Making Risk Models Operational for Situational Awareness and Decision Support

PR Paulson, G Coles, S Shoemaker

Pacific Northwest National Library
Richland, WA
USA

Abstract. We present CARIM, a decision support tool to aid in the evaluation of plans for converting control systems to digital instruments. The model provides the capability to optimize planning and resource allocation to reduce risk from multiple safety and economic perspectives.

1. Opportunities

Modernization of nuclear power operations control systems, in particular the move to digital control systems, creates an opportunity to modernize existing legacy infrastructure and extend plant life. We describe here decision support tools that allow the assessment of different facets of risk and support the optimization of available resources to reduce risks as plants are upgraded and maintained. The methodology can be applied to the design of new reactors such as small nuclear reactors (SMR), and be helpful in assessing the risks of different configurations of the reactors. Our tool provides a low cost evaluation of alternative configurations and provides an expanded safety analysis by considering scenarios early in the implementation cycle where cost impacts can be minimized. The effects of failures can be modeled and thoroughly vetted to understand their potential impact on risk. The process and tools presented here allow for an integrated assessment of risk by supporting traditional defense in depth approaches while taking into consideration the insertion of new digital instrument and control systems.

The primary discriminators of our approach is the consideration of multiple perspectives when assessing risk, so that decisions can be made at a policy level where appropriate, and to quantify the uncertainty that is sometimes unavoidable when assessing the reliability of digital components and cyber-security measures.

2. Design for an Operational Model

In this section, we outline the technical problems that must be addressed by assessing risk of upgrades to digital control and cyber systems in nuclear power plants.

Address Range of Threats: Assessment of risk is made with respect to a set of design basis threats [1]. Each threat is described in terms of the intent of the attacker and the capability of the attacker. For natural events or hazards not addressed by the design basis threats we specify the *effect* of the threat (gets equipment wet, causes abrasion, etc.) and the *intensity* of the threat. Associated with each threat for a particular plant is the relative likelihood that the threat will occur. Each of these dimensions must be addressed, since the likelihood of a threat affecting an asset depends both on vulnerability of the asset to the threat, which depends on the intent or effect, and the likelihood that the asset will be exposed to the threat, which depends on the capability of the attacker or the intensity of the hazard.

Field Cod

Multi-criteria Risk Assessment: The consequences that a risk-assessment methodology addresses affect multiple stakeholders. This is not just a matter of balancing economic return with public safety, but of also addressing different aspects of safety, such as the safety of workers, safety of public. Further, within each of broad categories of safety risk, independent assessment of long-term radiation exposure, for example, may have to be considered separately from risks that require evacuation. And this just considers two possible perspectives of consequences; others might include environmental concerns, security, and lost power generation.

Consequences of threats depend on stakeholder perspective: Each organizational mission may require a different stakeholder to determine which system components, types of threats, and which states of compromise present threats. Addressing the consequences of risks as they affect these different missions requires experts from multiple fields of expertise.

Stakeholders assessing risk to their missions are concerned with the *intent* of an attacker or the type of natural threat that an asset is exposed to; they most likely do not have the expertise to determine whether a particular asset will be exposed to a threat.

Technical expertise determines propagation of risk: The interaction of system components and the propagation of risk through the system are also modeled. While the impact of a compromised asset on a mission is determined by a stakeholder, how risks propagate through the system is determined by technical expertise and is independent of the organization's mission. When technical expertise provides conflicting assessments because of inconclusive evidence this result should be presented to the end user.

Quantify uncertainty associated with estimates of vulnerability: Because of the non-linearity and evolving nature of digital systems, traditional approaches of determining failure rates are not applicable [2, p. 2-5]. This means that in cyber systems, the best available estimates of vulnerability will be the currently "best practices" used by cyber-security experts for reliability and security.

Field Coc

It is important that the confidence of the risk assessment, and the elements of the assessment that lead to increased or decreased confidence, be clearly delineated. An example of where there is uncertainty in risk assessments of cyber components is the unknown effect of the supply-chain of digital equipment [3]. While not traditionally considered a source of threats in cyber security, recent attacks have raised the concern that trapdoors could be injected into digital control units at any of several points along their supply chain.

Field Coc

3. Overview of CARIM

We propose that the goals set forth in Section 2 can be met by a system with the following characteristics:

- Consequence of a compromised state on each mission is separately assessed.
- Propagation of faults and risks through a system are assessed by technical subject experts
- Each threat is characterized by an *intent* or *effect*, which is used by stakeholders to assess risk to their missions, a capability or intensity, which is used by technical assets to determine the tendency of the threat to propagate, and a *likelihood*, which is used in resource-allocation planning to determine the priority of addressing the threat.
- Multi-criteria optimization is used to determine how resources can be allocated to address the concerns of all stakeholders
- Uncertainty in risk estimates is tracked quantitatively so that decision makers have a basis on how much weight to give particular risk assessments in their planning

In order to determine which components in a proposed upgrade require the most thorough evaluation, two aspects of plant components are considered. First, the impacts that the component has on some modeled risk are considered. Secondly, the effect that a change in the state of the component has on other components is considered so that total risk can be assessed effectively.

By modeling the relationships along which risk propagates, attention can be focused on components that have the greatest potential to contribute to an improved risk assessment. Using Pareto-optimal multi-criteria optimizations, we can create plans that reduce risk across all stakeholder perspectives. The solutions do not trade, for example, an increase in public safety risk for a decrease in risk to economic viability, regardless of what weights were assigned to the importance of each criterion.

Because it is often impractical to do full-coverage testing of digital systems, we may need to rely on expert judgment and best practices to derive the likelihood that threats propagate through the system. For each collection of component types, expert judgment is used to determine what relationships between components will cause risk to be propagated through the system model. Costs of eliciting this knowledge can be amortized across multiple risk systems that use similar underlying components.

3.1. The Dimensions of Risk Assessment

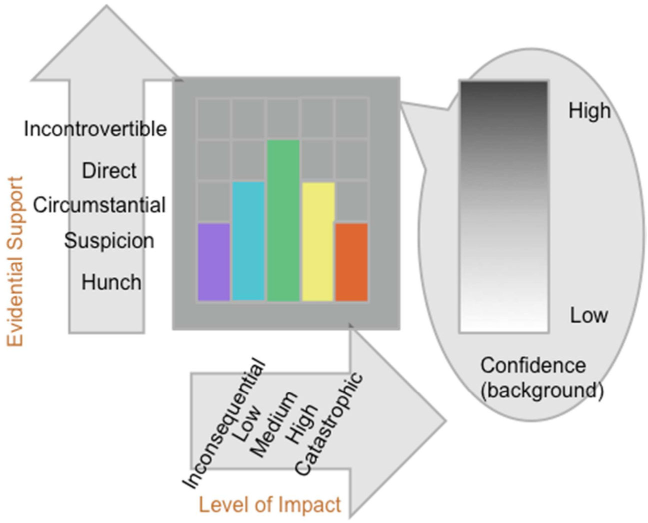


FIG. 1. Dimensions of Risk Assessment

To provide stakeholders as much information as possible about the risk assessment resulting from the model, we divide the risk assessment into 2 dimensions. The first dimension is the confidence, or certainty, to be afforded to the risk assessment. This measure is a self-assessment by the system and the experts providing input to the model of their level of confidence in their judgments given the available information and the ability to make successful predictions in their field. For example, a metallurgist might state with a great degree of confidence that a metal structure will withstand an impact at a particular level of forceand be able to cite experiments that have been performed to back up this judgment. On the other hand, an encryption expert will want to offers the judgment that a public/private key encryption mechanism with a sufficiently large and random key is unlikely to be cracked, but still wishes to hedge this judgment because they are not familiar with the implementation of the algorithm and are aware of new approaches to cracking hashing algorithms that may undermine public/private key encryption [4].

Field Cod

The second dimension of the risk assessment, the impact assessment, details the amount of evidence used to back up alternative hypothesis about the level of impact that is expected from the current set of threats and the state of compromise of the assets in the system. The impact assessment itself is broken down into two dimensions. The first dimension is the hypothesized level of impact, which ranges over an ordered set of discrete levels of impact, such as ‘inconsequential’ through ‘catastrophic’. The second dimension is a real value in [0, 1] that gives the relative amount of support for a particular risk level; the higher the value, the more evidence there is for that degree of risk.

4. Use Case Scenario

4.1. Assessing Direct Impact of Scenario on Stakeholders

Say we have 3 different stakeholders making decisions regarding operations at a plant. The first is in charge of worker safety, the second is in charge of environmental effects, and the last is in charge of plant production. These stakeholders are represented in the model as *risk perspectives*.

Say a cyber-vulnerability related to control of spent fuel pool level is in question. In one scenario, the level of water, while still well above the minimum required to prevent release of radiation, had dropped, indicating a potential problem in a pumping system. Stakeholders are asked how this potential state will impact their interests. We term the spent fuel pool an *asset* of the plant, and the fact that the water level has dropped a *state of compromise* of the asset. There are a number of factors to consider, such as how recently the fuel was discharged, time before vaporization or loss of level might occur, and the potential for recovery actions, but in the end the stakeholders must determine the consequence of diminished capacity; they do so by considering how this compromise state will affect their interest over a specific period of time, the *risk assessment period*. So in this example, the stakeholder in charge of worker safety will gage the risk of exposure to radiation for facility workers. The stakeholder in charge of environmental impacts (while aware of the near-term potential of drastic effects) needs to determine the potential for release of radioactivity in the event fuel is breached from overheating (also a public safety concern). The stakeholder in charge of production might say that this particular state of affairs has no direct impact upon current production because the plant is unlikely to shut down.

When assessing the upgrade to the control system for the pool's pump system, given the estimates of impacts from all stakeholders for all assets of the plant to the current set of threats types, we are in a position to determine what effect our actions will have in decreasing the potential impact across the interests of all stakeholders. In particular, we seek to determine which components of the upgrade have the greatest impact on the positions of the stakeholders. These components could be targeted for more rigorous acceptance tests or have more scrutiny applied to suppliers.

Our methodology makes a distinction between the direct impact of compromised assets and the potential impact of the compromised asset. The stakeholders described here are only concerned with the direct impact of the described scenario on their interests, not the potential impact of the scenario. The indirect impact of compromised assets is separately modeled by considering the degree to which the compromised state affects the other assets of the facility. This is accomplished by specifying the *mitigating role* each asset plays for any other asset in the facility. In this example, the spent fuel pool is keeps the spent fuel cool and provide shielding; the fuel rods are another asset of the plant. Further, the spent fuel pool makeup and recirculation pumps play part of the coolant-supply role for the fuel basin.

In addition to modeling the immediate impact of each compromised asset in terms of stakeholders, we model the effect of the state change in terms of the other assets of the plant. To model this aspect the model uses the current state of an asset playing a mitigating role—such as the coolant pump for a particular application—to determine the state of compromise of the mitigated asset. For example, if in the current risk assessment period a pump fails to run because of a fault in the power supply, but in the next risk assessment period we expect the pump to be operational, than we may be able to resume normal plant operations. The identification of the fillers of mitigating roles is completed during the construction of the operational risk assessment model.

4.2. *Assessing Propagation of Risk*

In order to determine the effect that an inoperative water pump will have on the effectiveness of the spent fuel pool, we consult with technical experts. These experts would consider whether there is a backup for the water pump, the current state of the spent fuel pool, and whatever considerations they deem appropriate to determine if an inoperative water pump will affect the state of a spent fuel pool it is providing water for. These technical personnel will consider not only the current state of the water pump but also the existence of other hazards, such as seismic activity that could affect the performance of the pump. Note that these experts are also considering only a limited number of influences: the state of the water pump and a limited set of external factors that could influence the effectiveness of the water pump in fulfilling its role to provide water to the spent fuel pool. Considerations of other factors that affect the state of the water pump are modeled separately. Again, the experts when providing their estimates consider the risk assessment period.

In some cases, technical experts must shade their assessments based on the maturity of the knowledge they are basing their judgment on. The expert might be fairly certain that in the event of a water pump failing that the control system will correctly fail back to a backup water pump. However, if the control systems for the plant have obtained from a new provider, there might be a chance that the new control components have been infiltrated with malware; this might lead the expert to make second guesses about the likelihood that the backup pump will be effective.

5. Using the Use Case to Design a Risk Modeling Methodology

Up until now, we've looked at a small slice of the risk assessment that would be done in response to an event in an operational setting. But we want to use the same method for planning and resource allocation for plant upgrades and redesign. For the assessment of direct impact, the steps are pretty much the same, except that the set of threats to be addressed is defined, each stakeholder is identified in advance, the assets identified that each stakeholder deems relevant to their interests, and the potential compromised states of the assets need to be identified.

For the propagation of risk, for each asset the mitigating relationships of the asset is identified and the potential compromised states of the mitigating assets are determined. In addition, the expected type and severity of each modeled threat or hazard is determined, and an estimate is made of the likelihood of an occurrence of that type and severity of threat. Finally, technical experts are used to determine the likelihood that the state of a mitigated asset will change in response to a change of the state in one of the mitigating assets in combination with a specific type/severity of threats.

5.1. Amortization of Modeling Costs

Collecting all the estimates, asset definitions, and possible mitigating relationships requires extensive interaction with stakeholder representatives and technical experts. Determining the threats, including their severity and likelihood of occurrence, will likely require original research into the historical record and consultation with actuaries. This cost will likely not be made up in the operation of a single plant. However, the knowledge collected about the structures of mitigations, types and compromised states of assets, identification of stakeholder classes, and estimates of risk propagation and final impact on shareholders, and types, severities, and relative likelihood of threats can be structured in such a way that it can be applied to many different plants and even reused within the risk model of a single plant.

5.2. Applying Appropriate Knowledge to a Specific Plant

The collected risk assessment knowledge is applied to a specific plant by identifying the plant's physical and cyber assets and determining which of the asset types that have been identified by domain experts each asset belongs to. Once this is done, the collected knowledge about possible mitigations can be used to elicit from plant managers relationships between plant assets that would allow risk to propagate between assets. The existing knowledge-base of types, severities of threats can be examined and a determination made about the likelihood that the threat may occur at the plant. Finally, the default assessments on the impact of compromise of assets on stakeholder interests can be tailored to the specific plant.

5.3. Generation, Maintenance, and Use of the Operational Risk Model

Once all plant-specific information has been collected, the operational risk model is generated. The operational risk model provides an assessment of risk for the current risk assessment period; the assessment is based on available information about the current state of assets and the current threat environment from the perspective of each stakeholder, as outlined in Section 3.1.

The risk model is updated to reflect changes in the state of assets and the threat profile of the plant. These changes require minimal changes in the running model; external systems such as intrusion detection systems could update the model automatically. In addition, changes to the systems architecture can result in changing the types of assets and their relationships to other assets. The model provides a mechanism to examine how such changes will affect the risk profile of the plant.

6. Details of Risk Model Development

The Risk Model Development Methodology consists of two phases. The first phase, development of the domain model, can be performedindependently of a specific plant. Once developed, the domain model can be used as the basis for multiple plant-specific models. The second phase is the development of the plant-specific model, in which particular assets, and their importance to the goals of the enterprise, are identified and classified according to the domain model. In this phase, the domain model can also be modified and amended as needed for the needs of the specific plant. The third phase, the generation and operational use of the plant-specific operational model, can be entirely automated.

6.1. Threats

Each threat or hazard has three components, the intent/type of threat/hazard, the capability/severity of the threat/hazard, and the likelihood that the plant will face the particular threat/hazard within the risk assessment period. The severity is used to determine the likelihood of compromise; the threat type is used to by stakeholders to determine the impact of compromise. For example, one type of hazard might be exposure to moisture. Specific types of attackers also are associated with intents; one threat type could be activist activity, another might be terrorist activity. The assumption here is that the activist organization’s goal is to demonstrate the ability to infiltrate cyber operations, while the terrorist organization intends to use the access to cause the plant to malfunction.

Each threat/hazard type has an associated totally ordered set of *threat severities*. The severity indicates the strength behind a particular threat; for example, the severities associated with exposure to moisture might be {high humidity, mist, driving rain, immersion}, while the severities associated with attackers would indicate the level of expertise of the attacker: { naïve, script-kiddie, sophisticated, state-backed}. The ordering of threat severities indicates that each high level of severity includes all the consequences of lower level of severities: a piece of equipment that has been submerged is at least as damaged as one exposed to driving rain.

6.1.1. Threat Likelihood

Given the known threats for a plant or domain, we impose a *threat likelihood* ordering on the threats. A threat-likelihood ordering is on an interval scale; we can specify, for example, that an attack by an eccentric billionaire mad scientist, while possible, is unlikely; however, an attack by a terrorist that wants to cause as much media attention is quite likely, thought such a threat might not have the capability associated with a state-backed effort.

6.2. Domain Model Development

The domain model consists of the identification of the capability/intensities of the threats to be modeled, the types of assets that are to be modeled in the domain, and the mitigation-relationships that exist between assets in the domain.

6.2.1. Asset Types

There is an asset typefor each distinct type of asset in the problem domain. The asset types differ from each other in how they compromises are mitigated within the domain and the types of compromise they are subject to. Examples of asset types include physical assets such as cooling pools, reactor cores, and physical CPUs and network cable to cyber assets such as server processes to abstract assets such as training programs and security policies. Each asset type is associated with a totally-ordered set of levels of *severity of compromise*. Severity of compromise are related to threat severities: the greater a threat severity, the more likely an asset type is to move from one level of compromise to a higher level. For a server, the level of compromise might be the access privileges obtained by a cyber-attacker, for a particular file, severity might indicate the degree of access obtained by an attacker: read-access, write-access, or ownership. For a storage pool, the level of compromise might indicate the storage pool temperature and whether it is rising. What compromises an asset type depends to a large degree on the detail desired in the model. For example, if multiple aspects of spent fuel pools, such as physical security and operational security, are to be modeled, then new asset types can be created for these aspects and related to the spent fuel pool through *mitigation relationships*.

6.2.2. Mitigation Relationships

The mitigation relationships specify, for each asset type, relationships to other asset types that affect the probability that a *mitigated asset* will change its level of compromise. Each *mitigation* *specification* specifies a *mitigating role* and the asset types if the *mitigated asset* and the *mitigating asset*. For example, the mitigations specified for a server might include a policy regarding security updates, the access control list settings for server software, and virus protection software. The mitigations for a cooling pool might include physical security and backup circulation pumps.

6.2.3. The Domain Model

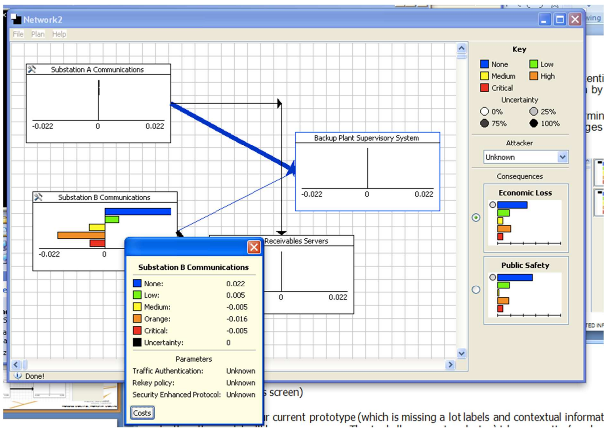
The domain model specifies, for each threat severity, mitigation relationship, and level of compromise of the mitigated asset type and the mitigating asset type, the likelihood that an asset of the mitigated type will move to the next most compromised level. This map is constrained in that it is always more likely that a more severe threat is more likely to cause a state change than a less severe threat. Uncertainty in the domain is modeled by associating a degree of uncertainty with the values in the domain model. A level of uncertainty is associated with the judgement of how a mitigation affects the relationship between a mitigating asset and a mitigated asset in the face of a particular threat severity.

6.3. Plant Model Development

A plant model of plant consists an ordered set of impact levels, a set of risk perspectives, which are associated with the missions of the plant’s organization, and the asset types, mitigation relationships, and current state of each asset in the plant.

The impact levels of a plant are a totally-ordered set of impact levels that can be used to partially normalize the impact to different stakeholders. The set of impact levels used in FIG. 1 , in ascending order, is `(inconsequential, low, medium, high, catastrophic).` Each risk perspective for a plant maps the compromised state of a particular asset exposed to a particular threat type to an impact level. In FIG. 3., the risk perspectives are `(public safety, economic cost, reliability, and competitive advantage).`

A plant model specifies a set of assets and an appropriate asset type for each asset. For each mitigation relationship in which that asset type is the mitigated asset, the plant model specifies which asset (of the appropriate asset type) is the mitigating asset for that relationship. The current state of the plant consists of the assignment of asset state to each asset.



Substation B Communications

🟦 None: 0.022
🟩 Low: 0.005
🟨 Medium: -0.005
🟧 Orange: -0.016
🟥 Critical: -0.005
■ Uncertainty: 0

Parameters

Traffic Authentication: Unknown
Rekey policy: Unknown
Security Enhanced Protocol: Unknown

Costs

FIG. 2. Tracing threats through mitigation relationships

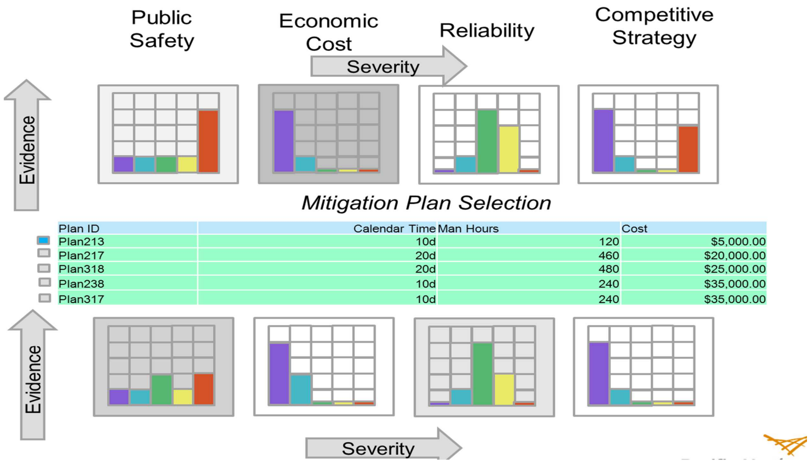
7. Using the Operational Model

Assessing Current Levels of Risk: The model provides a summary of the risk profile of the plant in terms of all risk perspectives. This allows each stakeholder to determine changes in the assessed risk to their mission and the level of uncertainty in the assessment. When multiple stakeholders see increases in their exposure the resource allocation tools described below can be used to find mitigations that can reduce exposure across all missions. If it is found that tradeoffs between missions must be made those decisions can be made at a policy level rather than delegating the decision to security analysts.

Tracing Exposure to Assets and Threats: when the risk profile indicates exposure, the model can be traced back along sources of the exposure to find which mitigations could be applied to immediately decrease exposure. This ability can be used both operationally as a mechanism to address threats as they arise and as a tool to explore interaction of plant assets and the plant's risk profile.

Resource Planning for Multiple Perspectives: The risk model can also trace back through propagated risk so that stakeholders can determine which assets are influencing the assessed impact to their interests (FIG. 2.). As new information becomes available, the model can be altered to modify the state of compromise of assets, to modify the mitigation relationships between assets, and to modify the likelihood of specific threats. The model can also be 'rolled forward' to future risk assessment periods to determine if the current state of the plant, while secure, might degrade unless actions are taken.

In addition to providing a snapshot of the current state of the plant, we use the model to determine what actions can be taken in order to decreased the assessed impact across all risk perspectives through the use of Pareto-optimal resource allocation based on the cost of adding mitigations and effectiveness of the mitigations in reducing assessed impact. Resource allocation plans are generated based on the projected cost, including financial cost and manpower resources, and the foreseen benefit to the risk profile of potential mitigations. As shown in FIG. 3., when a generated plans are summarized by cost in man-hours, calendar time, and dollar cost. Selecting a particular plan allows one to compare the current risk profile with the profile once the plan has been completed.



	Plan ID	Calendar Time	Man Hours	Cost
■	Plan213	10d	120	$5,000.00
□	Plan217	20d	460	$20,000.00
□	Plan318	20d	480	$25,000.00
□	Plan238	10d	240	$35,000.00
□	Plan317	10d	240	$35,000.00

FIG. 3. Resource Allocation Plans

8. REFERENCES

[1] International Atomic Energy Commission, *Development, Use and Maintenacne of the Design Basis Threat*, International Atomic Energy Commission: Vienna (2009).

[2] T. L. Chu, G.M.G., Mu. Ye, J. Lehner, P. Samanta, *Traditional Probabilistic Risk Assessment Methods for Digital Systems*, U.S. Nuclear Regulatory Commission (2008).

[3] McFadden, F.E. and R.D. Arnold. *Supply chain risk mitigation for IT electronics*. in *2010 10th IEEE International Conference on Technologies for Homeland Security, HST 2010, November 8, 2010 - November 10, 2010*. 2010. Waltham, MA, United states: IEEE Computer Society.

[4] Xie, S.-Y. and B. Xu. *A publicly verifiable authenticated encryption scheme without using one-way hash function*. in *6th International Conference on Machine Learning and Cybernetics, ICMLC 2007, August 19, 2007 - August 22, 2007*. 2007. Hong Kong, China: Inst. of Elec. and Elec. Eng. Computer Society.