

## Multi-Model Based Incident Prediction and Risk Assessment in Dynamic Cybersecurity Protection for Industrial Control Systems

Zhang Qi  
qiqi@hust.edu.cn  
October 15, 2015



Automation School,  
Huazhong University of Science and Technology,  
Wuhan.

Hello everyone, my name is Zhang Qi, and I am the Ph.D student of Professor Zhou Chunjie. I am very glad to be invited by Professor Yang Shuanghuang to make a presentation about my recent research.

My research interests are related to risk assessment and decision-making for industrial control systems. The title of my presentation is “Multi-Model Based Incident Prediction and Risk Assessment in Dynamic Cybersecurity Protection for Industrial Control Systems”.

### Outlines

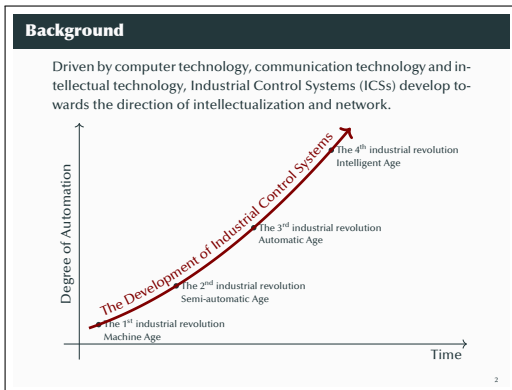
- Introduction
- Architecture
- Hazardous Incident Prediction
  - The Bayesian Network Based Knowledge Modeling
- Incident Prediction
- Dynamic Risk Assessment
  - Decouple of Incident Consequences
  - Classification of Incident Consequences
  - Quantification of Incident Consequences
  - Calculation of Dynamic Risk
- Simulation
  - Simulation Platform
  - Simulation and Result Analysis
- Conclusion and Prospect
  - Conclusion
  - Prospect

My presentation is separated into six parts:

- Firstly, I will introduce the background and the problems of risk assessment for industrial control systems.
- Secondly, I will give the architecture of our risk assessment solution for industrial control systems.
- Thirdly, I will elaborate the detail of our method.
- Then, I will show you the effectiveness of our approach by using a numerical simulation.
- At last, I will discuss the problems of our approach and introduce the future works.

## Introduction

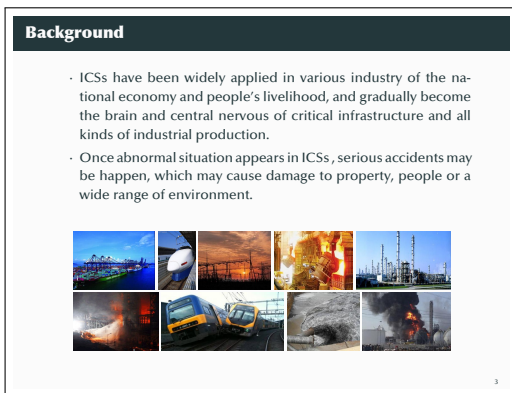
In this part, I will introduce the development history and the cybersecurity issues of industrial control systems. And, I will compare the cybersecurity issues of industrial control systems and traditional IT systems.



There are four great changes in the development of industrial control systems:

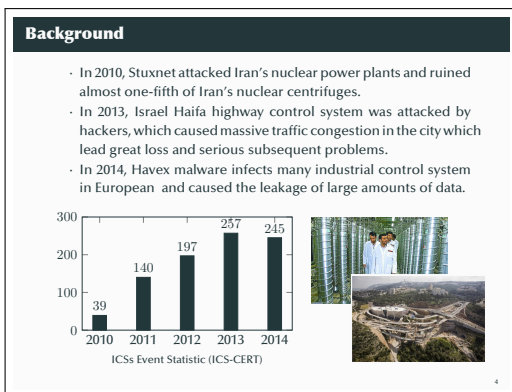
- Machine Age
- Semi-automatic Age
- Automatic Age
- Intelligent Age

From this figure, we can see that with the development of industrial control systems, the degree of automation is increasing. Intelligence and networking are the development trend of industrial control systems.



Nowadays, the industrial control systems have been widely applied in various industry, and they are becoming more and more important for the national economy and our life.

As mentioned before, the industrial control systems are evolving towards intelligence and networking. The rapid development of the industrial control systems reduce the difficulty of the development and the cost of construction, on the other hand, it has also introduced the cybersecurity issues into the industrial control systems.



For example, in 2010, the Stuxnet attacked Iran's nuclear power plants and ruined almost one-fifth of Iran's nuclear centrifuges. In 2013, Israel Haifa highway control system was attacked by hackers, which caused massive traffic congestion in the city which lead great loss and serious subsequent problems.

According to the statistical data from "Year in Review 2014" published by the ICS-CERT which is short for "Industrial Control Systems Cyber Emergency Response Team", the number of attacks for industrial control systems increases year by year. In 2010, there were

only 39 security incidents of industrial control systems, but in 2014, this number has grown to 245.

Unlike traditional IT systems, the security incidents of industrial control systems can cause irreparable harm to the physical systems being controlled and to the people dependent on them. Basically, protecting industrial control systems against cyber-attacks is vital to both the economy and stability of a nation. Therefore, the cybersecurity issue of industrial control systems must be taken seriously and solved as

soon as possible.

#### Problems – Timeliness and Availability

ICSs have rigorous requirements on timeliness and availability. The cybersecurity risks of ICSs are primarily from the potential loss caused by the cyber-attacks which demolish the timeliness and availability of the control system.

In order to achieve the destructive purpose, attackers generally need to follow part or all of these three steps:

1. infiltrate into the field network,
2. invalidate the system functions,
3. cause the hazardous incidents.

Therefore, the cybersecurity risk assessment of ICSs needs a novel and targeted risk model to analyze the risk propagation.

In recent years, considerable researches have been undertaken to study cybersecurity risk assessment methods. However, the cybersecurity risk assessment in the IT domain is not entirely applicable to industrial control systems because industrial control systems are relatively different from traditional IT systems in some aspects.

Firstly, the cybersecurity objectives are different. Traditional IT systems first require an ensuring of confidentiality, then integrity, and finally availability. For industrial control systems, in contrast, the priorities of

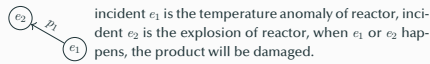
these three security objectives are first availability, then integrity, and finally confidentiality, because the timeliness and availability are the primary concerns. The malicious attacks induce the cybersecurity risk to industrial control systems by demolishing the timeliness and availability. Therefore, the risk assessment of industrial control systems needs a novel risk propagation analysis approach.

#### Problems – Overlapping amongst Consequences

The majority of existing quantitative risk assessment approaches used the following definition to calculate the risk  $\mathcal{R}$ .

$$\mathcal{R} = \sum_i S(e_i)P(e_i)$$

However, the overlapping amongst difference consequences may cause the error of risk value. For example,



Assume that  $P(e_1) = 1$ , so  $P(e_2) = p_1$ , then

$$\mathcal{R} = S(e_1) + p_1 S(e_2) = S(e_1) + p_1 S(e_1) = (1 + p_1)S(e_1) \geq S(e_1).$$

The majority of existing quantitative risk assessment approaches used this definition to calculate the risk, where  $S(e_i)$  is the severity of the incident  $e_i$  and  $P(e_i)$  is the probability of the incident  $e_i$ .

It is also worth noting that there is a problem when this definition is used in industrial control systems risk assessment. This is due to the fact that, for industrial control systems, different hazardous incidents may cause the same consequence; whereby, using this definition to assess risk will cause the severity of the same consequence to be accumulated multiple times. As a

result, there is an error in the risk assessment, which cannot be ignored. Even worse, the decision-making may generate a wrong policy with this inaccurate risk value.

For example, incident  $e_1$  is the temperature anomaly of reactor, incident  $e_2$  is the explosion of reactor, when  $e_1$  or  $e_2$  happens, the product will be damaged. Assume that  $P(e_1) = 1$ , so  $P(e_2) = p_1$ , then

$$\mathcal{R} = S(e_1) + p_1 S(e_2) = S(e_1) + p_1 S(e_1) = (1 + p_1)S(e_1) \geq S(e_1).$$

It is obviously wrong, because the risk of system can't be larger than the total value of all assets.

#### Problems – Unknown Attacks

Many ICSs run 24/7/365, and therefore the updates must be planned and scheduled days or weeks in advance. After the updates, exhaustive testing is necessary to ensure the high availability of the ICS.

This leads to inability of the attack knowledge of ICSs to be updated in time. Several attack knowledge-based risk assessments cannot work well on ICSs.

Therefore, the risk assessment should have the ability of assessing the risk caused by unknown attacks without the corresponding attack knowledge.

As continuous operation systems, the industrial control systems cannot tolerate frequent software patching or updates. This causes the database of attack signatures to lag far behind the rapid development of attacks. With this defect, several intrusion detection system based misuse detections would miss the unknown attacks.

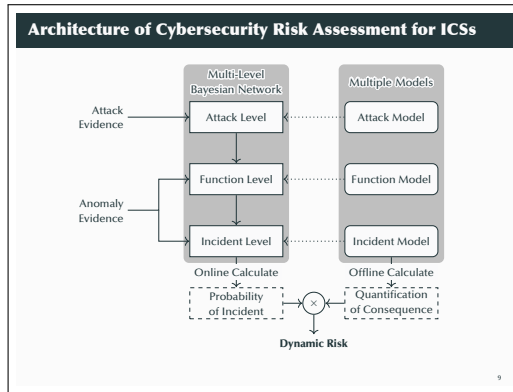
On the other hand, without the information about unknown attacks, such as purposes, consequences, and further steps, these unknown attacks and their consequences cannot be predicted accurately. As a result,

the risk assessment module will generate erroneous risk value, which may lead to a wrong decision.

#### Architecture

Based on the above analysis, the requirements of cybersecurity risk assessment for industrial control systems can be summarized. The risk assessment of industrial control systems needs:

- a novel and targeted risk model to analyze the risk propagation,
- a unified quantification approach to calculate the risk quantitatively without the error caused by overlapping amongst consequences,
- the ability of assessing the risk caused by unknown attacks without the corresponding attack knowledge.



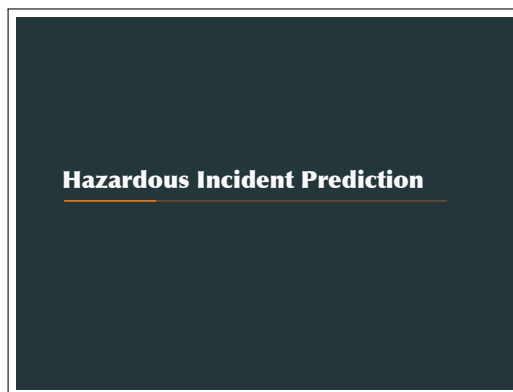
To meet the requirement of the risk assessment for industrial control systems, a dynamic cybersecurity risk assessment based on the multi-model is proposed.

To analyze the propagation of cybersecurity risk, the attack model, the function model, and the incident model are considered. Then, these three models are converted into a multi-level Bayesian network. This Bayesian network has three levels: the attack level, the function level, and the incident level.

There are two kinds of inputs for the dynamic cybersecurity risk assessment: attack evidence and anomaly

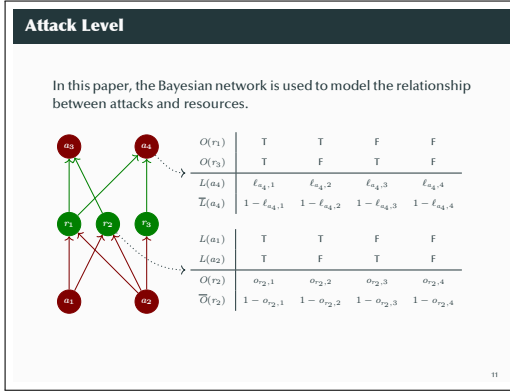
evidence. Attack evidence, which contains information about the type, target, and timestamp of the detected attack, is derived from intrusion detection system. Anomaly evidence, containing the information of the anomaly, such as the invalidation of a function, the occurrence of a hazardous incident, etc., can be obtained from the supervisor system of industrial control systems.

The dynamic cybersecurity risk assessment is divided into two phases: the hazardous incident prediction and the risk assessment. During the hazardous incident prediction phase, attack evidence and anomaly evidence are collected and marked in the multi-level Bayesian network. Then, probabilities of all the potential hazardous incidents can be calculated by analyzing the collected evidences and the multi-level Bayesian network. During the risk assessment phase, the consequences of the hazardous incidents are first classified, and then each type of consequence is quantified in the same unit. Secondly, the overlapping amongst hazardous incidents must be addressed, so the error caused by multiple accumulation of consequences can be eliminated. Finally, the probabilities and consequences of the hazardous incidents are combined into the cybersecurity risk.



Next, I will elaborate the proposed approach of risk assessment for industrial control systems from two parts:

- hazardous incident prediction
- dynamic risk assessment

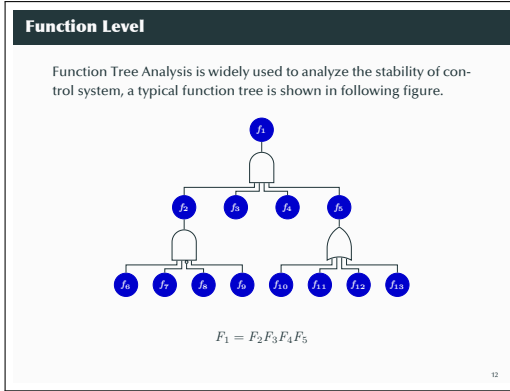


In the proposed approach, the Bayesian network is used to model the relationship between attacks and resources.

The left figure shows a typical Bayesian network of multi-step attack. In this Bayesian network, the attack nodes, which are colored red, represent attack strategies. the resource nodes, which are color green, represent resources. The enforcement of an attack strategy need some conditions. Only the conditions of an attack strategy is satisfied, may this attack strategy be launched. One the other hands, the enforcement of an

attack strategy may obtain another resources. So, using these two kinds of nodes, the Bayesian network can model the multi-step attack.

The Bayesian network uses the conditional probability table to describe the reachable probability. For example, attack node  $a_4$  has two conditions  $r_1$  and  $r_3$ . The first column of the conditional probability table of node  $a_4$  shows that when the attacker obtain the resources  $r_1$  and  $r_3$ , the probability that he launches attack  $a_4$  is  $\ell_{a_4,1}$ . Similarly, if he only has resource  $r_1$ , the probability is  $\ell_{a_4,2}$ .



Function Tree Analysis is widely used to analyze the stability of control system, a typical function tree is shown in following figure.

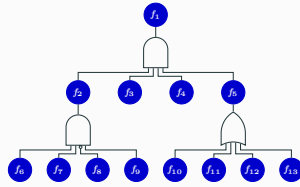
If the relationship amongst the functions lowercase  $f_1, f_2, f_3, f_4$  and  $f_5$  is uppercase  $F_1$  equals  $F_2F_3F_4F_5$ . In this slide, there are two kinds of letter F, where the lowercase  $f$  represents the system function, the uppercase  $F$  represents the status of system function  $f$ . For example, the uppercase  $F_1$  equals True means that the corresponding system function lowercase  $f_1$  is running normally, the uppercase  $F_1$  equals False means that

there is something wrong with the corresponding system function lowercase  $f_1$ .

Let's go back to the relationship amongst the functions lowercase  $f_1, f_2, f_3, f_4$  and  $f_5$ , if the relationship amongst the functions lowercase  $f_1, f_2, f_3, f_4$  and  $f_5$  is uppercase  $F_1$  equals  $F_2F_3F_4F_5$ . The function tree uses an and-gate to describe this relationship.

### Function Level

Function Tree Analysis is widely used to analyze the stability of control system, a typical function tree is shown in following figure.



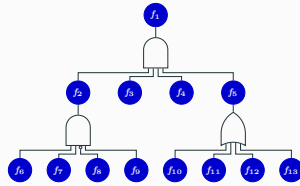
$$F_2 = F_6 F_7 \overline{F_8} F_9$$

12

If the relationship amongst the functions lowercase  $f_2$ ,  $f_6$ ,  $f_7$ ,  $f_8$  and  $f_9$  is uppercase  $F_2$  equals  $F_6 F_7 \overline{F_8} F_9$ . The function tree will uses an appropriate logical gate to describe this kind of relationship.

### Function Level

Function Tree Analysis is widely used to analyze the stability of control system, a typical function tree is shown in following figure.



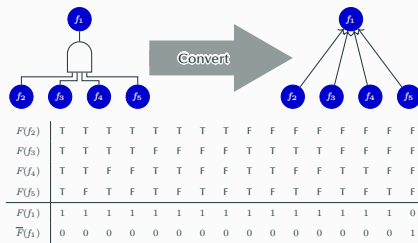
$$F_5 = F_{10} + F_{11} + F_{12} + F_{13}$$

12

Similarly, if the relationship amongst the functions lowercase  $f_5$ ,  $f_{10}$ ,  $f_{11}$ ,  $f_{12}$  and  $f_{13}$  is uppercase  $F_5$  equals  $F_{10} + F_{11} + F_{12} + F_{13}$ . The function tree will uses an or-gate to describe this kind of relationship.

### Function Level

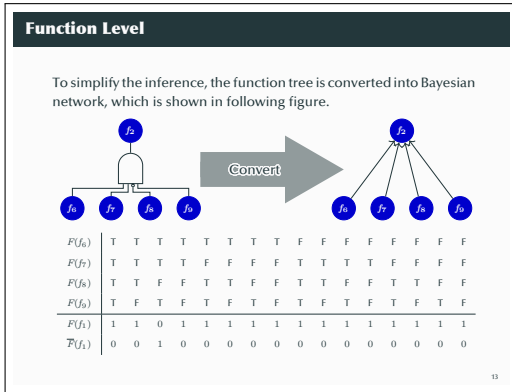
To simplify the inference, the function tree is converted into Bayesian network, which is shown in following figure.



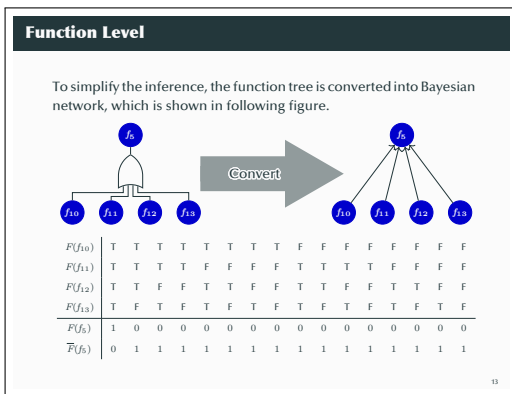
13

To simplify the inference, the function tree is converted into the Bayesian network, which is shown in following figure.

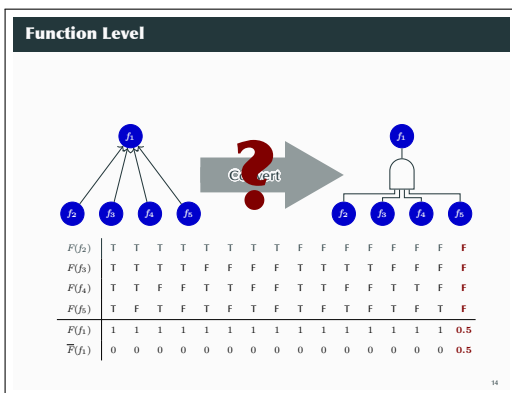
This and gate can be converted to a Bayesian network, in which  $f_2$ ,  $f_3$ ,  $f_4$  and  $f_5$  is the parent nodes of  $f_1$ . Of course, a conditional probability table is needed, too.



This kinds of gate can be also converted into a Bayesian network, but the conditional probability table is different. In fact, all kinds of logical gates can be converted into corresponding Bayesian networks.



For example, the or-gate can be converted into the following Bayesian network.

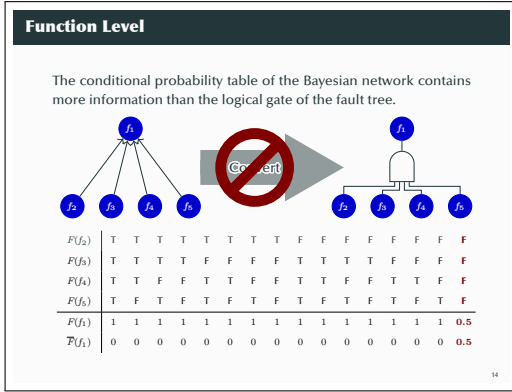


Now, let me digress for a moment. There is a question: can the Bayesian network be converted into the function tree?

The answer is YES, but not all the Bayesian networks can be converted into the corresponding function trees.

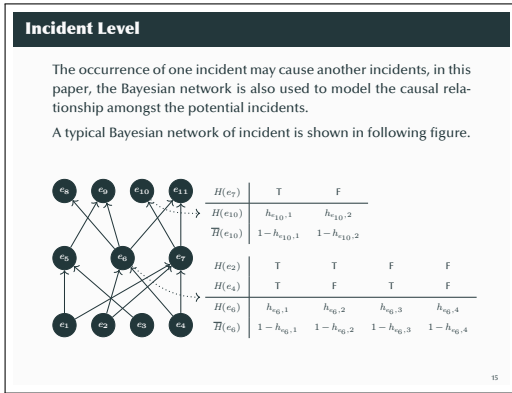
For example, the following conditional probability table can't be converted into a function tree.





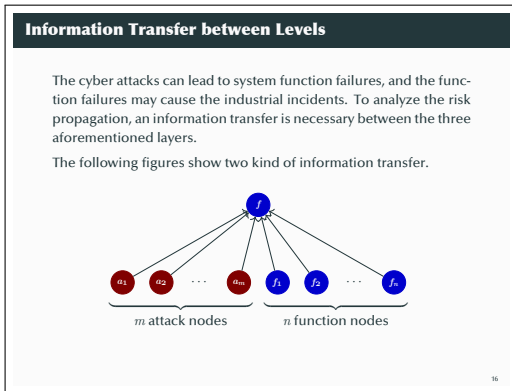
Because the conditional probability table of the Bayesian network contains more information than the logical gate of the fault tree. In other words, the logical gate cannot always accurately describe the relationship amongst functions.

the following conditional probability is an example.



The occurrence of one incident may cause another incidents, in the proposed approach, the Bayesian network is also used to model the causal relationship amongst the potential incidents. A typical Bayesian network of incident is shown in following figure.

Like the attack level, the incident node also needs a conditional probability table to describe the relationship amongst it and its parent nodes.



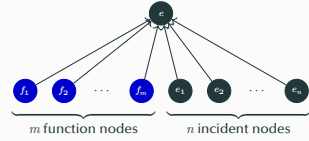
The attack level, the function level and the incident level have been introduced. Now let's talk about the information transfer between levels.

The cyber attacks can lead to system function failures, and the function failures may cause the industrial incidents. To analyze the risk propagation, an information transfer is necessary amongst the three aforementioned layers.

### Information Transfer between Levels

The cyber attacks can lead to system function failures, and the function failures may cause the industrial incidents. To analyze the risk propagation, an information transfer is necessary between the three aforementioned layers.

The following figures show two kind of information transfer.



16

### Collection of Evidence

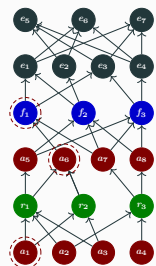
There are two kind of evidence need to be collected:

- **Attack Evidence**, contains the attack information, such as attack time, attack type, attack object, etc.
- **Anomaly Evidence**, contains the information about the anomaly, such as function failure, function restoration, incident occurrence, etc.

For each evidence, there exists a corresponding node in the multi-level Bayesian network. When the intrusion detection system or the monitoring system finds an evidence, the corresponding node will be marked in the multi-level Bayesian network.

17

### Calculation of Incident Probability



The left figure shows a typical multi-level Bayesian network.

Assuming that the evidence list is

$$a_1, a_6, f_1$$

Then the nodes  $a_1$ ,  $a_6$ , and  $f_1$  are marked with **red** dashed circles.

Finally, the algorithm named Probability Propagation in Trees of Clusters (PPTC) can calculate the probabilities of all the hazardous incidents.

18

## Dynamic Risk Assessment

### Decouple of Incident Consequences – Step 1

for each incident  $e_i$ , analyze its consequence and generate a consequence set

$$c_i = (c_{i1}, c_{i2}, \dots, c_{in}).$$

The meaning of  $c_i$  is that the occurring of the incident  $e_i$  will threaten the elements in consequence set  $c_i$ .

For example, the incident  $e_i$  is an explosion of a reactor, which may cause worker casualties, air pollution, facilities damages, and products loss. The consequence set of  $e_i$  is

$$c_i = (\text{workers, air, facilities, products}).$$

20

### Decouple of Incident Consequences – Step 2

21

### Decouple of Incident Consequences – Step 3

For each  $c'_j \in C'$ , generate a corresponding auxiliary node  $x_j$ . According to the **traceability** of  $C'$

$$\forall c' \in C', \exists c \in C, c' \subseteq c,$$

there must be a consequence set  $c_i \in C$ , where  $c'_j \subseteq c_i$ . So, for each  $c'_j \in C'$ , we can find the incident set

$$e_j = (e_{i_1}, e_{i_2}, \dots, e_{i_n}).$$

For each incident  $e_k$  of the incident set  $e_j$ , the corresponding consequence set  $c_k$  satisfies the following condition:

$$c'_j \subseteq c_k.$$

Therefore, the parent nodes of the auxiliary node  $x_j$  are incident nodes  $e_{i_1}, e_{i_2}, \dots, e_{i_n}$ .

22

### Decouple of Incident Consequences – Step 4

For each auxiliary node  $x_j$ , generate a conditional probability table. A typical conditional probability table of auxiliary node  $x_j$  is shown as following table.

$H(e_{i_1})$	T	T	T	...	F	F	F
$H(e_{i_2})$	T	T	T	...	F	F	F
$H(e_{i_3})$	T	T	T	...	F	F	F
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$H(e_{i_{n-2}})$	T	T	T	...	F	F	F
$H(e_{i_{n-1}})$	T	T	F	...	T	F	F
$H(e_{i_n})$	T	F	F	...	F	T	F
$H(x_j)$	1	1	1	...	1	1	0
$\bar{H}(x_j)$	0	0	0	...	0	0	1

23

### Classification of Incident Consequences

In this paper, there are three main kinds of incident consequences to be considered:

- **Harm to Humans:**
  - temporary harm,
  - permanent disability,
  - fatality.
- **Environmental Pollution:**
  - air pollution,
  - soil contamination,
  - water pollution.
- **Property Loss:**
  - damage of materials,
  - damage of products,
  - damage of equipment.

24

### Quantification of Incident Consequences

- **Harm to Humans  $Q_H$ :**

If the decision-maker would like to increase the cost of an investment by  $\Delta c$  to reduce the probability of a fatality by  $\Delta p$ ,

$$Q_H = \Delta c / \Delta p.$$

- **Environmental Pollution  $Q_E$ :**

The monetary loss of environmental pollution is defined as

$$Q_E = \text{Penalty} + \text{Compensation} + \text{HarnessCost}.$$

- **Property Loss  $Q_P$ :**

The cost of replacement is used to quantify the loss of property  $Q_P$ , such as the loss of materials, products, and equipment.

25

### Calculation of Dynamic Risk

Due to the following two reasons:

- there is no overlapping between the consequences of any two auxiliary nodes  $x_i$  and  $x_j$ ,  $i \neq j$ ,
- the auxiliary nodes contain all the consequences of incidents,

the dynamic cybersecurity risk can be defined as

$$\mathcal{R} = \sum_{i=1}^{m'} p(x_i) q(x_i),$$

where

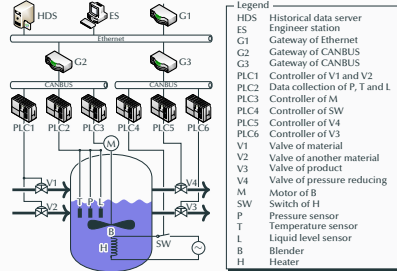
- $p(x_i)$  is the occurrence probability of the auxiliary node  $x_i$ ,
- $q(x_i)$  is the monetary loss of the auxiliary node  $x_i$ .

26

## Simulation

## Simulation Platform

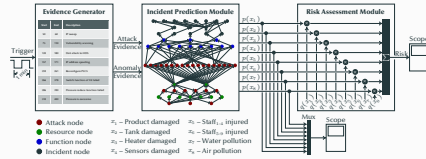
The simulation object is a chemical reactor whose control structure is shown as the following figure.



28

## Simulation Platform

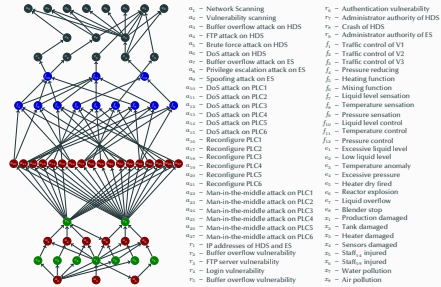
The simulation platform is implemented in Matlab, which consists of three modules: an evidence generator, an incident prediction module, and a risk assessment module.



29

## Simulation Platform

The multi-level Bayesian network of the chemical reactor is shown as following figure.



30

### Simulation Platform

The list of evidences is shown as following table.

Start	End	Description	Symbol
50	60	IP sweep	$L(a_1)$
75	110	Vulnerability scanning	$L(a_2)$
120	180	DoS attack to HDS	$L(a_6)$
157	171	IP address spoofing	$L(a_9)$
259	261	Reconfigure PLC5	$L(a_{20})$
266	378	Switch function of V4 failed	$F(f_4)$
286	390	Pressure reduce function failed	$F(f_{12})$
310	400	Pressure is excessive	$H(e_4)$

31

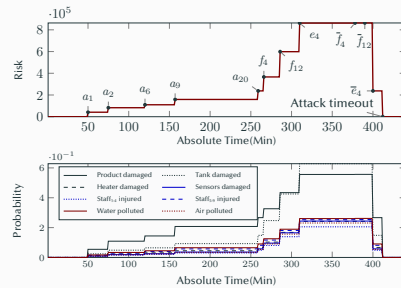
### Simulation Platform

The quantification of consequences is shown as following table.

Incident Symbol	Description of Incident	Quantification of Consequence(\$)
$x_1$	Product damaged	50,000
$x_2$	Tank damaged	500,000
$x_3$	Heater damaged	10,000
$x_4$	Sensors damaged	10,000
$x_5$	Staff <sub>1-4</sub> injured	800,000
$x_6$	Staff <sub>5-9</sub> injured	1,000,000
$x_7$	Water pollution	200,000
$x_8$	Air pollution	200,000

32

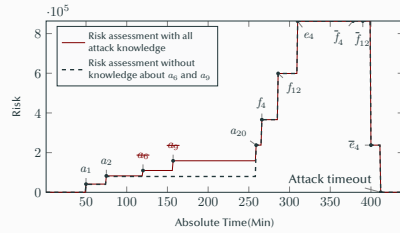
### Simulation and Result Analysis



33

### Simulation and Result Analysis

Then an identical multi-step attack on the system is launched to the system. The new cybersecurity risk curve is shown the dashed line in the following figure.



34

### Simulation and Result Analysis

Some parameters of the following figure:

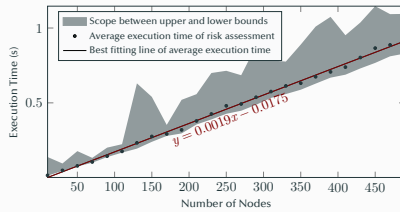
- The average execution time of a risk assessment is 94.1ms.
- The minimum execution time of a risk assessment is 89.9ms.
- The maximum execution time of a risk assessment is 131.6ms.



35

### Simulation and Result Analysis

This means that the execution time of the risk assessment scales linearly with the increase of the node size of the multi-level Bayesian network.



36



## Conclusion and Prospect

### Conclusion

- By considering the characteristics of ICSs, a novel multi-level Bayesian network was proposed, which integrated a knowledge of attack, system function, and hazardous incident.
- The attack knowledge and system knowledge were combined to analyze the potential impact of attacks, so the proposed approach had the ability of assessing the risk caused by unknown attacks.
- A unified quantification approach for a variety of consequences of industrial accidents was introduced. Furthermore, the proposed approach could eliminate the error of risk caused by the overlapping amongst hazardous incidents.
- By using a simplified chemical reactor control system in Matlab environment, the designed dynamic risk assessment approach was verified.

38

### Prospect

There are some shortcomings of the proposed risk assessment approach need to be improved.

- **Current research work has no ability for self-learning.**
- **The sub-second computation time cannot meet some hard real-time systems requirements.**

In the future, a dynamic cybersecurity risk assessment, which can automatically adjust the conditional probability and structure of the multi-level Bayesian network by analyzing the real-time data, will be researched, and several approximate inference methods will be attempted in the risk assessment.

39

**Thank You!**

**Thank You!**

You can obtain this slide from my Github:  
[zqmillet@github.com:Presentation.for.Loughborough.University](https://github.com/zqmillet/Presentation.for.Loughborough.University)

And I have pushed the code of the simulation to my Github, too.  
[zqmillet@github.com:Multi-level.Bayesian.Network](https://github.com/zqmillet/Multi-level.Bayesian.Network)

41

**Any Questions?**