EAI Endorsed Transactions
on Security and Safety
Research Article ICST.ORG

Assessing Security, Capacity and Reachability of a Heterogeneous Industrial Network during Planning Phase

Apala Ray[1,2,*], Johan Åkerberg [1,2], Mats Björkman [2], Mikael Gidlund [3]

[1]ABB AB; Corporate Research
[2]Mälardalen University; School of Innovation, Design, and Engineering
[3]Mid Sweden University

Abstract

In an industrial plant, there is usually a mix of devices with different levels of security features and computation capabilities. If a mix of devices with various degrees of security features and capabilities communicate, the overall network dynamics with respect to security and network performance will be complex. A secure communication path with high latency and low bandwidth may not satisfy the operational requirements in a plant. Therefore, there is a need to assess the relation of security and network performance for overall plant operation. In this work we focus on identifying an optimal flow path between two devices in a multi-hop heterogeneous network. We propose a model and an algorithm to estimate and generate a network path identified by flow performance indicators of a heterogeneous communication network. Through an example, we show how the flow performance metrics change with security, capacity and reachability of the devices in the network.

Received on 30 May; accepted on 01 June 2016; published on 08 December 2016

Keywords: Security Modeling, Network Assessment, Routing, Path Planning

doi:10.4108/eai.8-12-2016.151728

1. Introduction

The goal of industrial automation is to automate the operations involved in industrial processes, such as pulp and paper, water and wastewater, food and beverages, mining etc., with minimal or reduced human intervention. A growing concern of cyber threats towards industrial plants has prompted industrial practitioners to focus on secure communication solutions. In the initial phase of industrial automation, networks in industrial plants were not connected with the outside world. In the last decades, there is a growing trend to connect different sites in industrial plants via Internet to one or more centralized control servers for the purpose of reducing cost of operations. With the advent of industrial embedded system development, new hardware and concepts for devices with secure technology, reliable communication, and high-end computing capabilities are being developed. Technological advances in terms of computing power and communication capabilities bring operational benefits inside plants, but also increase the exposure to cyber security attacks. For example, when a smart sensor is developed to improve the capability of traditional analog sensors, it can come with wireless communication support for easy installation. This has opened the opportunity of remote monitoring and configuration of those devices but also brought security threats from the wireless domains.

Advancement of communication technologies also creates heterogeneity in plant networks. Typically an industrial plant has a mix of different industrial communication protocols and these industrial protocols are usually common across several industries. These industrial communication protocols require specific hardware and software for robust and reliable operation inside the plants. Along with advanced devices, there will be traditional devices inside plants without firmware and with no computational power, as there will not be enough

★A poster version of this paper is published in the Springer LNICST proceedings of SECURECOMM 2015

*Corresponding author. Email: apala.ray@in.abb.com

business motivation to replace all traditional devices for not having advanced technology support. In addition, with the inherent benefit of wireless communication, there will be parallel infrastructures with wired and wireless communication inside plants. These typically make an industrial plant a heterogeneous network where devices with varying capabilities communicate with each other for the successful operation of the plant. Therefore, there is a requirement to explore solutions which can allow overall reliable and secure operation of plants considering the heterogeneity of industrial communication networks. Along with cyber security requirements of industrial plants, it is also necessary to consider other important requirements of plants in terms of network performance. For example, the exchanged information in monitoring/supervision related applications is generally not sensitive to packet losses and jitter, rather it is designed for maximum throughput. On the other hand, closed loop control and interlocking and control are sensitive to jitter and delays, where network predictability of worst case delays is a paramount aspect. In this paper, we will use the term "reachability" as a timeliness performance metric where higher reachability implies lower delay. Availability is another key parameter which both industrial plant operation and security need to achieve. A widely used security concept is to segregate secure networks from insecure networks by the use of firewalls and demilitarized zones (DMZs). However, the scenario in industrial communication can be different, where a communication network itself has devices with varying security and reliability capabilities. From a system level perspective, instead of isolated communication islands in the core network, it is useful to have an understanding of a network-wide security, capacity and reachability of devices with both old and new technology. Therefore, it is important to understand the network capabilities during network design to assess the required network performance in a heterogeneous networked system.

Heterogeneous industrial networks like smart grid networks aim for increased capacity, reliability and efficiency of existing electricity grids by using cyber technology. The heterogeneous traffic generated by the smart grid can be routed through Internet, wireless, cellular networks, dedicated fiber optic, or power line networks [1]. The various available communication networks provide different performances in terms of bandwidth, delay and packet delivery rate. Therefore, identifying a communication network path which can support heterogeneous traffic generated by the smart grid is an important design decision [2]. From a very high level, the goals of networking and security are somewhat contradictory. Where networking research aims to enable rapid packet transfer between one node to another node, achieving a high average throughput, security research aims to inspect a packet before forwarding it to the next node. Moving from these two extreme viewpoints, we need to consider network design and the security requirements as complementary to each other. A secure communication data flow with high latency and low bandwidth may not satisfy the required operational benefits in a plant. In a multi-hop heterogeneous network, data communication between source and destination can be possible through multiple subnetworks involving devices with varying capabilities. The problem is that some sub-system of networks can score high on one particular performance parameter, whereas score very low on other performance parameters. For example, mesh networks with ZigBee communication might have limitation of data rate but can improve reachability through the inherent flexibility of extending a network. Fiber optic cable based communication can support high data rates but brings in additional cost of fiber optic laying and thus reduces the flexibility of extending a network. If the decision of choosing a communication flow path between a source and destination is done based on one performance criteria only, such as security or path reachability or link capacity, then one segment of a network may be overloaded. This may create instability in the overall plant operation, for example, a highly secure path may introduce large delay or provide less bandwidth. Therefore we focus on analyzing the systems globally, identifying flow paths based on application requirements and directing resources efficiently to increase the confidence in the system. The network planning phase should capture the properties of the system and identify constraints on the network to achieve an overall secure solution. This ensures efficient network planning keeping security in mind and gathering information from the industrial network regarding capabilities and vulnerabilities to identify the best path for the data.

In this paper, we focus on identifying a secure, high capacity and reachable path between two sub-systems in a multi-hop heterogeneous network. We explore how a network flow path can be chosen inside a plant between two sub-systems, where the network will balance the required levels of communication security, capacity and reachability. For an efficient flow path estimation, our model requires the topology of the system along with the performance related attributes as input. We also identify the key performance indicators required for successful operations of a network. Then we individually analyze the effect of each key performance indicator on a network flow. Based on this analysis, we explore the effects of a local performance indicator of each sub-system on the global performance indicator of a flow path keeping overall security, capacity and reachability in the system. This helps us to rank each communication flow path based on the key performance indicators of the network. This information is useful when designing a plant with a service level agreement. Thus we identify a secure,

high capacity and reachable path by computing the final weight of each flow path from each key performance parameter of the sub-systems involved in that network flow.

The paper is organized as follows. Section 2 discusses related work. Section 3 presents the network model and section 4 analyzes the model from the simulation and observations. Finally, conclusions are presented in section 5 along with future work.

2. Related Work

Managing heterogeneous devices in industrial networks opens up an interesting area for managing security in industrial plants with the notion of "trust". The trust can be associated with the confidence of an entity about the behavior of a particular entity. There has been an extensive research on trust management and in the research communities, different trust models have been proposed by researchers for distributed systems, point-to-point networks, agent based systems, pervasive computing, MANETs etc. Many of the available trust models rely on authentication or cryptography [3–6]. However, in this paper we focus on the trust management scenarios where all the entities inside the industrial plants cannot support authentication through cryptography. There are some works related to the network-wide security analysis for computer networks or Internet. A unified framework is provided to study the effects of packet filters, routing policy, and packet transformations on the network reachability [7]. A formal method approach for verification of security constraints on networks with dynamic routing protocols is proposed in [8]. Problems associated with the design and security analysis of network protocols which use cryptographic primitives are addressed in [9]. There is another set of work where different models are used to assess network security. Security monitoring and incident modeling by combing automated analysis of data from security monitors and system logs with human expertise is shown in [10]. In [11], an approach of cyber security research through experiments has been shown. In [12], a backscatter analysis method is presented which can provide an estimate of worldwide denial-of-service activity. There is some research on attack graph construction and performance evaluation [13]. In [14], a two layer attack graph is proposed. Attack models also can be used to assess network security. A hierarchical attack representation model is proposed in [15], where a two-layer hierarchy is proposed to separate the network topology information from the vulnerability information of each host. A ranking scheme to identify a relevant portion of the attack graph is proposed in [16]. In [17], a framework for an experimentation environment for network industrial control system is proposed which can concurrently reproduce physical and cyber systems. In [18], integrating the automation system and electric power networks for assessing vulnerabilities is presented. A simulation based security impact assessment method is proposed in [19]. The interdependencies of Process Control Systems with ICT systems and security challenges in SCADA systems is discussed in [20]. In [21], an approach to the security assessment based on the attack graphs is proposed. There is also a set of trust models that rely on experience and recommendation from already connected devices [22, 23]. In these types of trust models, time is an important parameter as in real life scenarios building trust takes time. Therefore, when a new device joins a network, other neighbors need direct experiences with this device in order to recommend this new device to others in the network. This brings down the trust management problem to its initial value assignment.

In this paper, we provide a model which can be used during network design to identify optimized network paths. We study the overall network dynamics with respect to security and robustness. Further, we analyze how to assess a secure, high capacity and reachable path between two end nodes in a multi-hop heterogeneous network. We do not intend to make run-time analysis of network traffic or generate attack graphs; rather we focus to assign initial paths in the planning phase based on network architecture and device property. To summarize, we provide a model to estimate and generate a network path identified by flow performance indicators of a heterogeneous communication network.

3. Method for Secure, Reachable and High Capacity Path Identification

In this section, we focus to identify a secure, reachable and high capacity flow path between two end sub-systems. To identify the flow path, we need to model the network topology of the system along with the performance attributes. Once we have a mathematical model, then we provide an algorithm to find a balanced path for the given network. In this work, we suggest to assign initial trust in the planning phase based on network architecture and device properties. Figure 1 presents the architecture of the component required for path identification.

The components involved in this architecture are a set of global performance parameter mappers and security assessment algorithms. The local performance metrics, such as security attributes, link bandwidth and distance are associated with the sub-system and link properties. From the local performance metrics, the global performance parameters, such as flow security, flow bandwidth, and flow reachability are mapped and this input along with the flow identifier is fed into the security, capacity and reachability assessment

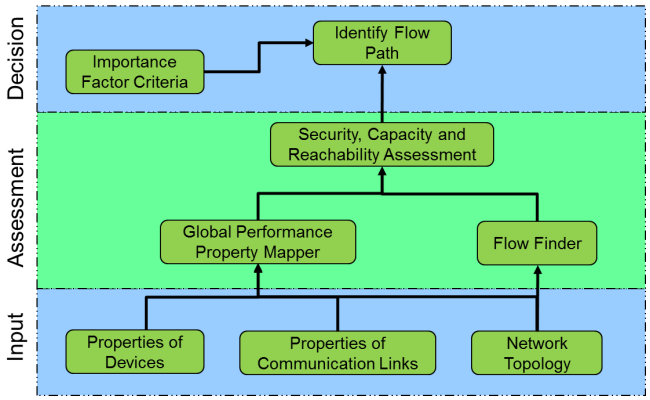

Figure 1. Architecture for Secure and Robust Path Identification

component. Along with the importance criteria of network, the optimized flow path is identified.

3.1. Formulation of communication network model:

In this section, we specify a network model to calculate the normalized flow value between two sub-systems in order to be able to compare the effect of performance parameters on identification of the chosen flow path. We represent a communication network of sub-systems by using a directed acyclic graph. A network (N) is a set of sub-systems (D) and the link (L) between sub-systems. A sub-system can contain an internal network of devices for communication purpose. For example, a ZigBee mesh network or a local area network is considered as a sub-system. We also denote a sub-system as a node and we will use both the terms interchangeably throughout this paper.

$$N = \{D, L\} \tag{1}$$

We consider the sub-systems to be the vertices in the graph and the links between communicating sub-systems are the edges.

$$D = \{d_1, d_2, d_3, ...d_{n-1}, d_n\} \tag{2}$$
$$L = \{l_{d_1,d_2}, l_{d_2,d_3}, ...l_{d_{n-1},d_n}\} \tag{3}$$

Links are ordered pairs of sub-systems. Therefore, l_{d_1,d_2} implies that the link is directed from sub-system d_1 to sub-system d_2.

For a sub-system d_i in the network, the number of inward directed links is known as the indegree and is denoted as $deg^-(d_i)$. The number of outward directed links is called the outdegree and is denoted as $deg^+(d_i)$. A sub-system d_i with $deg^-(d_i) = 0$ is known as a *source sub-system* in the network and a sub-system d_i with $deg^+(d_i) = 0$ is known as a *sink sub-system* in the network. We focus to estimate the optimized path between a *source sub-system* and a *sink sub-system*.

A flow path (f_{d_i,d_j}) from source sub-system d_i to the sink sub-system d_j is an ordered pair of links including all intermediate sub-systems $\{d_i, d_{i+1}, ..., d_{j-1}, d_j\}$ as,

$$f_{d_i,d_j} = \{l_{d_i,d_{i+1}}, l_{d_{i+1},d_{i+2}}...., l_{d_{j-1},d_j}\} \tag{4}$$

In a network, there will be multiple paths possible between two end sub-systems. If there are n paths possible between two sub-systems d_i and d_j, then we denote each flow path between two sub-systems as, $f^k_{d_x,d_y}$, where k is the index of available n path between sub-systems $d_x, d_y \in D$.

3.2. Formulation of input specifications:

Having defined the sub-systems, links and flows in the network, we formulate the local performance metrics of sub-systems so that it is easy to understand their effect on the global performance parameters of a network.

First we define the performance metric *node assurance value* $\alpha(d_i)$, as a probabilistic measure of confidence about the security properties of a sub-system $d_i \in D$. The *node assurance value* is used to determine the trustworthiness of sub-systems in a heterogeneous network. The inherent security properties of a sub-system are used to assign the initial *node assurance value*. The capabilities of each sub-system in terms of physical protection, cryptographic capability, key distribution support etc. can be used to assess the initial assurance value of a sub-system. Then we can estimate the probabilistic confidence about the security properties of a sub-system $d_i \in D$ and flow path.

$$0 \leq \alpha(d_i) \leq 1 \tag{5}$$

In our previous work, we have shown a model to estimate and predict *node assurance value* in a heterogeneous communication network [24]. To get the *node assurance value* of all the nodes in a network, we traverse the network graph from all leaf nodes. We update the *node assurance value* of every node based on the *node assurance value* of their incoming nodes. Our model aligns with the intuition that the vulnerable node can reduce its assurance value if the previous nodes in the flow path are also vulnerable. This model also captures that the assurance value of a node may be lowered when the node communicates with heterogeneous network devices.

The next performance metric we define is the *link capacity* $\beta(l_{d_i,d_j})$, as the bandwidth of the link between two sub-systems $d_i, d_j \in D$ and $l_{d_i,d_j} \in L$. The *link capacity* is the theoretical upper bound on the rate at which messages can be reliably transmitted over a communication link between two sub-systems. In this network model, we do not consider the throughput of

the system which can be an average rate of successful message delivery over a communication link. Therefore, the throughput of a link will be less than the *link capacity*.

$$0 < \beta(l_{d_i,d_j}) < \infty \tag{6}$$

We define another performance metric *hop count* $\gamma(f_{d_i,d_j})$, as the number of hops between two sub-systems $d_i, d_j \in D$. A hop count refers to the number of intermediate sub-systems from source through which message should pass to reach destination. This can be basic measurement of reachability in a network. The hop count is the cardinality of the flow path set f_{d_x,d_y}.

$$\gamma(f_{d_i,d_j}) = |f_{d_x,d_y}|, \text{ where } 1 \leq \gamma(f_{d_i,d_j}) < \infty \tag{7}$$

3.3. Formulation of security, capacity and reachability assessment:

In this section, we focus on formalizing the parameters for security, capacity and reachability assessment. Once we have defined the three local performance metrics *node assurance value*, *link capacity* and *hop count*, we define global performance parameters of a network.

First we define *flow assurance value* $\lambda(f_{d_x,d_y})$ for any flow path by multiplying *node assurance values* α of sub-systems in that flow. This gives an estimation of the vulnerability of a flow path from the sub-systems involved in that particular flow. If a flow has more vulnerable sub-systems in its path, the chance of being affected by those sub-systems is also higher for that flow. If we assume that the flow path $f_{d_x,d_y} = \{l_{d_x,d_1}, l_{d_1,d_2}, l_{d_2,d_y}\}$, then the *flow assurance value* $\lambda(f_{d_x,d_y})$ of flow f_{d_x,d_y} is,

$$\lambda(f_{d_x,d_y}) = \alpha(d_x)\alpha(d_1)\alpha(d_2)\alpha(d_y) \tag{8}$$

Next we define *flow capacity value* $\zeta(f_{d_x,d_y})$ for any flow path by taking the minimum *link capacity values* β of links in that flow. The minimum *link capacity value* provides a theoretical upper bound on the rate at which messages can be reliably transmitted between source and destination though multi-hop path. The capacity of a flow cannot support bandwidth more than this minimum value. If we assume that the flow path $f_{d_x,d_y} = \{l_{d_x,d_1}, l_{d_1,d_2}, l_{d_2,d_y}\}$, then the *flow capacity value* $\zeta(f_{d_x,d_y})$ of flow f_{d_x,d_y} is,

$$\zeta(f_{d_x,d_y}) = \min(\beta(l_{d_x,d_1}), \beta(l_{d_1,d_2}), \beta(l_{d_2,d_j})) \tag{9}$$

We also define *flow reachability value* $\eta(f_{d_x,d_y})$, as the inverse hop count between two sub-systems $d_x, d_y \in D$ in the flow path f_{d_x,d_y}. The *hop count* is a simplistic measure of reachability in the network. With increase of the *hop count*, the reachability is decreased.

$$\eta(f_{d_x,d_y}) = \frac{1}{\gamma(f_{d_i,d_j})} \tag{10}$$

In this work, we aim to find out the optimal path optimizing *flow assurance value*, *flow capacity value* and *flow reachability value*, given the set of k paths between source sub-system to the destination subsystem. This is a multi-objective optimization (MOO) problem. For a nontrivial multi-objective optimization problem, generally there does not exist a single solution that simultaneously optimizes each objective. However, there will be a requirement to select only one or a reduced number of final solutions, as the "decision maker" needs to identify the most preferred one among the solutions. This is often a non-trivial task for an operator and certain guidelines are necessary. One of the solutions is to combine the multiple objectives into one single-objective scalar function. One such approach is known as the weighted summation method.

We define *weight value* $\omega(x)$, as the importance factor based on the application requirement in the industrial plant. This gives a relative weight of a particular characteristic in a state and allows to give more importance to one characteristic over another characteristic.

$$\omega(\lambda) = \text{importance factor for flow assurance value} \tag{11}$$

$$\omega(\zeta) = \text{importance factor for flow capacity value} \tag{12}$$

$$\omega(\eta) = \text{importance factor for flow reachability value} \tag{13}$$

We define *flow value* $\psi(f_{d_x,d_y})$, as the weighted summation of *flow assurance value*, *flow capacity value* and *flow reachability value*. This is the combination of global performance parameters of a network *flow assurance value*, *flow capacity value* and *flow reachability value*.

$$\psi(f_{d_x,d_y}) = \lambda(f_{d_x,d_y})\omega(\lambda) + \zeta(f_{d_x,d_y})\omega(\zeta) + \eta(f_{d_x,d_y})\omega(\eta) \tag{14}$$

3.4. Computation of Normalized Flow Value:

We have seen that there might be more than one flow path between two sub-systems $d_x, d_y \in D$ in a given network. If there are n paths possible between two sub-systems d_i and d_j, then we denote each flow path between two sub-systems as, $f^k_{d_x,d_y}$, where k is the index of available n path between sub-systems $d_x, d_y \in D$. For each flow path $f^k_{d_x,d_y}$, we have different global performance parameters *flow assurance value*, *flow capacity value* and *flow reachability value* based on local performance metrics *node assurance value*, *link capacity* and *hop count*. The range of *node assurance value*, *link capacity* and *hop count* are in different scales. Therefore, we use normalization to adjust $\lambda(f_{d_x,d_y}), \zeta(f_{d_x,d_y}), \eta(f_{d_x,d_y})$ which are measured in different scales to a *0 to 1* scale to calculate the flow

value $\psi(f_{d_x,d_y})$. We used normalization to factor *sum to unity* so that these values sum to 1.

We define *normalized flow assurance value* $\tilde{\lambda}(f^k_{d_x,d_y})$ of a flow path $f^k_{d_x,d_y}$, as the *flow assurance value* of that flow path divided by the summation of all *flow assurance values* of n flow paths.

$$\tilde{\lambda}(f^k_{d_x,d_y}) = \frac{\lambda(f^k_{d_x,d_y})}{\sum_{k=0}^{n-1} \lambda(f^k_{d_x,d_y})} \tag{15}$$

We define *normalized flow capacity value* $\tilde{\zeta}(f^k_{d_x,d_y})$ of a flow path $f^k_{d_x,d_y}$, as the *flow capacity value* of that flow path divided by summation of all *flow capacity values* of n flow paths.

$$\tilde{\zeta}(f^k_{d_x,d_y}) = \frac{\zeta(f^k_{d_x,d_y})}{\sum_{k=0}^{n-1} \zeta(f^k_{d_x,d_y})} \tag{16}$$

We define *normalized flow reachability value* $\tilde{\eta}(f^k_{d_x,d_y})$ of a flow path $f^k_{d_x,d_y}$, as the *flow reachability value* of that flow path divided by summation of all *flow reachability values* of n flow paths.

$$\tilde{\eta}(f^k_{d_x,d_y}) = \frac{\eta(f^k_{d_x,d_y})}{\sum_{k=0}^{n-1} \eta(f^k_{d_x,d_y})} \tag{17}$$

We define *normalized flow value* $\tilde{\psi}(f^k_{d_x,d_y})$ of a flow path $f^k_{d_x,d_y}$, as the *flow value* of that flow path divided by summation of all *flow values* of n flow paths.

$$\tilde{\psi}(f^k_{d_x,d_y}) = \frac{\psi(f^k_{d_x,d_y})}{\sum_{k=0}^{n-1} \psi(f^k_{d_x,d_y})} \tag{18}$$

3.5. Algorithm to calculate Flow Value:

Algorithm 1 shows a procedure of calculating *flow values* of all possible paths between a source and a destination in a network. This is modified version of [25] to compute *flow assurance value*, *flow capacity value* and *flow reachability value*.

To discover all flow paths between two given sub-systems d_x and d_y, we start the search from the source sub-system d_x and observe all outgoing links. Then we proceed by exploring the first child sub-system until the destination sub-system is found or the algorithm finds other sink sub-systems. The process continues from the next recent child sub-system which was not explored in this method. The moment we find that the algorithm reaches the destination sub-system, we can realize that one of the flow paths is discovered which is denoted as $f^k_{d_x,d_y}$. Then we compute the global performance parameters *flow assurance value* $\lambda(f^k_{d_x,d_y})$, *flow capacity value* $\zeta(f^k_{d_x,d_y})$ and *flow reachability value* $\eta(f^k_{d_x,d_y})$.

Algorithm 1 Calculation of λ, ζ, η all Flow Paths

```
1: procedure CalcAllFlowValues(G,S,D)
2: % S:= source node and D:= destination node
3:     VisitedStack := List, VisitedStack := [S]
4:     StoreStack := List, StoreStack := [S]
5:     FinalList := List
6:     while StoreStack is not empty do
7:         currentNode := get the last entry of StoreStack;
8:         pull one node from outgoing nodes for currentNode,
9:         index := 0
10:        if node is Null then
11:            Delete the current index of StoreStack
12:            Delete the current index of VisitedStack
13:        else:
14:            if node == D then
15:                K := append node with VistedStack
16:                FinalList(index) = K
17:                index = index +1
18:                Compute the Flow Assurance Value for the FinalList
19:                Compute the Flow Capacity Value for the FinalList
20:                Compute the Flow Reachability for the FinalList
21:            else if node is not in the VisitedStack then
22:                append node in VisitedStack
23:                append node in StoreStack
24:            end if
25:        end if
26:    end while
27: end procedure
```

After all the possible paths between two given sub-systems d_x and d_y are discovered, we normalize the global performance parameters and estimate the weighted sum of *flow value* $\psi(f^k_{d_x,d_y})$ based on global performance parameters and their importance factor for assurance, capacity and reachability. Next based on this *flow value*, the ranks of the flow paths are obtained.

4. Analysis

In this section we analyze the proposed method of calculating flow values by giving an example and observing the effect of local performance metrics on global performance parameters.

4.1. An example - Calculation of Flow Value

In this section, we apply the proposed approach on an example network to analyze the network flow value. We consider a small network and explain how the flow value of each flow path is estimated. Then we also present the estimated rank for all possible flows between two sub-systems in the network.

As we can see from Figure 2, there are 9 sub-systems in the network. The *node assurance values* are mentioned inside the circle and the *link capacity values* are mentioned near the link. We can see that Nodes 7 and 8 are source nodes as $deg^-(7) = 0$ and $deg^-(8) = 0$. Node 1 and 2 are sink nodes as $deg^+(1) = 0$ and $deg^+(2) = 0$

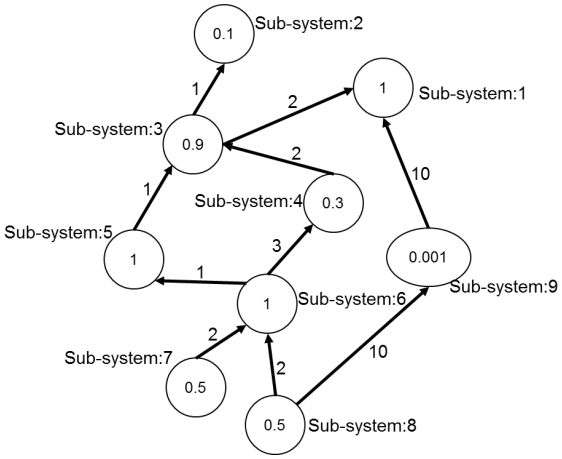

Figure 2. A Sample Network for Calculating Flow Value

In this example, we can see node assurance values for all sub-systems as below:

$$\alpha(1) = \alpha(5) = \alpha(6) = 1$$
$$\alpha(7) = \alpha(8) = 0.5$$
$$\alpha(2) = 0.1; \alpha(3) = 0.9; \alpha(4) = 0.3; \alpha(9) = 0.001$$

We can also get the link capacity value of all the links.

$$\beta(l_{7,6}) = \beta(l_{8,6}) = \beta(l_{4,3}) = \beta(l_{3,1}) = 2$$
$$\beta(l_{6,5}) = \beta(l_{5,3}) = \beta(l_{3,2}) = 1$$
$$\beta(l_{8,9}) = \beta(l_{9,1}) = 10; \beta(l_{6,4}) = 3$$

In this example, we would like to compute the route value for the flow path $f_{8,1}$ between *source sub-system* 8 and *sink sub-system* 1. Using the algorithm, we have found that there are three flows

$$f^1_{8,1} = \{l_{8,6}, l_{6,5}, l_{5,3}, l_{3,1}\}$$
$$f^2_{8,1} = \{l_{8,6}, l_{6,4}, l_{4,3}, l_{3,1}\}$$
$$f^3_{8,1} = \{l_{8,9}, l_{9,1}\}$$

From Table 1, we can see that in the first scenario, there is an equal importance factor for *flow assurance value*, *flow capacity value* and *flow distance value*.

$$\omega(\lambda) = \omega(\beta) = \omega(\zeta) = 1$$

Then out of three flow paths, $f^3_{8,1}$ is considered as the best path, and $f^1_{8,1}$ is the second best path and $f^2_{8,1}$ is the third best path. This is according to intuition that, the flow path $f^3_{8,1}$ has only 2 hops and flow capacity is 10, whereas the other two flow paths have flow capacity in the order of 1 and 2. Therefore, the flow path $f^3_{8,1}$ scores high though the flow assurance value is comparatively lower than the other two flow paths.

However, if we increase the importance factor for *flow assurance value*, then the ranking of the flow path changes.

$$\omega(\lambda) = 2; \omega(\beta) = \omega(\zeta) = 1$$

In this scenario, out of three flow paths, $f^1_{8,1}$ is considered as the best path, and $f^3_{8,1}$ is the second best path and $f^2_{8,1}$ is the third best path. This is also according to the intuition that in this scenario the rank of the flow path $f^3_{8,1}$ is low as the flow assurance value is much lower than the other flow path $f^1_{8,1}$.

4.2. Effect of local performance metrics on flow value computation:

In this section, we will study how the flow value $\psi(f_{d_x,d_y})$ between d_x and d_y changes based on the change in local performance metrics node assurance value $\alpha(d_i)$, link capacity $\beta(l_{d_i,d_j})$ and hop count $\gamma(f_{d_x,d_y})$. We have seen that *flow value* is defined as,

$$\psi(f_{d_x,d_y}) = \lambda(f_{d_x,d_y})\omega(\lambda) + \zeta(f_{d_x,d_y})\omega(\zeta) + \eta(f_{d_x,d_y})\omega(\eta)$$

The change in *flow value* $\psi(f_{d_x,d_y})$ with respect to *node assurance value* $\alpha(d_i)$, *link capacity* $\beta(l_{d_i,d_j})$ and *hop count* $\gamma(f_{d_x,d_y})$

$$\begin{cases} \frac{\partial(\psi(f_{d_x,d_y}))}{\partial(\alpha(d_i))} = \frac{\partial(\lambda(f_{d_x,d_y}))}{\partial(\alpha(d_i))} \\ \frac{\partial(\psi(f_{d_x,d_y}))}{\partial(\beta(l_{d_i,d_j}))} = \frac{\partial(\zeta(f_{d_x,d_y}))}{\partial(\beta(l_{d_i,d_j}))} \\ \frac{\partial(\psi(f_{d_x,d_y}))}{\partial(\gamma(f_{d_x,d_y}))} = \frac{\partial(\eta(f_{d_x,d_y}))}{\partial(\gamma(f_{d_x,d_y}))} \end{cases} \tag{19}$$

Now we demonstrate the change in *flow value* for a particular flow path f_{d_x,d_y} which has three sub-systems $\{d_x, d_m, d_y\}$ and two links $\{l_{d_x,d_m}, l_{d_m,d_y}\}$.

We know that, the *flow capacity value* $\zeta(f_{d_x,d_y})$ of flow f_{d_x,d_y} can be re-written as,

$$\zeta(f_{d_x,d_y}) = \min(\beta(l_{d_x,d_m}), \beta(l_{d_m,d_y}))$$
$$= \frac{\beta(l_{d_x,d_m}) - \beta(l_{d_m,d_y}) - \left|\beta(l_{d_x,d_m}) - \beta(l_{d_m,d_y})\right|}{2}$$

Therefore, the change in *flow assurance value*, *flow capacity value* and *flow reachability value* with respect to *node assurance value* $\alpha(d_m)$, *link capacity* $\beta(l_{d_x,d_m})$

Table 1. Calculation of flow value

Imp. Factor (ω)	k	$f^k_{8,1}$	L	$\tilde{\lambda}(f^k_{8,1})$	$\tilde{\zeta}(f^k_{8,1})$	$\tilde{\eta}(f^k_{8,1})$	$\tilde{\psi}(f^k_{8,1})$
$\omega(\lambda)=1$ $\omega(\zeta)=1$ $\omega(\eta)=1$	1	$f^1_{8,1}$	$\{l_{8,6}, l_{6,5}, l_{5,3}, l_{3,1}\}$	0.77	0.077	0.25	0.37
	2	$f^2_{8,1}$	$\{l_{8,6}, l_{6,4}, l_{4,3}, l_{3,1}\}$	0.23	0.15	0.25	0.21
	3	$f^3_{8,1}$	$\{l_{8,9}, l_{9,1}\}$	9.0×10^{-2}	0.77	0.5	**0.42**
$\omega(\lambda)=2$ $\omega(\zeta)=1$ $\omega(\eta)=1$	1	$f^1_{8,1}$	$\{l_{8,6}, l_{6,5}, l_{5,3}, l_{3,1}\}$	0.77	0.077	0.25	**0.47**
	2	$f^2_{8,1}$	$\{l_{8,6}, l_{6,4}, l_{4,3}, l_{3,1}\}$	0.23	0.15	0.25	0.22
	3	$f^3_{8,1}$	$\{l_{8,9}, l_{9,1}\}$	9.0×10^{-2}	0.77	0.5	0.32

and *hop count* $\gamma(f_{d_x,d_y})$ is given as,

$$\begin{cases} \frac{\partial(\lambda(f_{d_x,d_y}))}{\partial(\alpha(d_m))} = \alpha(d_x)\alpha(d_y) \\ \frac{\partial(\zeta(f_{d_x,d_y}))}{\partial(\beta(l_{d_x,d_m}))} = \frac{\frac{\beta(l_{d_m,d_y})-\beta(l_{d_x,d_m})}{\left|\beta(l_{d_m,d_y})-\beta(l_{d_x,d_m})\right|}+1}{2} \\ \frac{\partial(\eta(f_{d_x,d_y}))}{\partial(\gamma(f_{d_x,d_y}))} = \frac{1}{[\gamma(f_{d_x,d_y})]^2} \end{cases} \tag{20}$$

Figure 3 shows the graphical presentation of change in flow value of flow path $f^2_{8,1}$ with change in node assurance value. Figure 4 shows the graphical presentation of change in flow value of flow path $f^2_{8,1}$ with change in link capacity and Figure 5 shows the graphical presentation of change in flow value of flow path $f^2_{8,1}$ with change in hop count.

In the first case, we gradually increase the *node assurance value* of Node 3 from 0.1 to 1. In the second case we gradually increase the *link capacity* of $l_{6,4}$ from 1 to 10 and in the third case, we linearly increase the *hop count* of the flow path $f^2_{8,1}$.

We can see from the graph that with an increase of *node assurance value* the *flow value* gradually increases until the *node assurance value* reaches 1. With the increase of *link capacity*, the *flow value* increases until it reaches the minimum of the rest of *link capacity* set in the flow path. Once the *link capacity* reaches the minimum of the set, there is no change in the flow value. The increase of *hop count* decreases the *flow reachability* and in turn decreases the *flow value*.

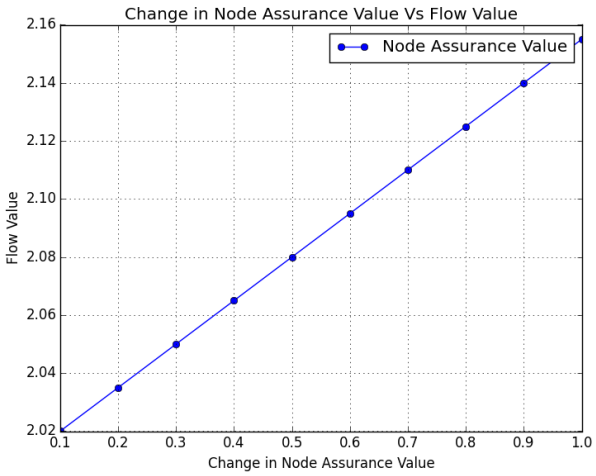

Figure 3. Change in Flow Value with change in Node Assurance Value

4.3. Observations

We have implemented the graph traversal algorithm with different routing metrics and simulated some network scenarios for different directed acyclic graphs. From the study, we will present propositions about the final flow value computation framework and give the proof of their correctness.

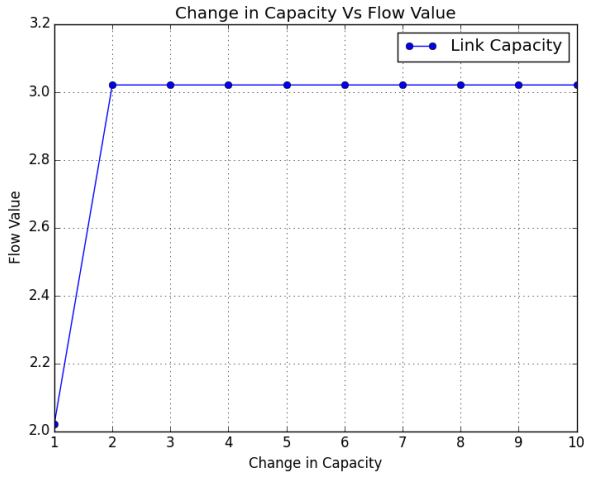

Figure 4. Change in Flow Value with change in Link Capacity Value

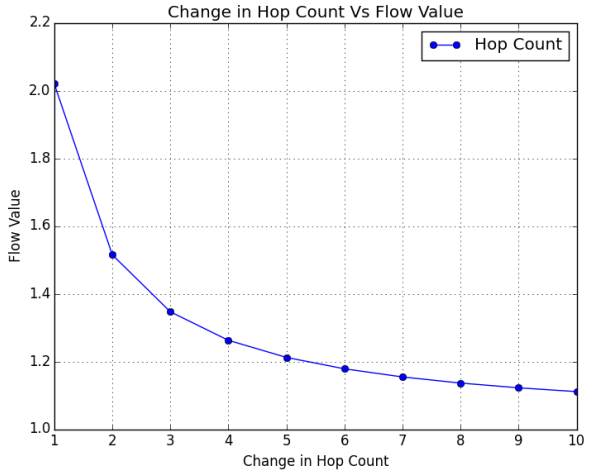

Figure 5. Change in Flow Value with change in Hop Count

Proposition 1: *The final flow value increases with the increase of the node assurance value of each sub-system.*

Proof: let us assume that a flow path f_{d_x,d_y} has three sub-systems $\{d_x, d_m, d_y\}$ and two links $\{l_{d_x,d_m}, l_{d_m,d_y}\}$. From equation 20, we can see that the change in a node assurance value $\alpha(d_m)$ will result in a change in the final flow value with a factor $\alpha(d_m)\alpha(d_x)\alpha(d_y)$. Therefore, with the increase of the *node assurance value* the final *flow value* will increase steadily.

Proposition 2: *The final flow value depends on the capacity of a link.*

Proof: Let us assume that a flow path f_{d_x,d_y} has three sub-systems $\{d_x, d_m, d_y\}$ and two links $\{l_{d_x,d_m}, l_{d_m,d_y}\}$. The link capacity value $\beta(l_{d_x,d_m}) > \beta(l_{d_m,d_y})$. This implies that the $\beta(l_{d_x,d_m})$ was the minimum link bandwidth and $\beta(l_{d_m,d_y})$ is the next minimum bandwidth. From equation 20, we can see that the change in a link capacity value $\beta(l_{d_x,d_m})$ results in a change in the final flow value with a factor $\beta(l_{d_x,d_m})0.5[\frac{\beta(l_{d_m,d_y})-\beta(l_{d_x,d_m})}{|\beta(l_{d_m,d_y})-\beta(l_{d_x,d_m})|}+1]$. This implies that the final flow value will grow linearly with the increase of $\beta(l_{d_x,d_m})$ until $\beta(l_{d_x,d_m})$ becomes equal to $\beta(l_{d_m,d_y})$. Once $\beta(l_{d_x,d_m})$ becomes equal to $\beta(l_{d_m,d_y})$, then with further increase of $\beta(l_{d_x,d_m})$ there will be no change in the final flow value.

Proposition 3: *The final flow value decreases with an increase of hop count in a path.*

Proof: From equation 20, we can see that the change in a hop count $|f_{d_x,d_y}|$ results in decrease in the final flow value with a factor $\frac{1}{[\gamma(f_{d_x,d_y})]^2}$. This implies that the final flow value will decrease gradually with the increase of $|f_{d_x,d_y}|$, as $\eta(f_{d_x,d_y}) = \frac{1}{|f_{d_x,d_y}|}$

We have observed that if we need to improve the *flow value*, we can use more trustworthy nodes in the communication path. The effect of hop count on *flow value* is straight forward. The increase of hop count will result in an increase in delay and in turn reducing the reachability. Whereas, the *flow value* depends on the minimum link bandwidth in a path. Hence, if a flow path has a link with very low bandwidth, then improving the other parts of the path with higher bandwidth will not improve the *flow value*. This implies that, if a flow path has a secure path with low link bandwidth and higher number of hops compared to a flow path with a lower number of hops and higher bandwidth, the rank of the flow path with the high security path will score less. Similarly, if we have a high number of intermediate nodes with low capacity and high trustworthiness between the source node and destination node, we might not get a high rank flow path. If there is a bottleneck with a low capacity link in the network, the increase of trustworthiness of nodes will not improve the flow path value. This type of assessment can help plant operators to decide on the network design for plant commissioning.

5. Conclusions and Future Work

This paper introduces a concept to balance secure, high capacity and reachable flow paths in a heterogeneous industrial network during planning phase. Using this method, we study the effects of local performance indicators of each node on global performance indicators of a flow path keeping overall security and robustness in the system. This model can assist plant operators to rank each communication flow path based on security, capacity and reachability. This model also aids to compute an initial realistic guess for different alternate options of network paths. We have observed that if there is a bottleneck with a low capacity link in the network, the increase of trustworthiness of sub-systems will not improve the flow path value. Similarly, if we

have a high number of intermediate sub-systems with low capacity and high security between the source sub-system and destination sub-system, we might not get a high rank flow path. This type of information is useful when designing a plant with a service level agreement. We demonstrate the characteristics of our model through an analytical example and simulation results.

In this work, we mainly focus to identify the optimized network flow path between source and destination in a heterogeneous network. This modeling is done during the network formation stage and is not aware of the run-time information. In this model, we consider the link capacity between two sub-systems as a local performance metric. This is the theoretical upper bound on the rate at which messages can be reliably transmitted. In this network model, we do not consider the throughput of the system which can be an average rate of successful message delivery over a communication link. This throughput can only be available to the network operator during run-time when the message sending rate is also available along with the fixed topology. Furthermore, the information used in our model as local performance metrics are static. It would be interesting to study the network dynamics considering the dynamic message transactions between the nodes in the network. Therefore, we need to analyze the working flow paths rather than all possible flow paths. Then we can validate the performance of a network after choosing the identified flow path as described in our model. We plan to explore this option in our next work. We also plan to explore the applicability of this solution in an industrial use case where security, network capacity and reachability needs to be optimal for successful network operation. Another interesting area to explore is the scenario with bi-directional network links. This is challenging to incorporate as both the nodes have the possibility to affect each other, therefore finding trust values for these situations requires further research.

Acknowledgments

This work has been supported by the Swedish Knowledge Foundation (KKS) through ITS-EASY, Embedded Software and Systems Industrial Research School, affiliated with the School of Innovation, Design and Engineering (IDT) at Mälardalen University (MDH, Västerås, Sweden) as well as by the ABB Communication Research Area.

References

[1] V. Gungor, D. Sahin, T. Kocak, S. Ergut, C. Buccella, C. Cecati, and G. Hancke, Smart Grid Technologies: Communication Technologies and Standards, Industrial Informatics, IEEE Transactions on, vol. 7, no. 4, pp. 529539, Nov 2011.

[2] M. Levorato and U. Mitra, Optimal allocation of heterogeneous smart grid traffic to heterogeneous networks, in Smart Grid Communications (SmartGridComm), 2011 IEEE International Conference on, Oct 2011, pp. 132137.

[3] L. Zhou and Z. J. Haas, Securing ad hoc networks, IEEE Network, vol. 13, no. 6, pp. 2430, Nov 1999.

[4] A. Josang and S. J. Knapskog, A metric for trusted systems, Proceedings 21st NIST-NCSC National Information Systems Security Conference, pp. 1629, 1998.

[5] S. Capkun, L. Buttyan, and J. P. Hubaux, Self-organized public-key management for mobile ad hoc networks, IEEE Transactions on Mobile Computing, vol. 2, no. 1, pp. 5264, Jan 2003.

[6] J. H. Cho, K. S. Chan, and I. R. Chen, Composite trust-based public key management in mobile ad hoc networks, in Proceedings of the 28th Annual ACM Symposium on Applied Computing, ser. SAC 13. New York, NY, USA: ACM, 2013, pp. 19491956.

[7] G. Xie, J. Zhan, D. Maltz, H. Zhang, A. Greenberg, G. Hjalmtysson, and J. Rexford: On static reachability analysis of ip networks : in INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE, March 2005, pp. 21702183, vol. 3.

[8] P. Matousek, J. Rab, O. Rysavy, and M. Sveda: A formal model for network-wide security analysis: in Engineering of Computer Based Systems, 2008. ECBS 2008. 15th Annual IEEE International Conference and Workshop on the, March 2008, pp. 171181.

[9] A. Datta: Security analysis of network protocols: Compositional reasoning and complexity-theoretic foundations: in PhD Thesis, 2005.

[10] A. Sharma, Z. Kalbarczyk, J. Barlow, and R. Iyer: Analysis of security data from a large computing organization: in Dependable Systems Networks (DSN), 2011 IEEE/IFIP 41st International Conference on, June 2011, pp. 506517.

[11] J. Mirkovic, T. V. Benzel, T. Faber, R. Braden, J. T. Wroclawski, and S. Schwab: The deter project: Advancing the science of cyber security experimentation and test: in Technologies for Homeland Security (HST), 2010 IEEE International Conference on, 2010, p. 7

[12] D. Moore, C. Shannon, D. J. Brown, G. M. Voelker, and S. Savage: Inferring internet denial-of-service activity: ACM Trans. Comput. Syst., vol. 24, no. 2, pp. 115139, May 2006

[13] O. Sheyner, J. Haines, S. Jha, and R. Wing: "Automated generation and analysis of attack graphs": in Technical Report CMU, 2002

[14] A. Xie, Z. Cai, C. Tang, J. Hu, and Z. Chen: "Evaluating network security with two-layer attack graphs": In Proc. of annual computer security applications conference (ACSAC 2009) pp. 127 -136

[15] J. Hong and D.-S. Kim: Harms: Hierarchical attack representation models for network security analysis: in Proceedings of the 10th Australian Information Security Management Conference, Novotel Langley Hotel, Perth, Western Australia Dec 2012.

[16] V. Mehta, C. Bartzis, H. Zhu, E. Clarke, and J. Wing: "Ranking Attack Graphs": in Recent Advances in Intrusion Detection, Lecture Notes in Computer Science, Vol. 4219, pp. 127-144, 2006

[17] B. Genge, C. Siaterlis, I.N. Fovino, and M. Masera:, "A cyber-physical experimentation environment for the security analysis of networked industrial control systems.",: Computers and Electrical Engineering38 (5) , 1146-1161, 2012

[18] T. Chee-Wooi, L. Chen-Ching, and G. Manimaran: "Vulnerability Assessment of Cybersecurity for SCADA Systems," Power Systems, IEEE Transactions on, vol. 23, pp. 1836-1846, 2008.

[19] N. Papakonstantinou, S. Sierla, K. Charitoudi, B. O'Halloran, T. Karhela, V. Vyatkin, and I. Turner: "Security impact assessment of industrial automation systems using genetic algorithm and simulation": Emerging Technology and Factory Automation (ETFA), 2014 IEEE , vol., no., pp.1,8, 16-19 Sept. 2014

[20] C. Alcaraz, G. Fernandez, and F. Carvajal: "Security aspects of SCADA and DCS environments": 7130. Jan. 1, 2012. 120-149 p. 7130. (Lecture Notes in Computer Science). ISBN: 9783642289194.

[21] I. Kotenko and E. Doynikova: "Evaluation of Computer Network Security based on Attack Graphs and Security Event Processing": in Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), pp-14–29, vol. 5, No. 3, Sep 2014

[22] S. I. Ahamed, M. M. Haque, M. E. Hoque, F. Rahman, and N. Talukder, Design, analysis, and deployment of omnipresent formal trust model (ftm) with trust bootstrapping for pervasive environments, Journal of Systems and Software, vol. 83, no. 2, pp. 253270, Feb. 2010.

[23] S. Adal, K. Chan, and J. Cho, TANDEM: a trust-based agent framework for networked decision making, Computational & Mathematical Organization Theory, vol. 21, no. 4, pp. 461490, 2015.

[24] A. Ray, J. Åkerberg, M. Björkman, and M. Gidlund, Towards Security Assurance for Heterogeneous Industrial Networks, in 41st Annual Conference of the IEEE Industrial Electronics Society (IECON 2015), November 2015.

[25] R. Sedgewick, Algorithms in C, Part 5: Graph Algorithms, Addison Wesley Professional, 3rd ed., 2001