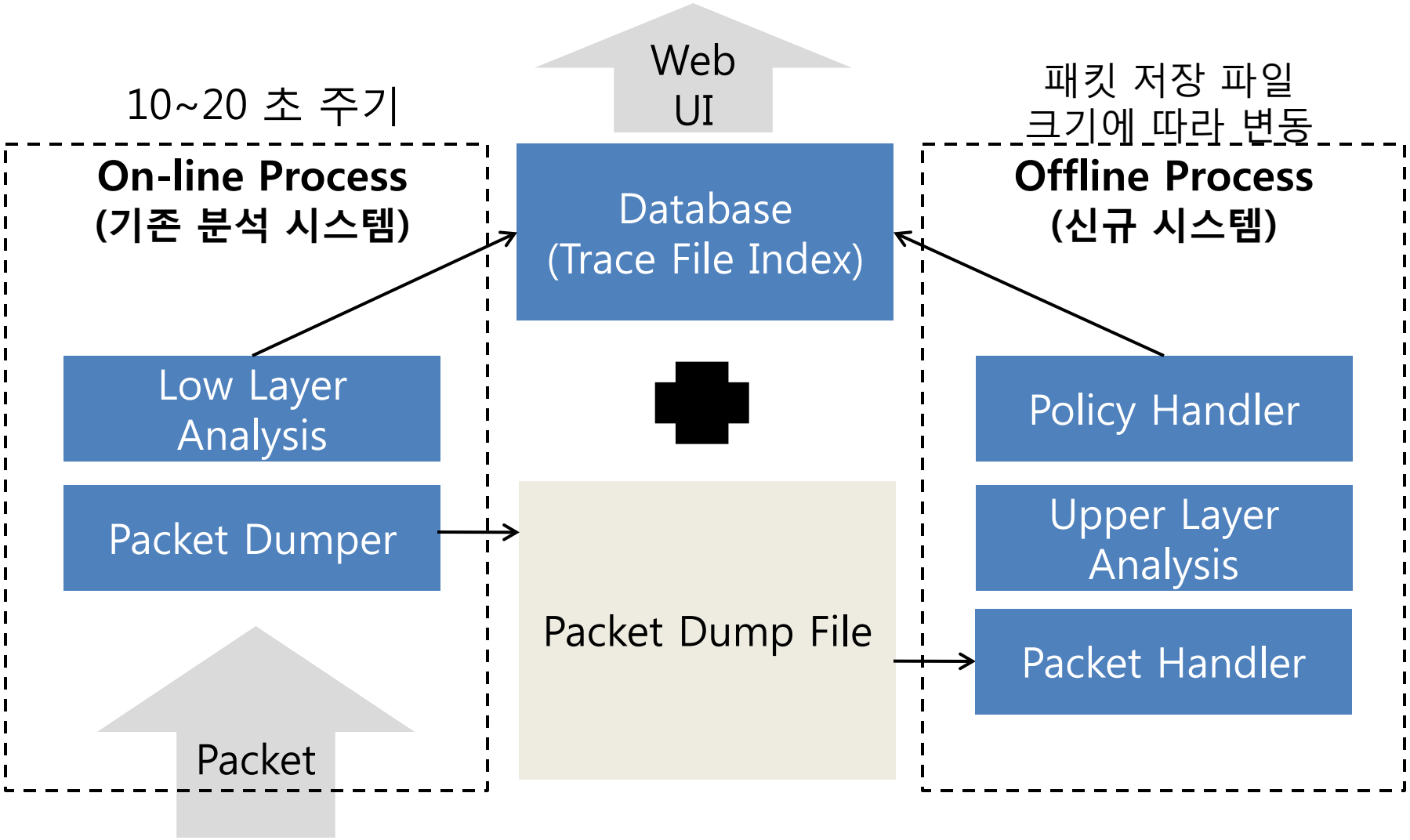

- **Measuring out-of-order packet delivery and retransmission occurrences in every session**
  - **At B, 92% of sessions encounter retransmission packets at least once while online**
  - **Despite of the result, no major operational difficulty was reported during the time of data collection**

- **Yet, it is not clear whether such phenomenon is bound to this particular case**
  - **A further investigation is needed for future work**

# APPENDIX B.

# 장애 시스템 재 설계

## ❖On/Offline 병합구조



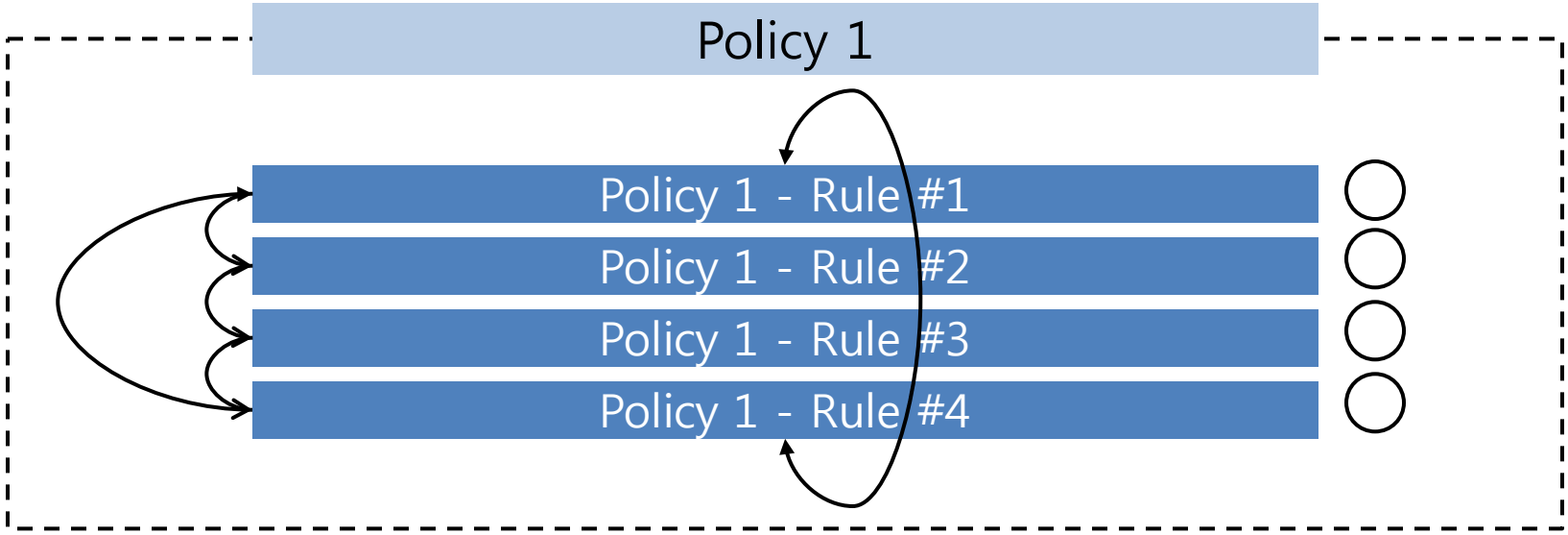
# Policy

## ❖ Policy 정의

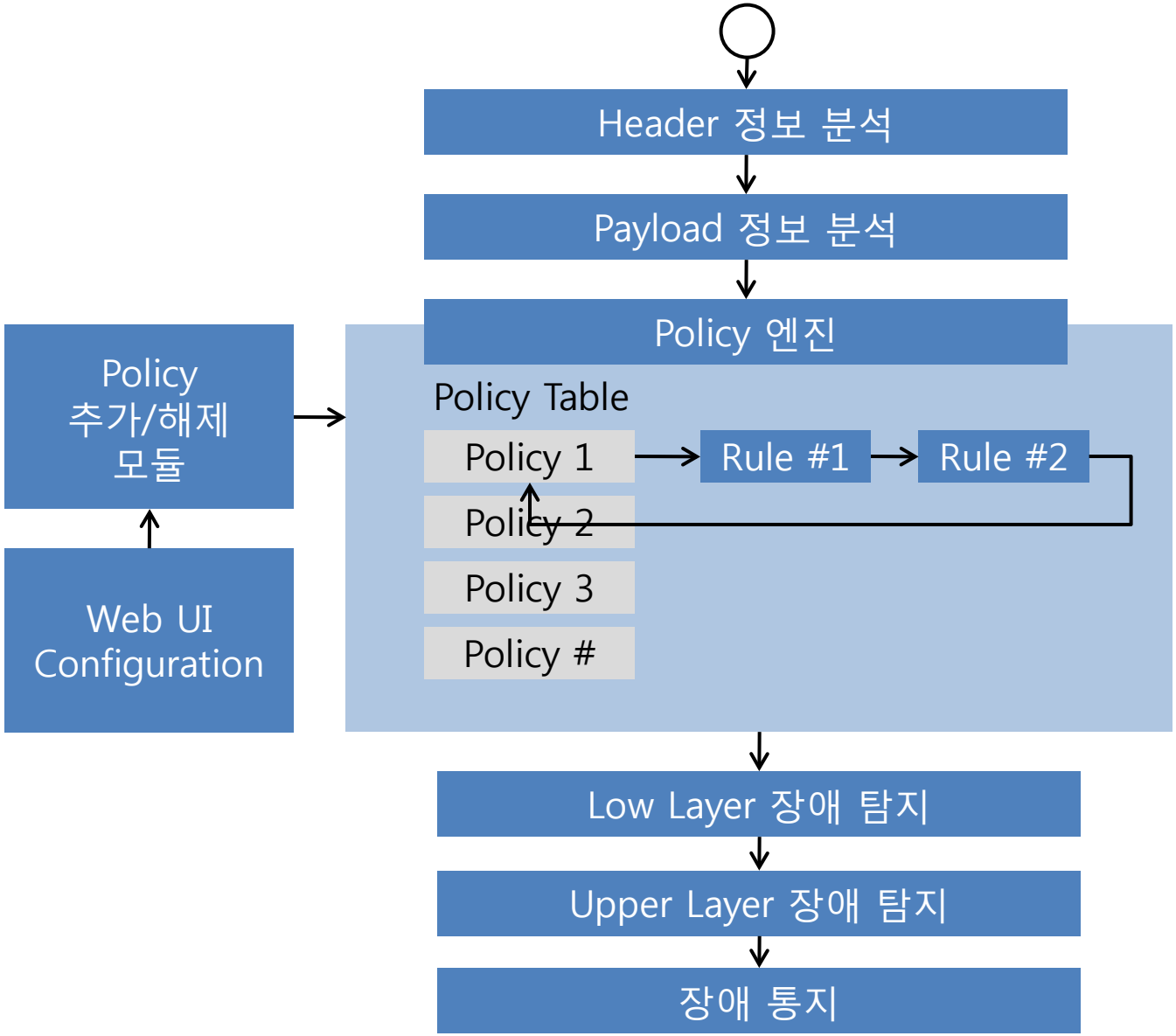
- 응용레벨의 장애 패턴의 표현을 위한 하나 이상의 Rule 의 집합체로써, 각각 또는 연속된 Rule 의 위반여부에 따른 장애 판단의 기준을 명시

## ❖ Policy 구성 요건

- 하나 이상의 Rule 로 구성된다
- Policy 에 속하는 Rule 은 순차적 적용 순서를 가진다



# Policy 적용 프로세스/엔진




Packet

Layer 1~4 패킷 분석
기존 시스템

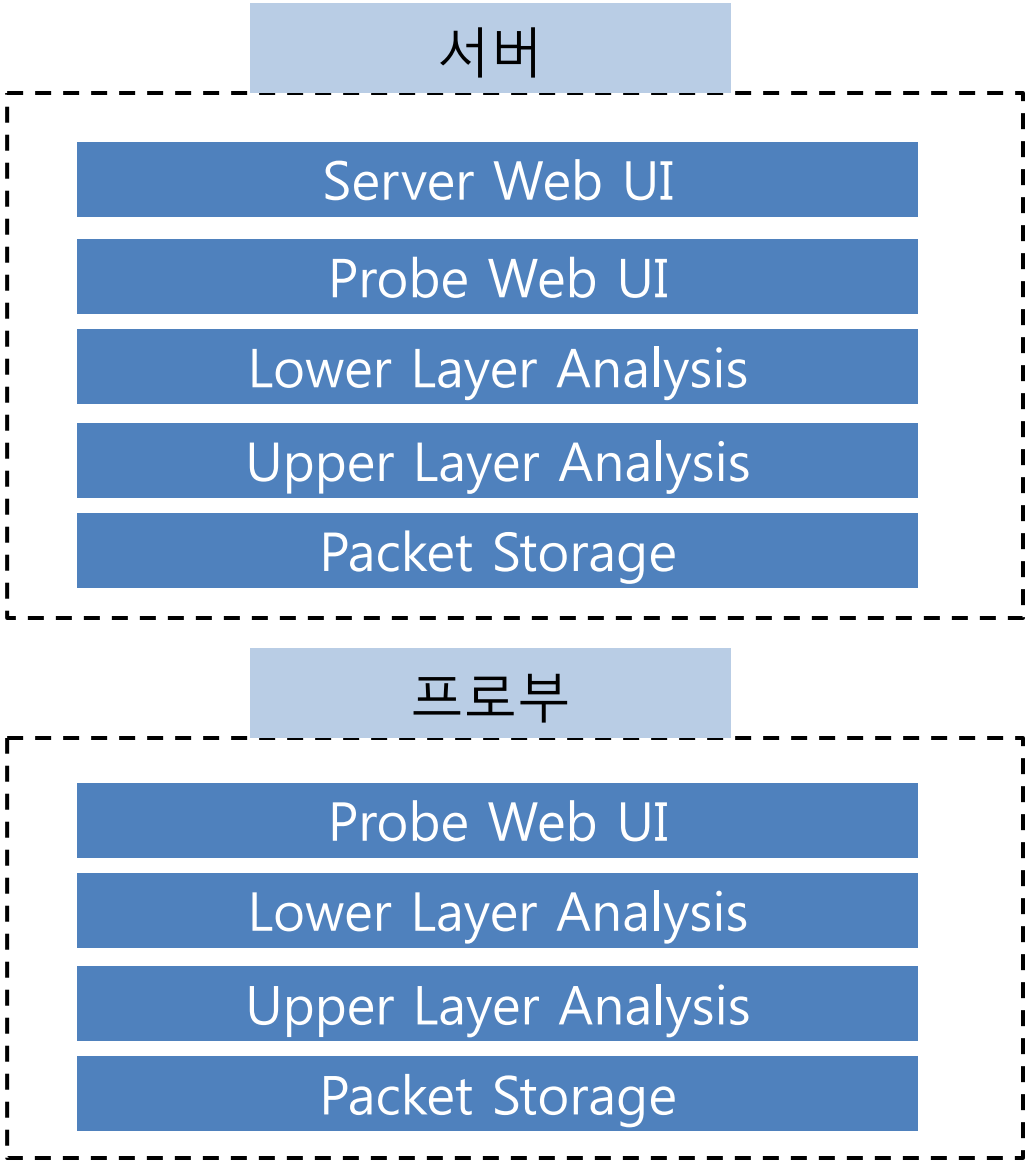
Layer 4~ 패킷 분석
신규 시스템

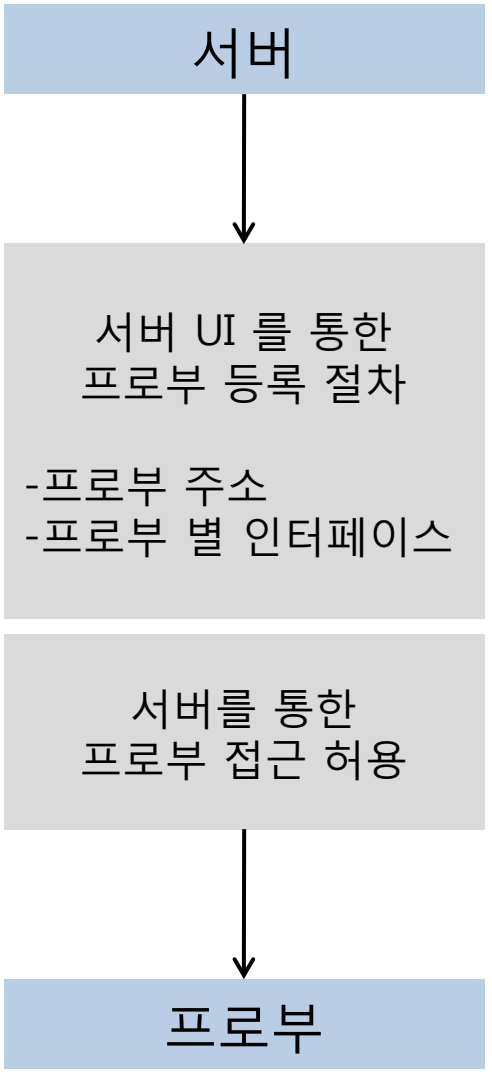
Layer 1~4 장애 탐지

Layer 4~ 장애 탐지

Alarm 생성

# 분산화를 위한 시스템 재 설계



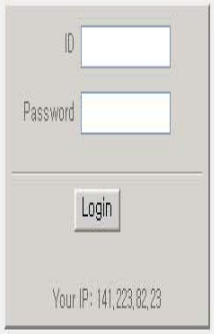



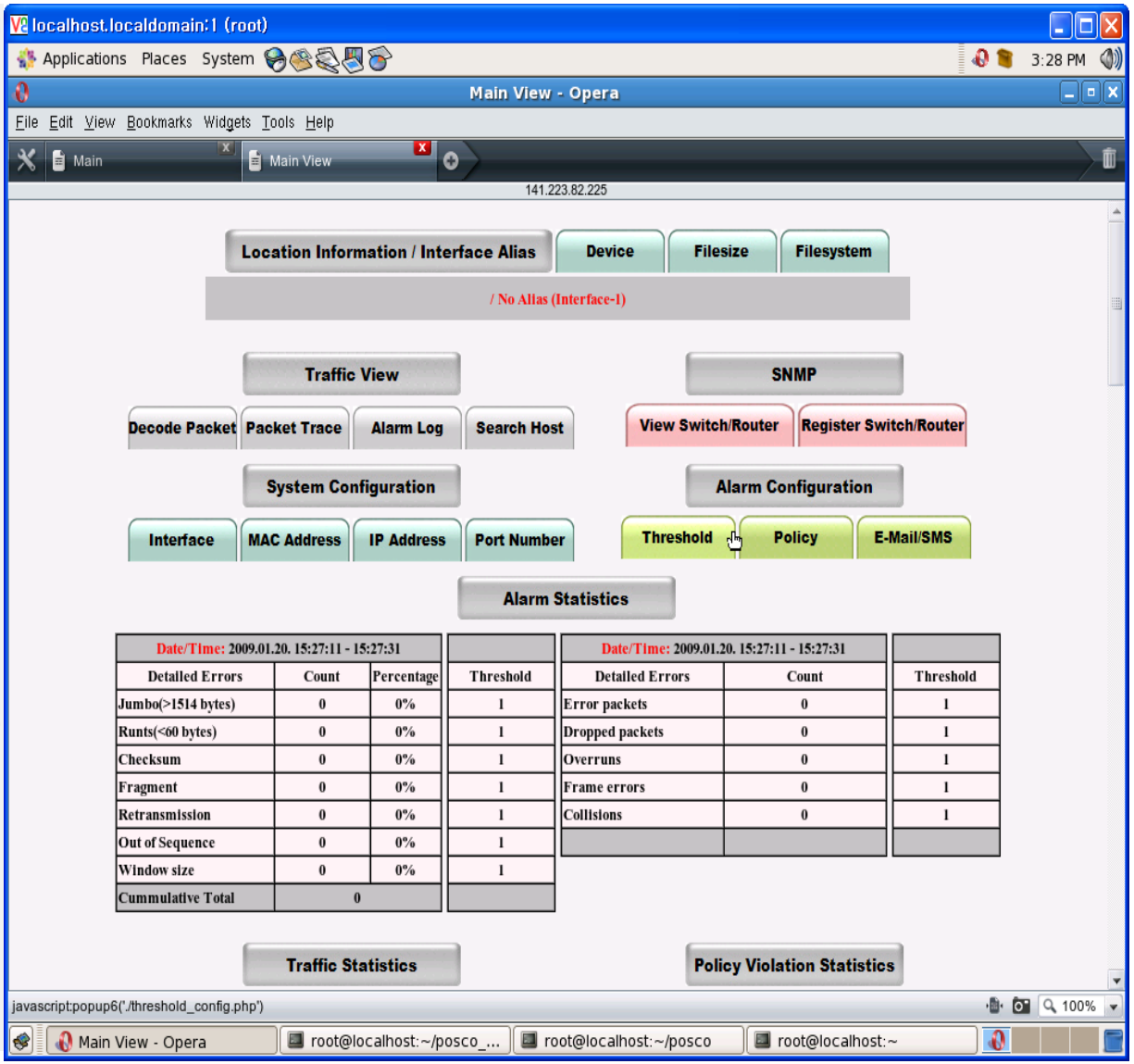
# Screenshot (1/5)

POSTECH 2020

ARACHNE

Adaptive Remote Analyzer for Comprehensive & Hierarchical POSCO control NEtworks

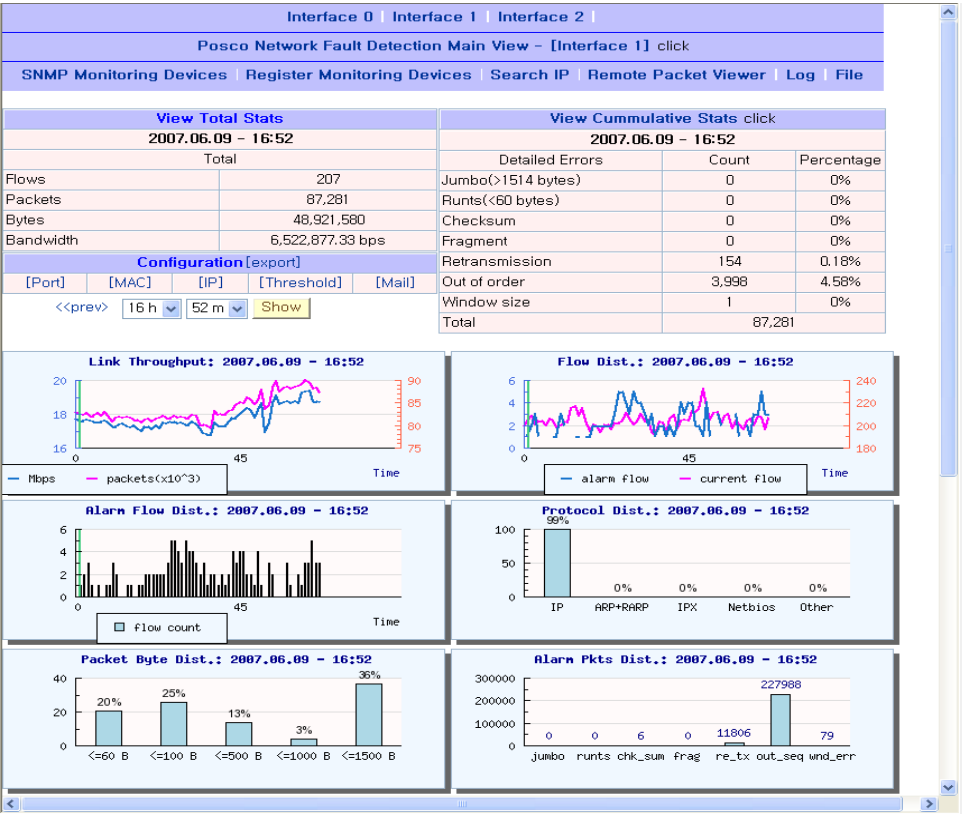


# Screenshot (2/5)

POSTECH 2020

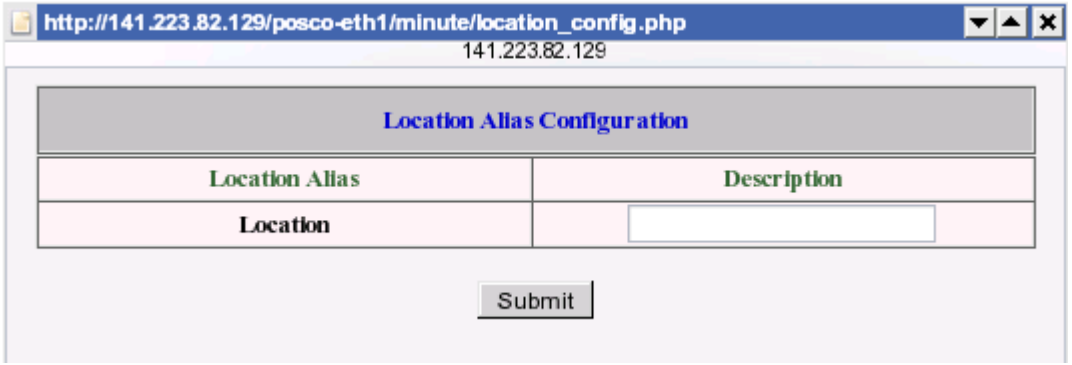
## ❖ 그래프 및 알람 정보 화면

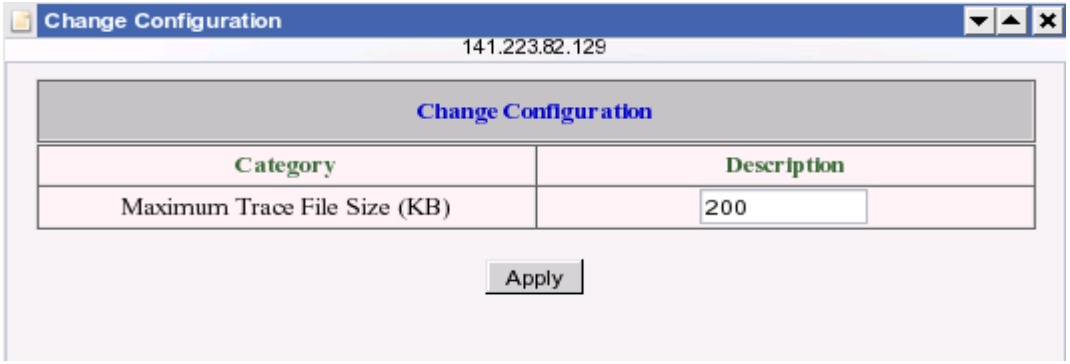


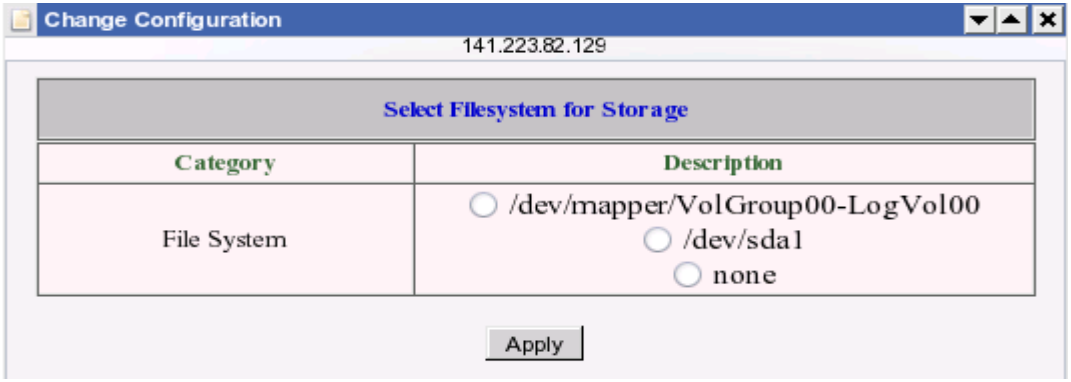

# Screenshot (3/5)

POSTECH 2020

❖ 시스템 Configuration 화면



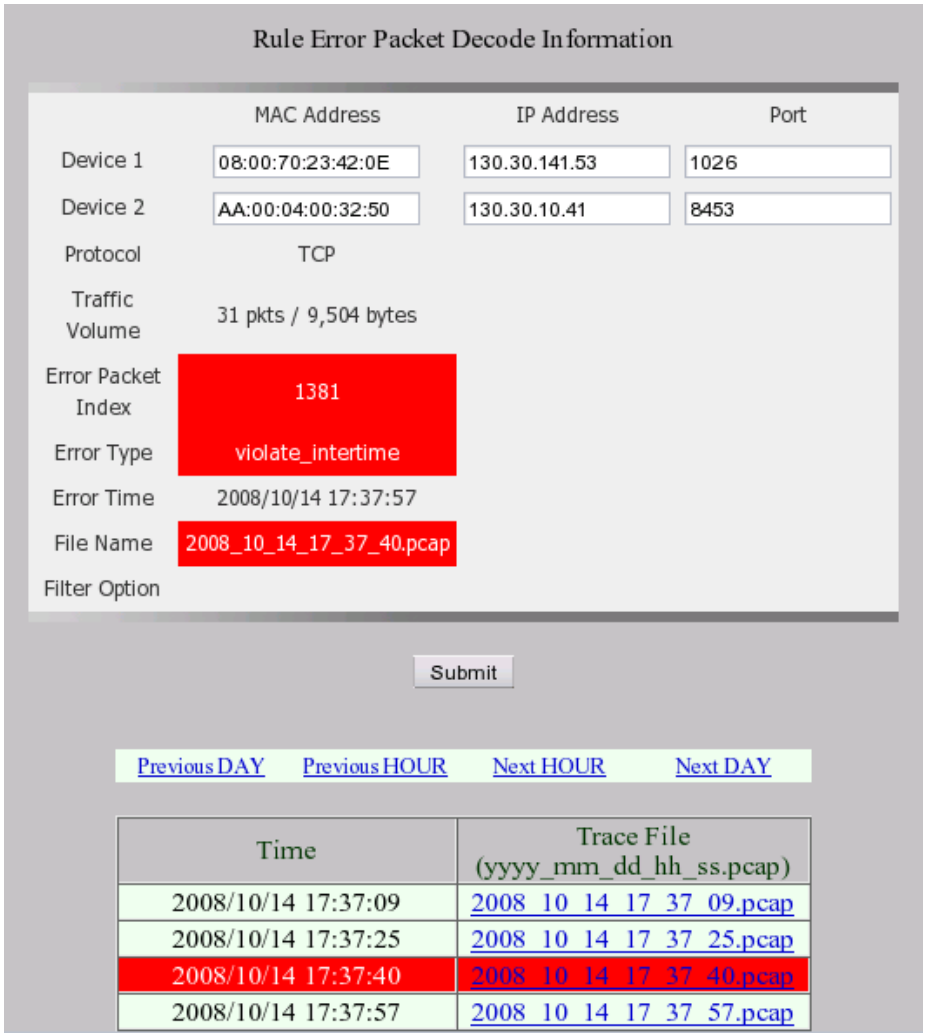



# Screenshot (4/5)

POSTECH 2020

## ❖ 패킷 디코딩 화면 & Policy 위반 패킷 디코딩 화면

# Screenshot (5/5)

❖ Policy 위반 플로우 & Policy 등록 화면

**Policy Violation Flows**

[Top10] [Top20] [Top40] [Top60] [Top80] [top100] [all]

| Time | Src IP | Dst IP | Src Port | Dst Port | Protocol | Pkts | Bytes | Alarm | Decode |
|---|---|---|---|---|---|---|---|---|---|
| 2008/10/14 17:37:57 | 130.30.141.53 08:00:70:23:42:0E -------- -------- | 130.30.10.41 AA:00:04:00:32:50 -------- -------- | 1026 | 8453 | TCP | 31 | 9,504 | violate_intertime | |
| 2008/10/14 17:37:48 | 130.30.141.53 08:00:70:23:42:0E -------- -------- | 130.30.10.41 AA:00:04:00:32:50 -------- -------- | 1026 | 8453 | TCP | 31 | 9,504 | violate_cycle | |
| 2008/10/14 17:37:46 | 130.30.141.53 08:00:70:23:42:0E -------- -------- | 130.30.10.41 AA:00:04:00:32:50 -------- -------- | 1026 | 8453 | TCP | 31 | 9,504 | out_of_order_rule | |
| 2008/10/14 | 130.30.141.53 08:00:70:23:42:0E | 130.30.10.41 AA:00:04:00:32:50 | 1026 | 8453 | TCP | 36 | 9,906 | violate_intertime | |

Policy Management Information

**Policies**

Choose Policy: reply_error — Delete — ViewAll

**Policy Information**

Policy Name: reply_error — Cycle: 2000

Rule Seq.

Src. MAC Address — Dest. MAC Address

Src. IP Address(port) — Dest. IP Address(port) — Register

Signiture

Offset — Word Size

Rule intertime — Rule nexttime

Regularity

**Rule Lists**

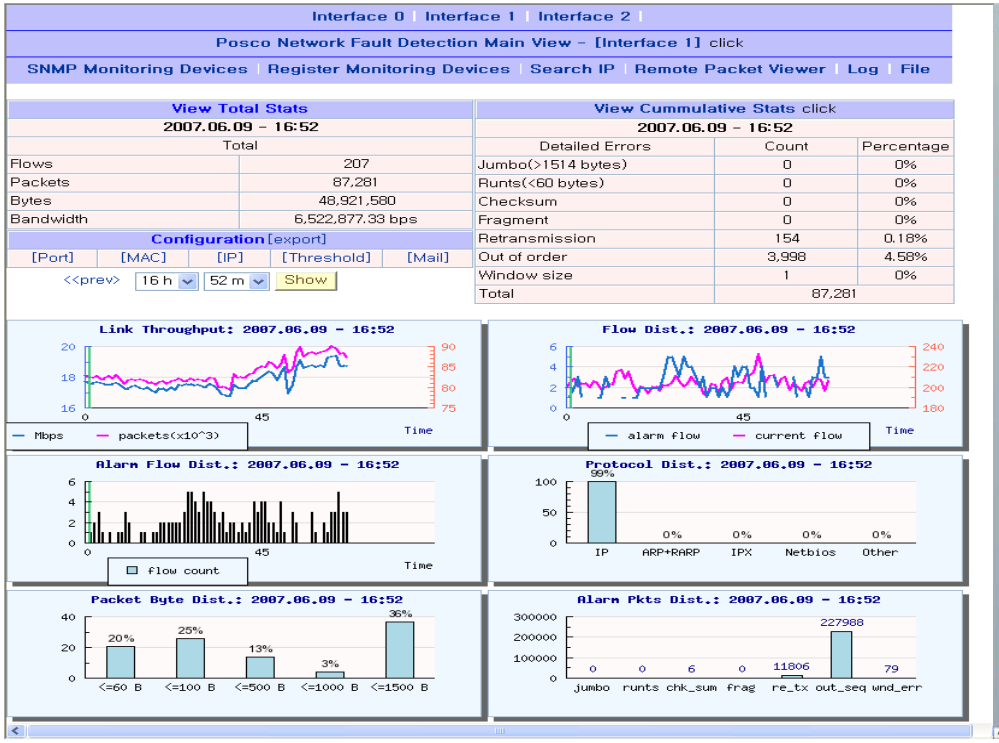
| Rule Seq | SrcMac | DstMac | SrcAddr | SrcPort | DstAddr | DstPort | Signature | Offset | WordSize | Regularity | Rule InterTime | Rule NextTime |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 08007023420e | aa0004003250 | 130.30.141.53 | 1026 | 130.30.10.41 | 8453 | 6000 | 0 | 2 | 0 | 500 | 500 |
| 2 | 08007023420e | aa0004003250 | 130.30.141.53 | 1026 | 130.30.10.41 | 8453 | e000 | 0 | 2 | 0 | 500 | 500 |

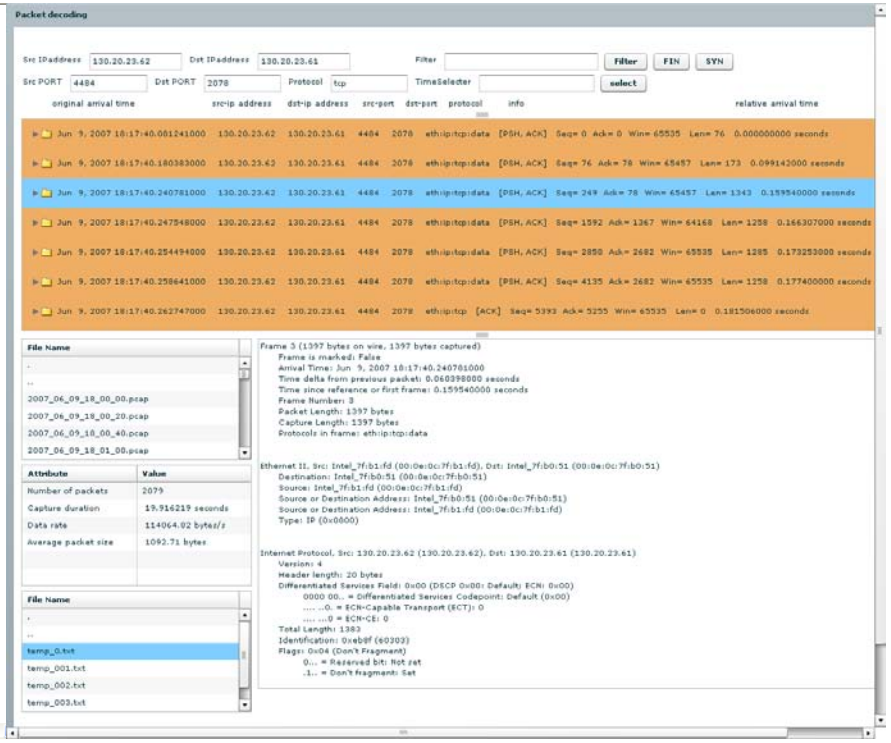
# 시스템 적용

## ❖ 2 후판 공장 적용

- 2007 년 2월 ~ 지속
- 2008 년 1, 3 후판 공장 이동
- 사례) 가열로 PLC 통신 전문 송신 이상 분석.조치 (2007년 10월 24일) 외 2건
- 테스트 용 트래픽 다수 수집

## ❖ Screenshots– System UI



# 시스템 – 소프트웨어 환경

POSTECH 2020

❖ **Open source, GNU 라이선스 product 의 활용**

- 확장성 고려 및 추가 설치 비용의 최소화를 위한 대비
- 검증 된 프로그램의 활용
- OS 교체 실시
- 분산 서버 설치 시 추가 라이선스 비용 없음

| | 이름 | 라이선스 | 비고 |
|---|---|---|---|
| OS | CentOS 5, RHEL 4/5 | GNU<br>Commerial | Dell Server<br>Dell Laptop |
| 개발환경 | Linux 2.6.x, gcc-3.4.x | | |
| 개발언어 | C, PHP, Html, Perl | | |
| 필요패키지 | httpd-2.2.2<br>php-5.2.5<br>mysql-5.0.27<br>jpgraph-1.13<br>mrtg-2.15.0<br>net-snmp-5.1.2<br>net-snmp-utils-5.1.2<br>wireshark | GNU | |

# 시스템 – 하드웨어 환경

POSTECH 2020

- **타워 형 분석 서버 (분산형태) 2 기**
  - 설치장소: 포스코 3 후판 전산소
  - Dell Optiflex 755/760

- **Aggregator 네트워크 탭 4 기**
  - 필요 UTP 케이블 일체

- **이동 형 분석 서버 1 기**
  - Dell 노트북 Vostro 1510


# Damage Cost Function

## ❖ From the earthquake damage cost function

$$E[C_d] = \int_0^1 \left[ \int_0^\infty C_d(x) f_{X|L}(x|l)\,dx \right] (P|F, t_l) f_L(l)\,dl$$

$$\approx C_0(x'_{ki}) \cdot \int_0^L \frac{1}{L} e^{-t} dt$$

- ***l*** **is loss impact of a fault category (0-1)**
  - **1 is a complete process failure**
- $C_0$ **is the maximum damage cost (L=1)**
- **Future cost is uniformly distributed where**

$$(P|F, t_l) = \int_0^L \frac{1}{L} e^{-t} dt$$

## ❖ Replacing with our fault prediction model (Weibull distribution)

$$E[C_d] = \sum_0^1 C_d(x) \sum_{n=1}^{\infty} \left[ \int_0^L \frac{n}{L} e^{-t} dt \right] f(t_l; n, \eta)$$

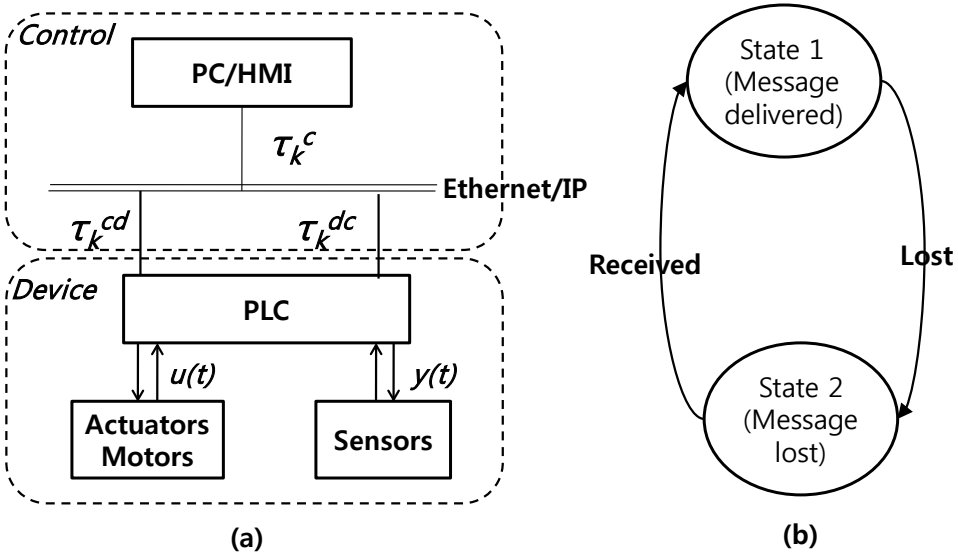
$$= \sum_0^1 C_d(x) \sum_{n=1}^{\infty} \left[ \int_0^L \frac{n}{L} e^{-t} dt \right] \frac{n}{\eta} \left( \frac{t_l}{\eta} \right)^{n-1} e^{-\left(\frac{t_l}{\eta}\right)^n}$$

# Modeling ICN Application-layer Fault



- **Based on state changes regarding the transmission sequence between control and device**

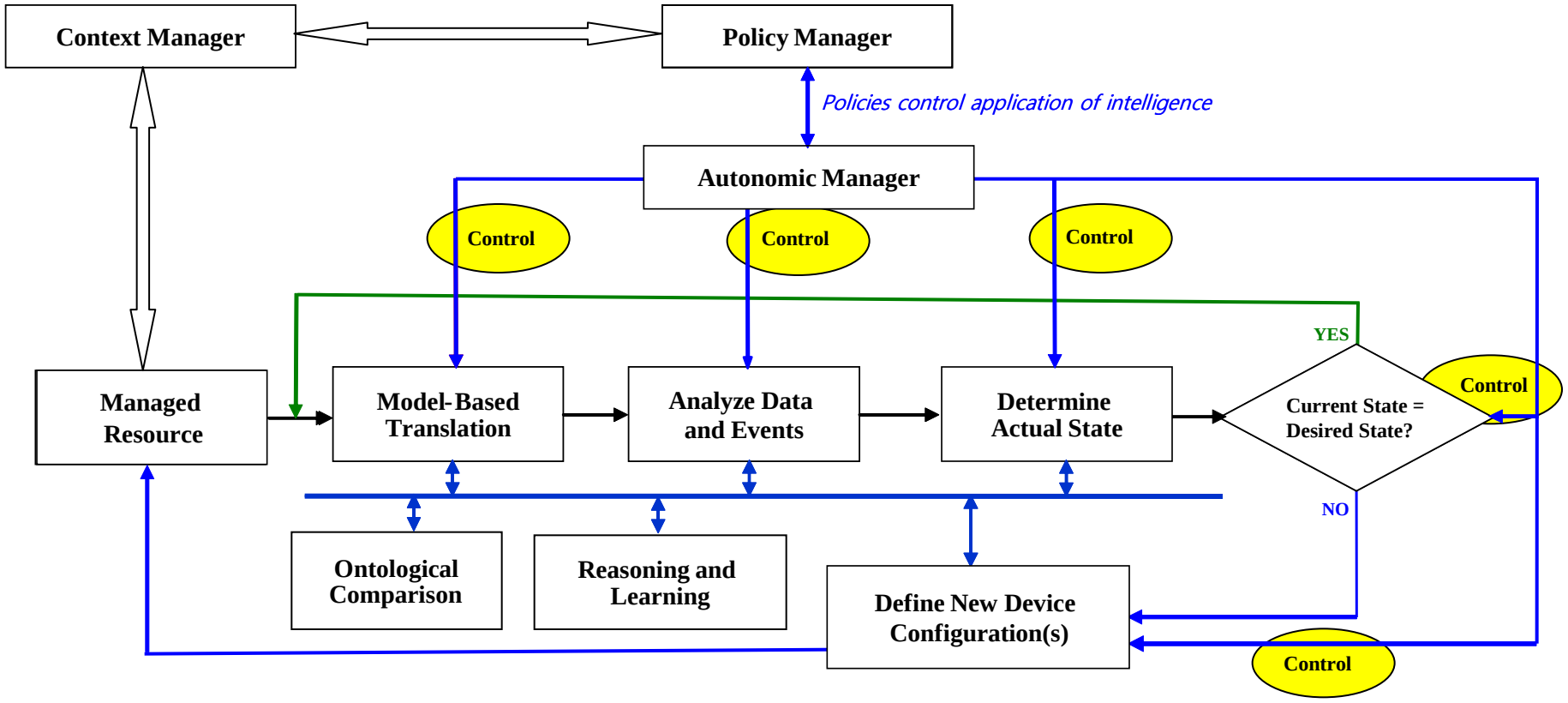
$$kh, (k+1)h, (k+2)h, \dots, (k+n)h$$




$$\max\left\{\frac{1}{2}h-\frac{1}{\kappa h},0\right\}<\frac{\tau}{h}<\min\left\{\frac{1}{\kappa h},1\right\}$$

$$\max\left\{\frac{1}{2}h-\frac{1}{\kappa},0\right\}<\tau<\min\left\{\frac{1}{\kappa},h\right\}$$

# FOCALE Autonomic Architecture



# Autonomic ICN Fault Diagnosis (1/2)

POSTECH 2020

- **Model-Based Translation**
  - Two state model translation – Fault/No Fault
  - Metrics are in good state (threshold and boundary of user rules) – referring threshold table
  - System translation in metrics from user-defined alarm rules
- **Analyze Data and Events**
  - Fault detection algorithms
- **Ontological Comparison – Perhaps omitted from control loop**
  - Metric reorganization, relationship definition
  - A set of alarm cases and relationship to a set of fault cases
- **Reasoning and Learning**
  - ML decisions for reference (only for end-to-end metrics)
  - User opinion as input (for later, prediction engine based on rule occurrence in temporal aspect)
- **Define New Device Configuration (Action)**
  - Counteraction, Instruction to Operator, or Simply a notification
  - E.g. Device isolation, Replacement, Observation (from trouble shooting case study)

# Autonomic ICN Fault Diagnosis (2/2)

POSTECH 2020

## State Decision

- Two states: Non-fault and Fault (ON/OFF model)
- State of alarm, network fault, failure

## Autonomic Manager

- Deliver user-defined rule and handling periodic updates of rules
- Deliver a notification from analysis module to State Decision
- Handle a minor system configuration (location name change, etc.)
- Feed user opinions to reasoning module

## Policy Manager

- User input via Web
- Define conditions for pre-cautionary alarms

## Context Manager

- Network metrics, Components of rule definition: cyclic information, temporal aspect

## Managed Resource

- PC, PLC, Devices, Detection system itself (capturing probe)