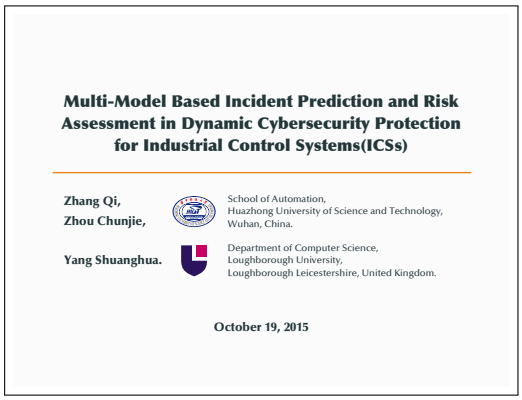
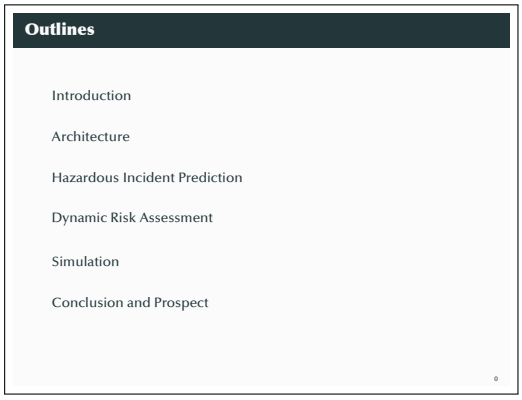



Hello everyone, my name is Zhang Qi, and I am the Ph.D student of Professor Zhou Chunjie. I am very glad to be invited by Professor Yang Shuanghuang to make a presentation about my recent research.

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

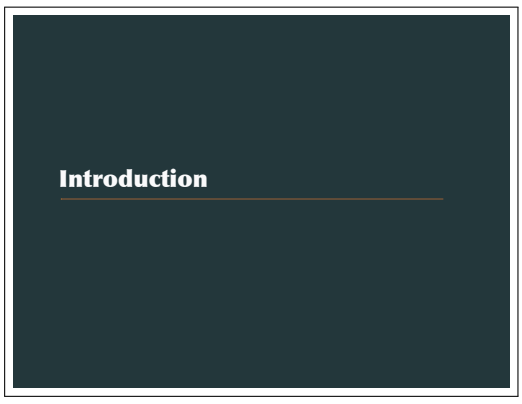
The purpose of my presentation is to introduce my research, and I hope I can get your good advices.

I shall take about 20 minutes of your time. If you have any questions, please feel free to interrupt.

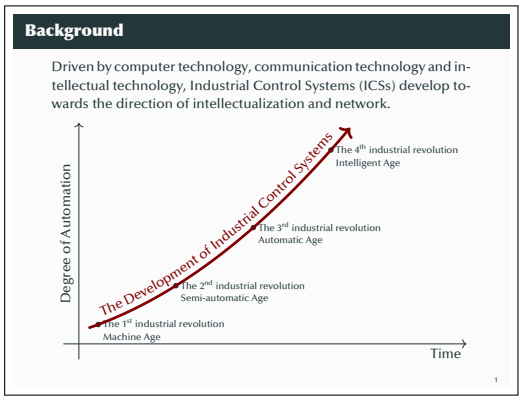



My presentation is separated into six parts:

- 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.



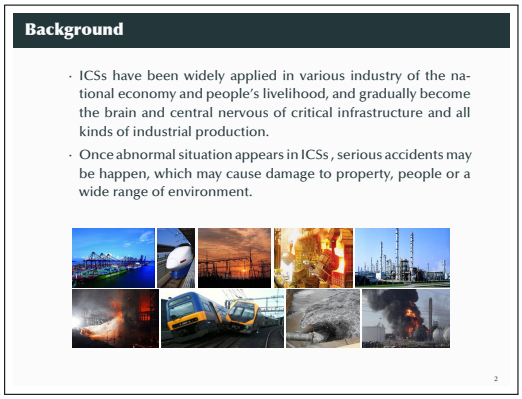

In this part, I will introduce the development history and the cybersecurity issues of industrial control systems. And, I will compare the cybersecurity issues of industrial control systems and traditional IT systems.




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

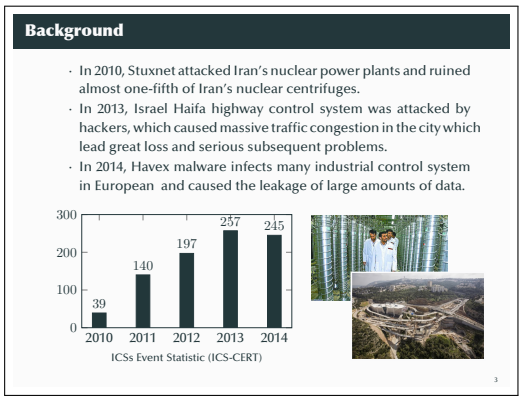
- Machine Age
- Semi-automatic Age
- Automatic Age
- Intelligent Age

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.

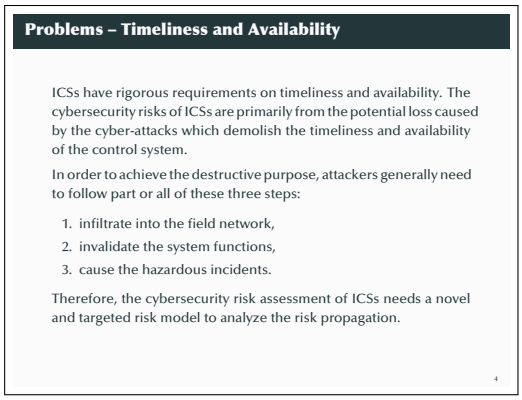
As mentioned before, the industrial control systems are evolving towards intelligence and networking. The rapid development of the industrial control systems reduce the difficulty of the development and the cost of construction, on the other hand, it has also introduced the cybersecurity issues into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

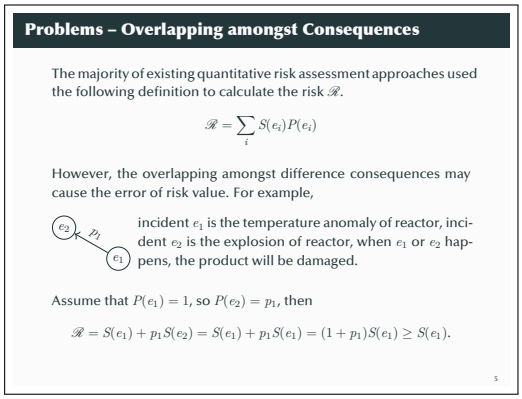
According to the statistical data from "Year in Review 2014" published by the ICS-CERT which is short for "Industrial Control Systems Cyber Emergency Response Team", the number of attacks for industrial control systems increases year by year. In 2010, there were only 39 security incidents of industrial 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.




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.




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 $e_i$.

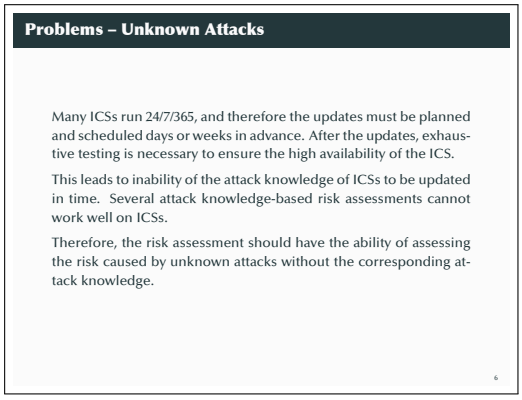
It is also worth noting that there is a problem when this definition is used in industrial control systems risk assessment. This is due to the fact that, for industrial control systems, different hazardous incidents may cause the same consequence; whereby, using this definition to assess risk will cause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

$$\mathscr{R} = S(e_1) + p_1 S(e_2) = S(e_1) + p_1 S(e_1) = (1 + p_1)S(e_1) \geq S(e_1).^{\text{I}}$$

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

I. the risk is equal to capital S e sub 1 plus p sub 1 multiplied by capital S e sub 2 is equal to capital S e sub 1 plus p sub 1 multiplied by capital S e sub 1 is equal to 1 plus p sub 1 multiplied by capital S e sub 1 is greater than or equal to capital S e sub 1

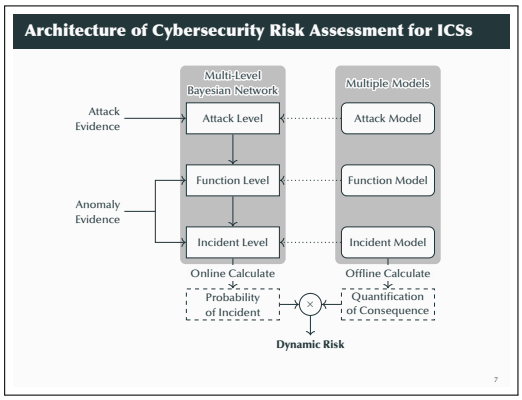



As continuous operation systems, the industrial control systems cannot tolerate frequent software patching or updates. This causes the database of attack signatures to lag far behind the rapid development of attacks. With this defect, several intrusion detection system based misuse detections would miss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


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.

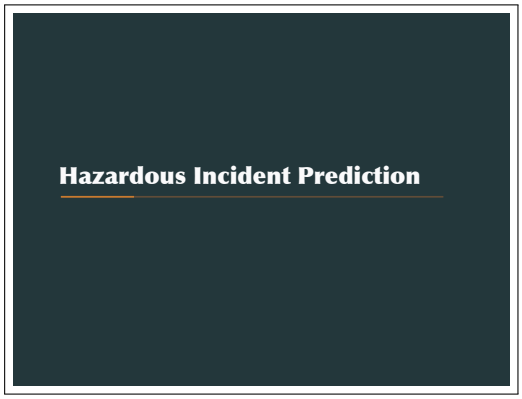



To meet the requirement of the risk assessment for industrial control systems, a dynamic cybersecurity risk assessment based on the multi-model is proposed.

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.

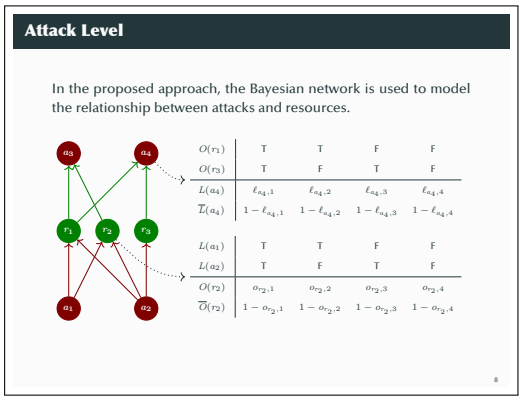
There are two kinds of inputs for the dynamic cybersecurity risk assessment: attack evidence and anomaly evidence. Attack evidence, which contains information about the type, target, and timestamp of the detected attack, is derived from intrusion detection system. Anomaly evidence, containing the information 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

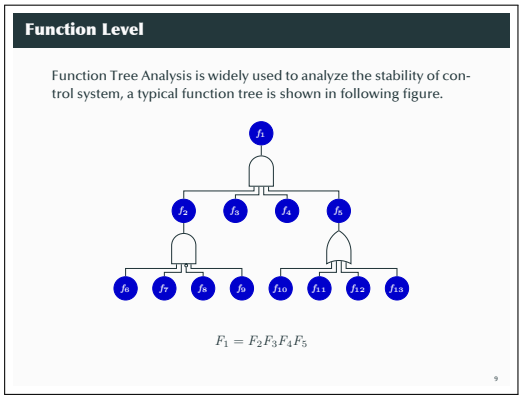



In the proposed approach, the Bayesian network is used to model the relationship between attacks and resources.

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 $\ell_{a_4,1}$. [I] Similarly, if he only has resource $r_1$, the probability is $\ell_{a_4,2}$. [II]

I. the probability that he launches attack a sub 4 is l sub a sub 4 1
II. Similarly, if he only has resource r sub 1, the probability is l sub a sub 4 2

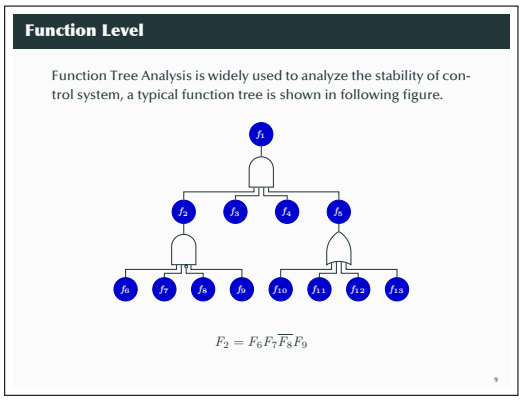



Function Tree Analysis is widely used to analyze the stability of control system, a typical function tree is shown in following figure.

If the relationship amongst the functions $f_1$, $f_2$, $f_3$, $f_4$ and $f_5$ is $F_1 = F_2F_3F_4F_5$. [I] In this slide, there are two kinds of letter `F`, where the lowercase `f` represents the system function, the capital `F` represents the status of system function $f$. For example, the $F_1$ is equal to `True` means that the corresponding system function $f_1$ is running normally, the $F_1$ is equal to `False` means that there is something wrong with the corresponding system function $f_1$.

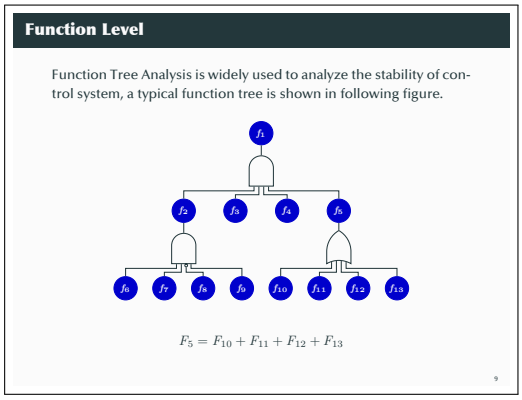
Let's go back to the relationship amongst the functions $f_1$, $f_2$, $f_3$, $f_4$ and $f_5$, if the relationship amongst the functions $f_1$, $f_2$, $f_3$, $f_4$ and $f_5$ is $F_1 = F_2F_3F_4F_5$. The function tree uses an and-gate to describe this relationship.

I. If the relationship amongst the functions lowercase f sub 1, f sub 2, f sub 3, f sub 4 and f sub 5 is capital F sub 1 is equal to capital F sub 2 F sub 3 F sub 4 F sub 5



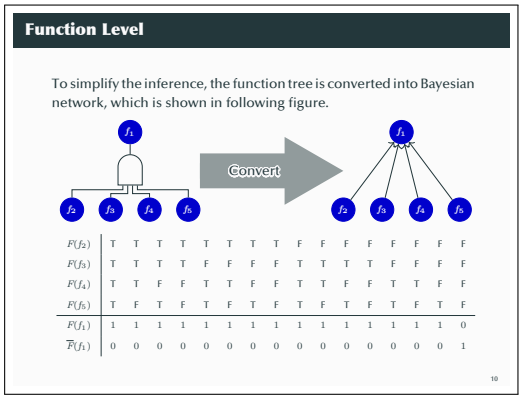

If the relationship amongst the functions $f_2$, $f_6$, $f_7$, $f_8$ and $f_9$ is $F_2 = F_6F_7\overline{F_8}F_9$. [I] The function tree will uses an appropriate logical gate to describe this kind of relationship.

I. If the relationship amongst the functions lowercase f sub 2, f sub 6, f sub 7, f sub 8 and f sub 9 is capital F sub 2 is equal to capital F sub 6 F sub 7 F sub 8 bar F sub 9



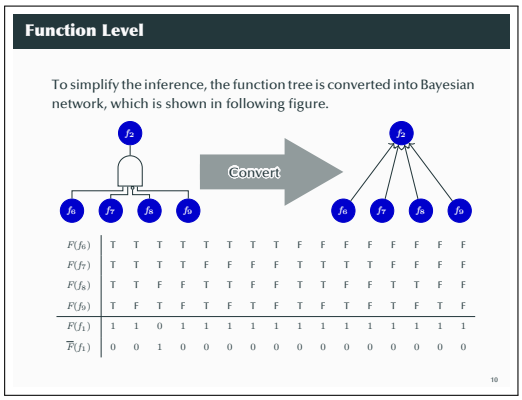

Similarly, if the relationship amongst the functions $f_5$, $f_{10}$, $f_{11}$, $f_{12}$ and $f_{13}$ is $F_5 = F_{10} + F_{11} + F_{12} + F_{13}$. [I] The function tree will uses an or-gate to describe this kind of relationship.

I. Similarly, if the relationship amongst the functions lowercase f sub 5, f sub 10, f sub 11, f sub 12 and f sub 13 is capital F sub 5 is equal to capital F sub 10 plus F sub 11 plus F sub 12 plus F sub 13

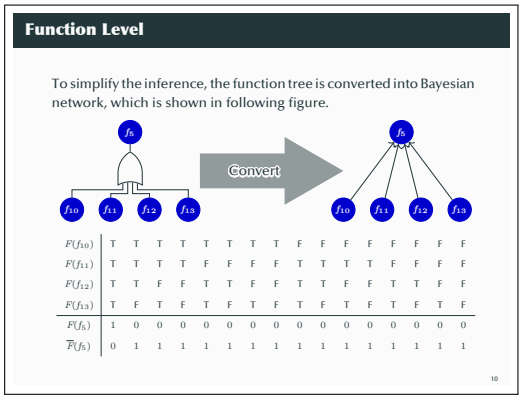



To simplify the inference, the function tree is converted into the Bayesian network, which is shown in following figure.

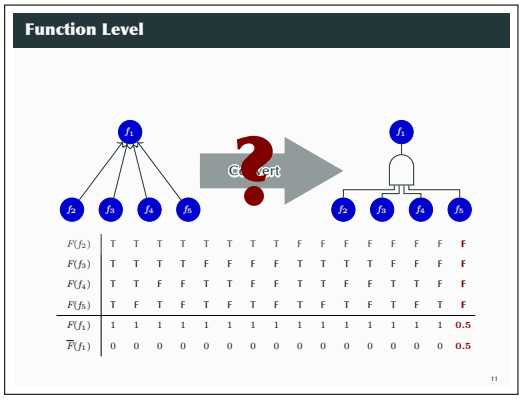
This and gate can be converted to a Bayesian network, in which $f_2$, $f_3$, $f_4$ and $f_5$ is the parent nodes of $f_1$. Of cause, a conditional probability table is needed, too.




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.



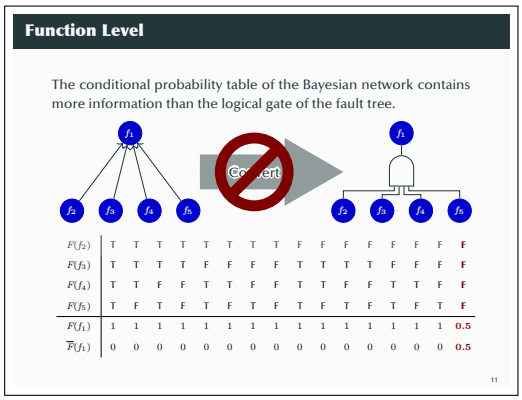

For example, the or-gate can be converted into the following Bayesian network.




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

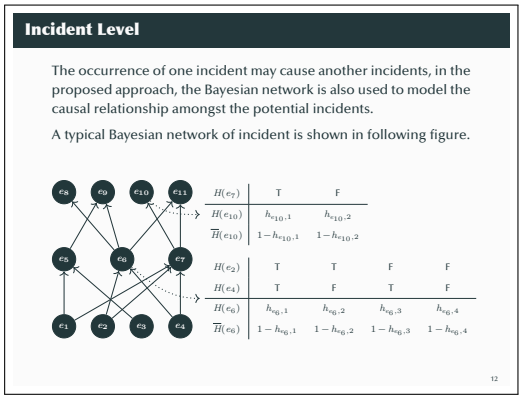
The answer is YES, but not all the Bayesian networks can be converted into the corresponding function trees.

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



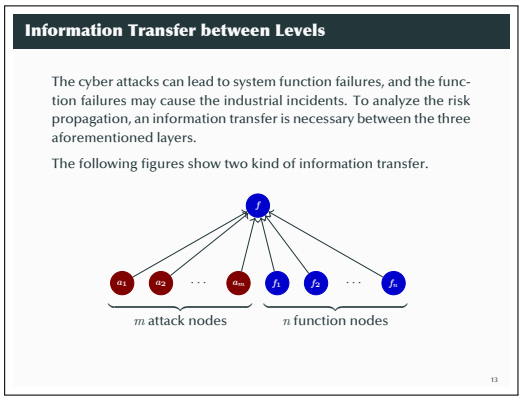

Because the conditional probability table of the Bayesian network contains more information than the logical gate of the fault tree. In other words, the logical gate cannot always accurately describe the relationship amongst functions.

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.

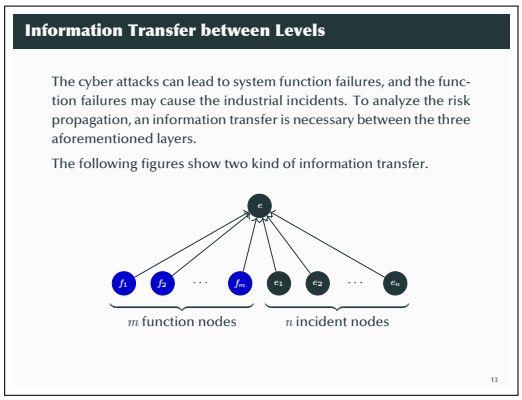



The attack level, the function level and the incident level have been introduced. Now let's talk about the information transfer between levels.

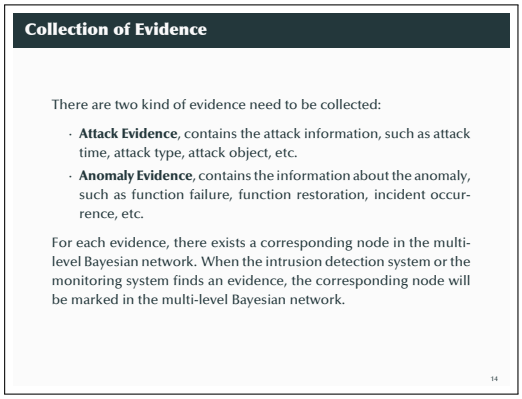
The cyber attacks can lead to system function failures, and the function failures may cause the industrial incidents. To analyze the risk propagation, an information transfer is necessary amongst the three aforementioned layers.

The following figures show the information transfer between attack level and function level.

We can see that the function failure can lead to another function failures, the launch of attack can also cause function failures. Therefore, we need add some parent nodes for function node $f$. Additional, the conditional probability table of function node $f$ need to be extended.



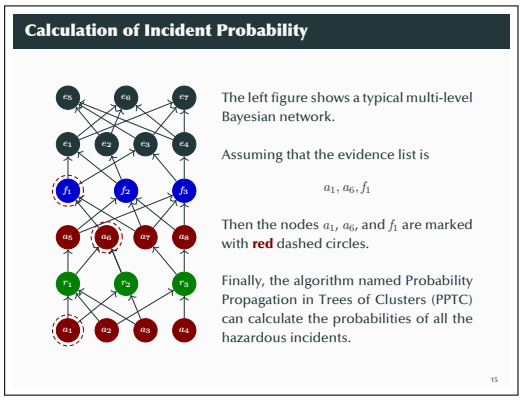

Similarly, the information transfer between function level and incident level is shown in the following figure.




There are two kind of evidence need to be collected attack evidence and anomaly evidence.

- The attack evidence contains the attack information, such as attack time, attack type, attack object, and so on. The attack evidence can be obtain by the intrusion detection systems.
- The anomaly evidence, contains the information about the anomaly, such as function failure, function restoration, incident occurrence, and so on. The anomaly evidence can be obtained by the monitoring systems.

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.

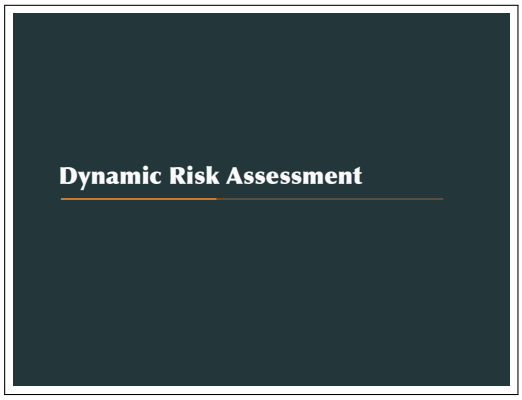



The right figure shows a typical multi-level Bayesian network. We can see it has three levels: attack level colored red and green, function level colored blue, and incident level colored black.

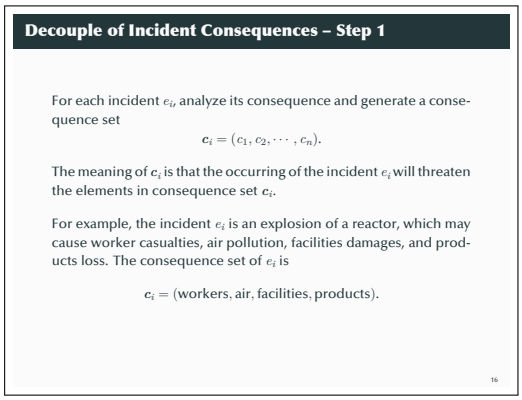
Assuming that there are three evidences which are detected. They are $a_1$, $a_6$ and $f_1$. Then the nodes $a_1$, $a_6$ and $f_1$ are marked with red dashed circles.

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

Now, all the potential hazardous incidences can be calculated. So in summary, first, we collect the evidences about the systems. The evidences contains attack evidences and anomaly evidences. Then we infer what gonna happen according the evidences and their relationship. We use the Bayesian network to model the relationship amongst the evidences.




Now we have obtained the probability of all potential hazardous incidents. Let us talk about another problem, how to calculate the dynamic cybersecurity risk?




As mentioned before, there may exist overlapping amongst different incident consequences. And the overlapping will cause the error of the cybersecurity risk. So the first thing to do is the decouple of consequences.

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

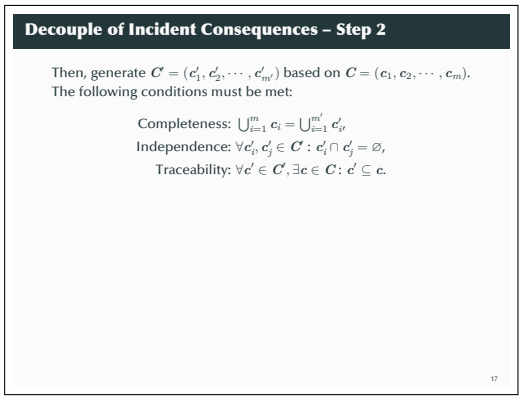
$$\boldsymbol{c}_i = (c_1, c_2, \cdots, c_n).^{\text{I}}$$

The meaning of $\boldsymbol{c}_i$ is that the occurring of the incident $e_i$ will threaten the elements in consequence set $\boldsymbol{c}_i$.

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

$$\boldsymbol{c}_i = (\text{workers}, \text{air}, \text{facilities}, \text{products}).^{\text{II}}$$

I. the set c sub i is equal to the set of c sub 1, c sub 2, dots, c sub n
II. c sub i is equal to the set of workers, air, facilities and products




Step 2, generate $\boldsymbol{C}' = (\boldsymbol{c}'_1, \boldsymbol{c}'_2, \cdots, \boldsymbol{c}'_{m'})$ based on $\boldsymbol{C} = (\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_{m'})$. [I]

It is noted that all letter $\mathtt{C}$ is in bold type, it means that they are all sets. the lowercase $\mathtt{c}$ is the set of consequences, the capital $\mathtt{C}$ is the set of lowercase $\mathtt{c}$.

When we generate the set $\boldsymbol{C}'$, there are three conditions that set $\boldsymbol{C}'$ must be met:

First, the completeness:

$$\bigcup_{i=1}^{m} \boldsymbol{c}_i = \bigcup_{i=1}^{m'} \boldsymbol{c}'_i.^{\text{II}}$$

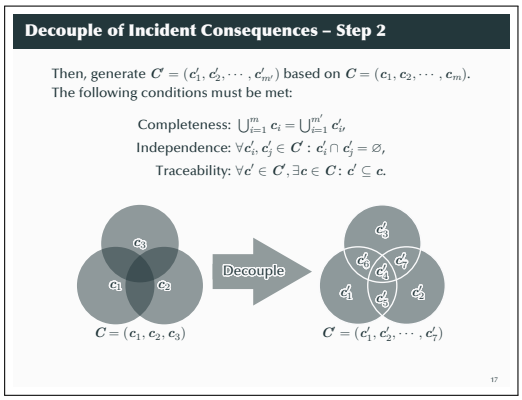
Second, the independence:

$$\forall \boldsymbol{c}'_i, \boldsymbol{c}'_j \in \boldsymbol{C}' : \boldsymbol{c}'_i \cap \boldsymbol{c}'_j = \varnothing.^{\text{III}}$$

Third, the traceability:

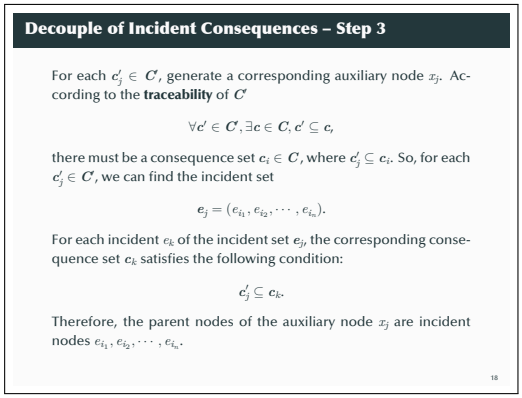
$$\forall \boldsymbol{c}' \in \boldsymbol{C}', \exists \boldsymbol{c} \in \boldsymbol{C} : \boldsymbol{c}' \subseteq \boldsymbol{c}.^{\text{IV}}$$

I. Step 2, generate set C prime which is equal to set c prime sub 1, set c prime sub 2, dots, set c prime sub m prime, based on set C is equal to set c sub 1, set c sub 2, dots, set c sub m
II. First, the completeness, the union of c sub 1 to c sub m is equal to c prime sub 1 to c prime sub m prime
III. Second, the independence, for each c prime i and c prime j in capital capital C prime, c prime i intersection c prime j is equal to empty set
IV. Third, the traceability, for each set c prime in capital C prime, there must be a set c in capital C, which c prime is a subset of or equal to c




For example, there are three hazardous incidents $e_1$, $e_2$, and $e_3$, and their consequences are $c_1$, $c_2$, and $c_3$. From the figure we can see that there are overlapping amongst these three consequences. The decoupled consequences are shown in the right figure.

Obviously, the set $\boldsymbol{C'}$ satisfies these three conditions.




For each $\boldsymbol{c'_j} \in \boldsymbol{C'}$, we generate a corresponding auxiliary node $x_j$. [I]

According to the traceability of $\boldsymbol{C'}$, there must be a consequence set $\boldsymbol{c_i} \in \boldsymbol{C}$, where $\boldsymbol{c'_j} \subseteq \boldsymbol{c_i}$. [II]

So, for each $\boldsymbol{c'_j} \in \boldsymbol{C'}$, we can find the incident set

$$\boldsymbol{e_j} = (e_{i_1}, e_{i_2}, \cdots, e_{i_n}).^{\text{III}}$$

For each incident $e_k$ of the incident set $\boldsymbol{e_j}$, the corresponding consequence set $\boldsymbol{c_k}$ satisfies the following condition: $\boldsymbol{c'_j} \subseteq \boldsymbol{c_k}$. [IV]

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

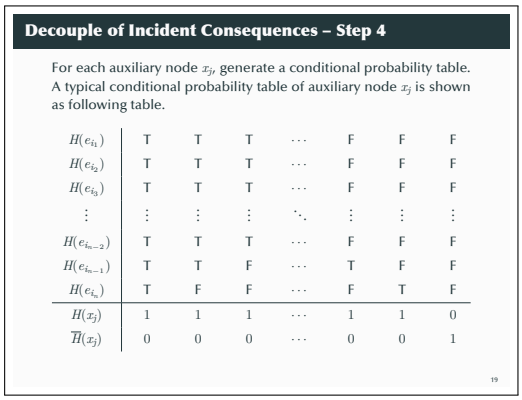
I. For each c prime sub j in capital C prime, we can generate a corresponding auxiliary node x sub j

II. According to the traceability of capital C prime, there must be a consequences set c sub i in capital C, where c prime sub j is a subset of or equal to c sub i.

III. So, for each c prime sub j in capital prime, we can fine the incident set e sub j is equal to e sub i sub 1, e sub i sub 2, dots, e sub i sub n

IV. For each incident e sub k of the incident set e sub j , the corresponding consequence set c sub k satisfies the following condition: c prime j is a subset of or equal to c sub k, where c prime sub j is an element of capital C prime

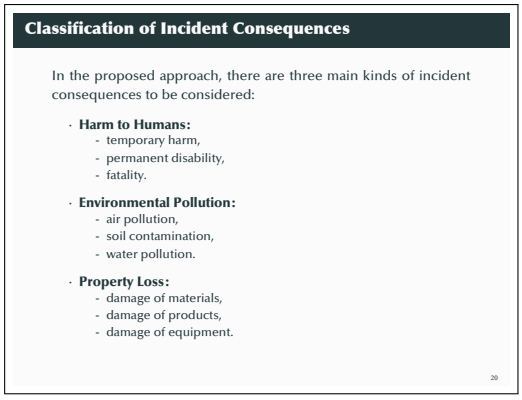
V. Therefore, the parent nodes of the auxiliary node x sub j are incident nodes e sub i sub 1, e sub i sub 2, dots, e sub i sub n




The last step, For each auxiliary node x sub j , generate a conditional probability table. A typical conditional probability table of auxiliary node x sub j is shown as following table.

$H(x_j)$ is equal to 0 only when all its parent nodes are `False`. In other words, if an incident happens, the auxiliary node will be damaged.

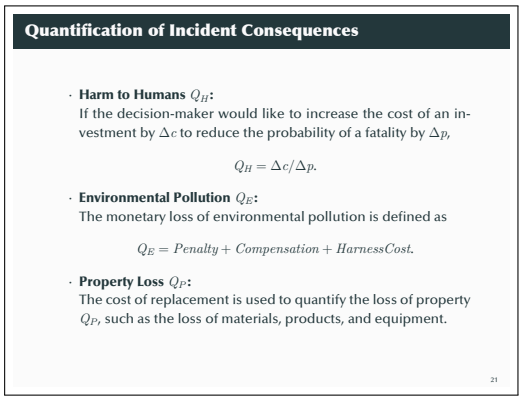
Now, the consequences has been decoupled. The completeness of set $C'$ ensures that there is no omission, and the independence of set $C'$ ensures that there is no overlapping.




Next, we will talk about the quantification of incident consequences.

Before the quantification of incident consequences, the adverse effects of an incident may be classified, because different adverse effects need different quantification methods.

For industrial control systems, the adverse effects of an incident may be classified into three categories: harm to humans, environmental pollution, and property loss.




Different loss of different incident consequences need to be accumulated into the cybersecurity risk, so the loss of incident need to be quantified into same unit. In the proposed approach, we quantify the different loss into monetary loss.

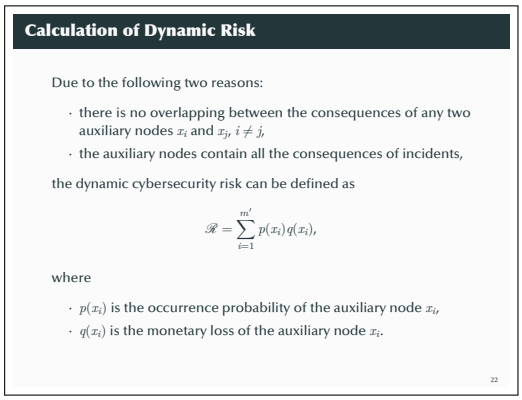
For the harm to humans, if the decision-maker would like to increase the cost of an investment by $\Delta c$ to reduce the probability of a fatality by $\Delta p$, in proposed approach, the quantification of harm to humans is defined as $Q_H = \Delta c/\Delta p$. [I]

The monetary loss of environmental pollution is defined as the sum of penalty, compensation, and harness cost.

- According to the environmental protection laws, if the occurrence of an incident causes environmental pollution, as the owner of industrial control system, the company must pay the penalty charge
- When environmental pollution occurs, it tends to influence the living conditions of residents near the plant, the downstream agricultural production, etc. As the relevant liable person, the company has the obligation to pay for compensation.
- To clear the polluted environment, as the polluter, the company must take action to improve the environment.

The monetary loss of property is defined as the cost of replacement.

I. For the harm to humans, 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, in proposed approach, the quantification of harm to humans is defined as Q sub H is equal Delta c over Delta p




Due to the following two reasons:

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

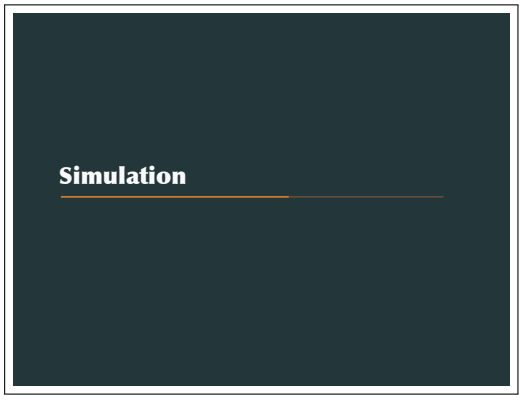
the dynamic cybersecurity risk can be defined as the following formula:

$$\mathscr{R} = \sum_{i=1}^{m'} p(x_i)q(x_i), \text{[II]}$$

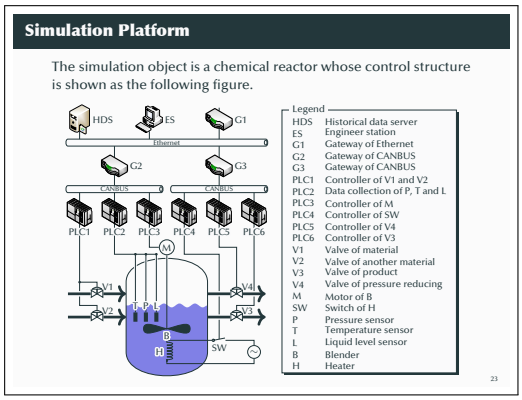
where

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

I. i is not equal to j
II. the risk is equal to the sum from i equals one to m prime p x sub i multiplied by q x sub i
III. p x sub i is the occurrence probability of the auxiliary node x sub i
IV. q x sub i is the monetary loss of the auxiliary node x sub i

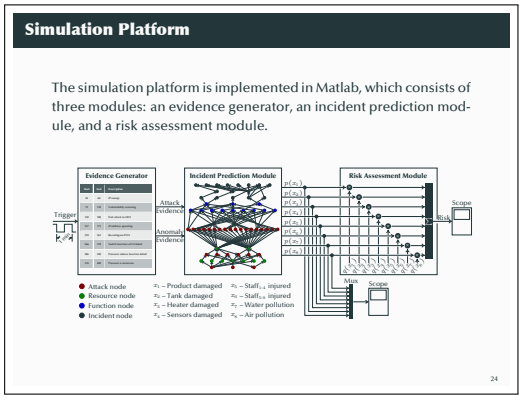



To illustrate how our approach validly calculates the cybersecurity risk in real-time, a numerical simulation is carried out.




A chemical reactor is a device for containing and controlling a chemical reaction and is widely used in the chemical industry. The representative structure of a chemical reactor control system is shown as the figure.

In this figure, the Ethernet connects to the enterprise network via gateway 1, which is not shown in this figure. Two CANBUS networks connect to the Ethernet via gateway 2 and gateway 3. In the Ethernet, there are an engineer station and a historical data server. The host in the enterprise network can access the historical data of the historical data server, but cannot access the engineer station. PLC1-6 are distributed into two CANBUS networks. The engineer station and the historical data server can obtain data from all of the PLCs, but only the engineer station can modify and configure PLCs.



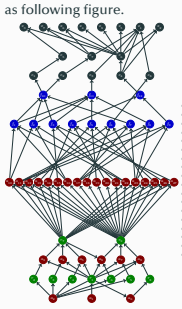

This figure shows the structure of the simulation platform.

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

The evidence generator is used to simulate the intrusion detection system and monitoring system. It uses an array to store an evidence list, which will be shown later. It has a time trigger. To the scheduled time, it will generate an evidence and send it to the incident prediction module.

**Simulation Platform**

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



$a_1$ – Network Scanning
$a_2$ – Vulnerability scanning
$a_3$ – Buffer overflow attack on HDS
$a_4$ – FTP attack on HDS
$a_5$ – Brute force attack on HDS
$a_6$ – DoS attack on HDS
$a_7$ – Buffer overflow attack on ES
$a_8$ – Privilege escalation attack on ES
$a_9$ – Spoofing attack on ES
$a_{10}$ – DoS attack on PLC1
$a_{11}$ – DoS attack on PLC2
$a_{12}$ – DoS attack on PLC3
$a_{13}$ – DoS attack on PLC4
$a_{14}$ – DoS attack on PLC5
$a_{15}$ – DoS attack on PLC6
$a_{16}$ – Reconfigure PLC1
$a_{17}$ – Reconfigure PLC2
$a_{18}$ – Reconfigure PLC3
$a_{19}$ – Reconfigure PLC4
$a_{20}$ – Reconfigure PLC5
$a_{21}$ – Reconfigure PLC6
$a_{22}$ – Man-in-the-middle attack on PLC1
$a_{23}$ – Man-in-the-middle attack on PLC2
$a_{24}$ – Man-in-the-middle attack on PLC3
$a_{25}$ – Man-in-the-middle attack on PLC4
$a_{26}$ – Man-in-the-middle attack on PLC5
$a_{27}$ – Man-in-the-middle attack on PLC6
$\tau_1$ – IP addresses of HDS and ES
$\tau_2$ – Buffer overflow vulnerability
$\tau_3$ – FTP server vulnerability
$\tau_4$ – Login vulnerability
$\tau_5$ – Buffer overflow vulnerability
$\tau_6$ – Authentication vulnerability
$\tau_7$ – Administrator authority of HDS
$\tau_8$ – Crash of HDS
$\tau_9$ – Administrator authority of ES
$f_1$ – Traffic control of V1
$f_2$ – Traffic control of V2
$f_3$ – Traffic control of V3
$f_4$ – Pressure reducing
$f_5$ – Heating function
$f_6$ – Mixing function
$f_7$ – Liquid level sensation
$f_8$ – Temperature sensation
$f_9$ – Pressure sensation
$f_{10}$ – Liquid level control
$f_{11}$ – Temperature control
$f_{12}$ – Pressure control
$e_1$ – Excessive liquid level
$e_2$ – Low liquid level
$e_3$ – Temperature anomaly
$e_4$ – Excessive pressure
$e_5$ – Heater dry fired
$e_6$ – Reactor explosion
$e_7$ – Liquid overflow
$e_8$ – Blender stop
$x_1$ – Production damaged
$x_2$ – Tank damaged
$x_3$ – Heater damaged
$x_4$ – Sensors damaged
$x_5$ – Staff$_{1-4}$ injured
$x_6$ – Staff$_{5-9}$ injured
$x_7$ – Water pollution
$x_8$ – Air pollution

25

This figure shows the multi-Level Bayesian network of reactor.

**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 | $H(e_4)$ |

26

This table shows the evidence list of the evidence generator.

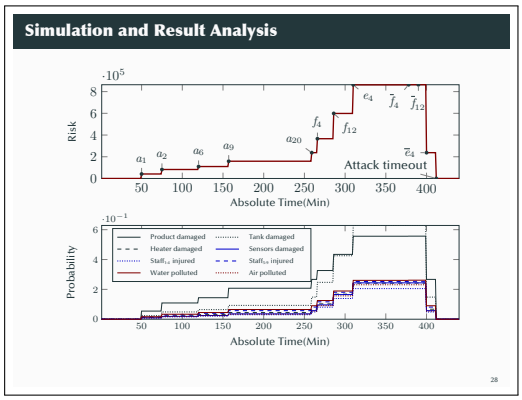
**Simulation Platform**

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$_{1-4}$ injured | 800,000 |
| $x_6$ | Staff$_{5-9}$ injured | 1,000,000 |
| $x_7$ | Water pollution | 200,000 |
| $x_8$ | Air pollution | 200,000 |

27

This table shows the quantifications of different consequences.

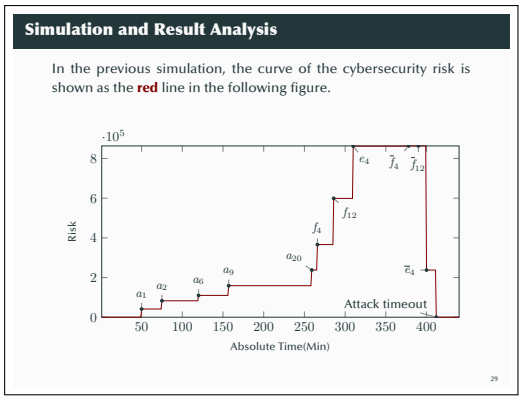



The first simulation.

A multi-step attack is launched on the chemical reactor control system. The evidences are collected and the cybersecurity risk is calculated every minute. Then the curves of the cybersecurity risk and probabilities of incidents $x_1, x_2, \cdots, x_8$ in the multi-level Bayesian network are shown in these two figure.

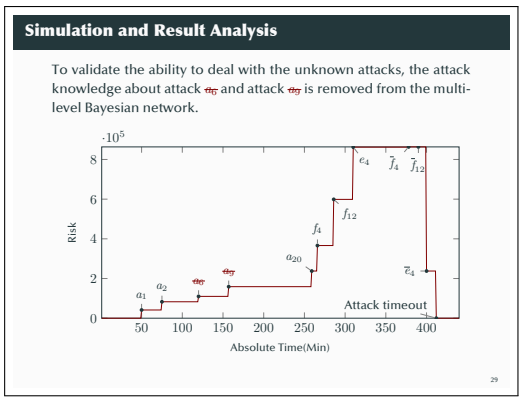
From these two figures, we can see:

- with the infiltration of multi-step attacks, the cybersecurity risk of the chemical reactor control system is increasing,
- when attack is aborted, the cybersecurity risk of the chemical reactor control system is decreasing,
- the curves of probabilities of incidents is in line with the trend of the curve of cybersecurity risk.

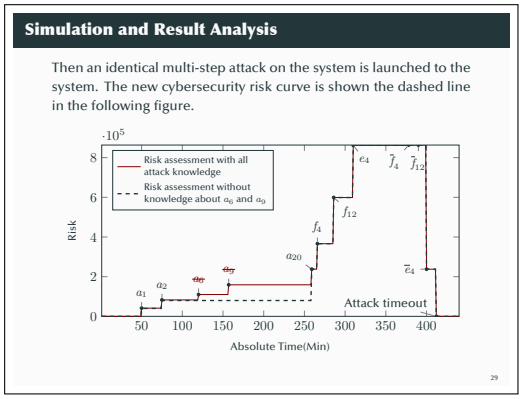



The second simulation.

In the previous simulation, the curve of the cybersecurity risk is shown as the red line in the following figure.

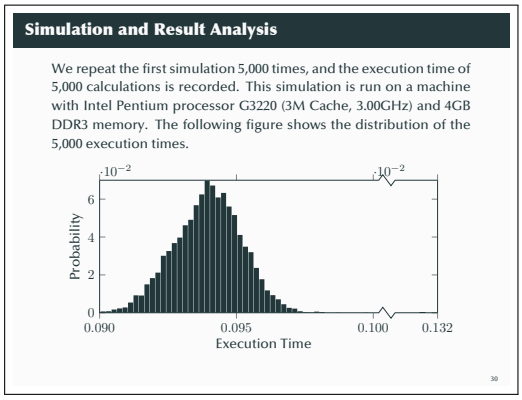



To illustrate the ability to deal with unknown attacks, the attack nodes $a_6$ and $a_9$ are removed from the multi-level Bayesian network. Thus, the incident prediction module does not know that an attacker can get the administrator authority of the engineer station through a DoS attack and an IP spoofing attack. In other words, $a_6$ and $a_9$ are unknown attacks to the incident prediction module. Additionally, the conditional probability table of the resource node $r_9$ also needs to be modified.



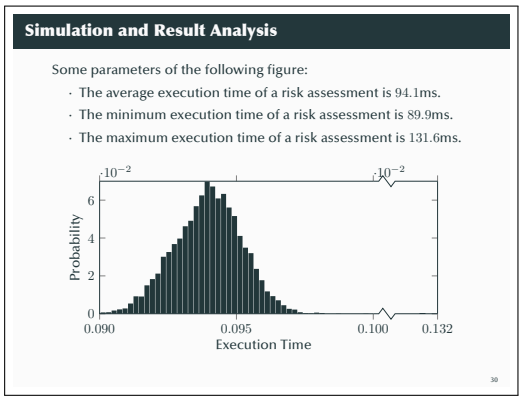

The same multi-step attack is launched to the chemical reactor control system again. Since there is no knowledge of attacks $a_6$ and $a_9$, the evidences of $a_6$ and $a_9$ must be removed from the evidence list. The cybersecurity risk value is recorded every minute, the new cybersecurity risk curve is shown the dashed line in the following figure.

This figure shows that, before the 120th minute, the risk of the second simulation is slightly below that of the first simulation. The reason is that, without the knowledge of $a_6$ and $a_9$, the probability of an attack obtaining the resource $r_9$ is lower in view of the incident prediction module. After the 120th minute and before the 259th minute, there is a difference between two of the risk curves. Since there is no evidences of $a_6$ and $a_9$, the risk of the second simulation in this range remains unchanged. After the 259th minute, we can see the risk curves of two simulations that have overlapped. This comparison shows that, without the knowledge of several atom attacks, there is no comparatively large deviation in the result of the risk assessment. Therefore, if there are a few unknown atom attacks in a multi-step attack, our approach can still generate a relatively accurate risk value.



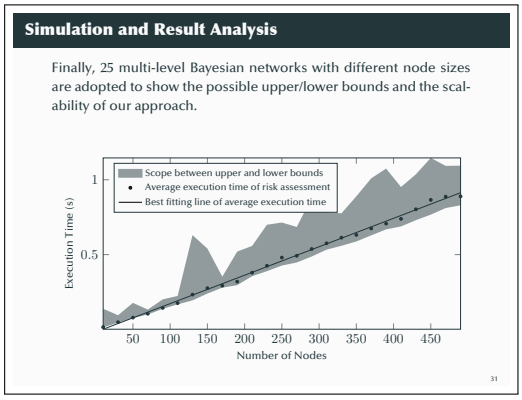

The third simulation.

We repeat the first simulation 5,000 times, and the execution time of 5,000 calculations is recorded. This simulation is run on a machine with Intel Pentium processor G3220 (3M Cache, 3.00GHz) and 4GB DDR3 memory. The following figure shows the distribution of the 5,000 execution times.



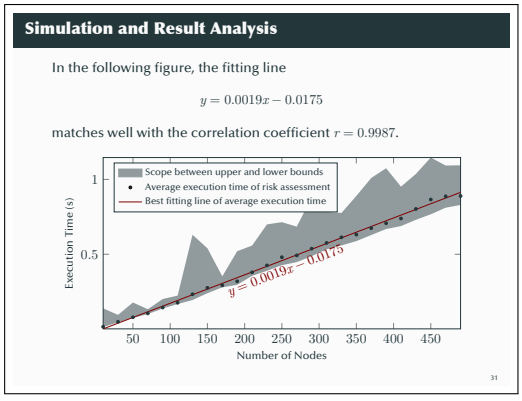

Some parameters of the following figure:

- 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.

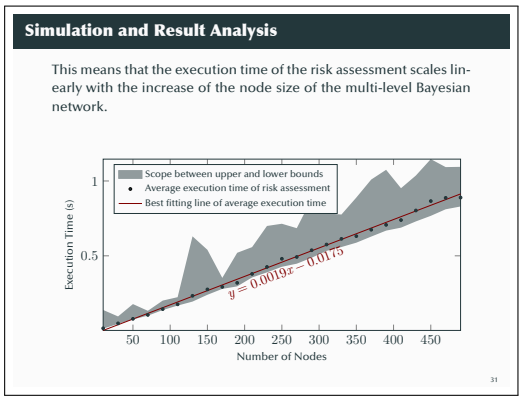



The final simulation.

25 multi-level Bayesian networks with different node sizes are adopted to show the possible upper/lower bounds and the scalability of our approach. The minimum node size is 10, and the maximum node size is 490, which can model extremely complicated control systems. For each multi-level Bayesian network, the risk assessment is repeated 200 times and all the execution time is recorded. The following figure shows the possible upper/lower bounds and the scalability of the proposed risk assessment approach.

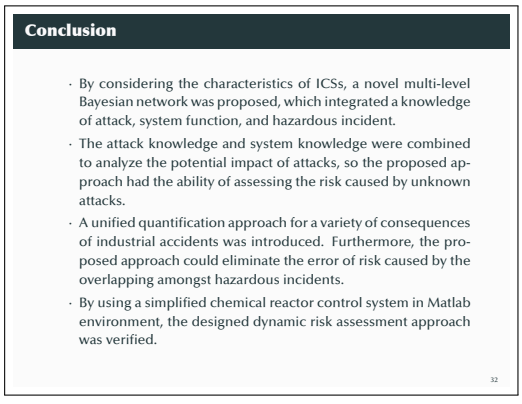



In this figure, a fitting line $y = 0.0019x - 0.0175$ matches well with the correlation coefficient $r = 0.9987$.




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.


Conclusion and prospect.



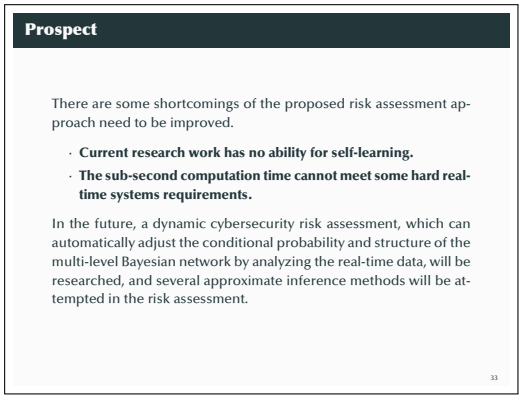

Now, let me make a conclusion for my presentation.

First, 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.

Second, 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.

Third, 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.

At last, by using a simplified chemical reactor control system in Matlab environment, the designed dynamic risk assessment approach was verified.

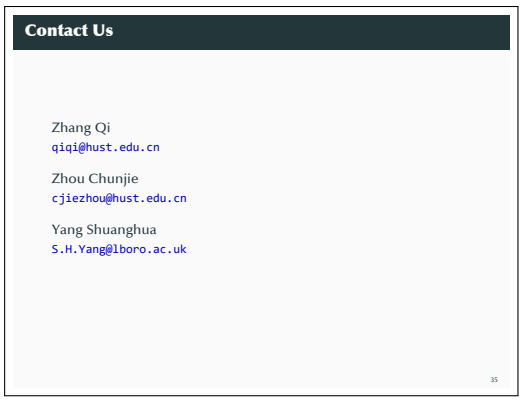



However, there are some shortcomings of the proposed risk assessment approach need to be improved.

- 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.


Thank you for your listening! I hope I have made myself understood.




If you have some questions to discuss with us, you can contact us by email.


We are very glad to welcome you to Huazhong University of Science & Technology.


Any questions?