53rd IEEE Conference on Decision and Control
December 15-17, 2014. Los Angeles, California, USA

Vulnerability Analysis for Distributed and Coordinated Destabilization Attacks[†]

V. Chetty*[1], N. Woodbury[1], E. Vaziripour[2], and S. Warnick[1]

[1]Information and Decision Algorithms Laboratories, Brigham Young University
[2]Internet Research Lab, Brigham Young University

Abstract— **This paper focuses on how the vulnerability of an LTI system to destabilizing attacks can be posed as its robustness to external disturbances. First, we extend existing work on single link attack models to a more generalized attack model that allows for multiple link attacks. This is done by extending the partial structure representation of dynamical structure functions to include external perturbations. Given the new model, we then discuss how to determine the vulnerability of the system for both coordinated and distributed destabilizing attacks on a system. Finally, we develop a separability result for vulnerability in feedback systems that will be useful in determining secure architectures for structured controller design.**

I. Introduction

A variety of systems–such as power grids, water supplies, and transportation networks–can be modeled by a network of physical and cyber components. Many of these infrastructures are critical to maintaining public health and safety and it is necessary that these systems are resilient to both malicious attacks and accidental outages. Measuring the vulnerability of these systems to an attack or an unplanned outage will help to design a more robust network infrastructure.

In general, malicious attacks are usually directed towards the most sensitive parts of a system in order to increase the amount of damage they can cause with minimum effort. Attacks can take many shapes and forms, such as physical or cyber attacks against a system. Certain attacks or unplanned outages can cause cascading failures across numerous systems [1]. For example, a software bug caused what is now known as the North-east blackout of 2003. This bug blacked out several large regions of the United States and Canada for up to two days. Backup generators failed, water pressure in several cities fell, phone systems became non-operational, and major water networks were contaminated as a result [2].

The recent economic crisis has also been compared to a blackout in a power grid. The comparisons claim that a small change within the global economic network–specifically a rise in the default rates on mortgages–led to a cascading failure throughout the entire economic network much like a downed power line can lead to cascading failure within a power grid [3].

Each of these examples demonstrate how a local disturbance can cause global cascading failure throughout a networked system. Networked systems should be designed to be robust to such disturbances, especially if the disturbances could be malicious attacks. Designing systems to be robust requires models of both the system being attacked as well as the method of the attack itself. We now explore a framework of attack models which will be useful in discussing the vulnerability of LTI systems in particular.

A. Attack Models and Scenarios

The work in [4] introduces a framework by which attack scenarios may be classified. The elements of an attack may include:

- *Attack Goals:* The purpose or intent of the attack which can be stated as the impact of a successful attack on a system. This purpose may be to steal information, change the state of the system, destabilize the system, etc.
- *Attack Policy:* The mechanism by which the attacker executes the attack. This policy is characterized as a point in a three dimensional attack space, where the dimensions are defined as follows:
  - *System Knowledge:* The scope of knowledge about the system available to the attacker before the attack begins.
  - *Disclosure Resources:* The set of states that can be measured by the attacker during the attack.
  - *Disruption Resources:* The set of states that can be modified or controlled by the attacker during the attack.
- *Additional Constraints:* Any other constraints on the attack. For example, one could consider a stealthiness constraint on many attacks which further limits the amount of resources available to the attacker and the size of the attack is unobservable to the system.

Equipped with this framework, we can categorize several types of well-known attacks.

B. Deception Attacks

Deception attacks define a class of attack models that have recently received wide attention. Within the framework

[†]This material is based on research sponsored by the Department of Homeland Security (DHS) Science and Technology Directorate, Homeland Security Advanced Research Projects Agency (HSARPA), Cyber Security Division (DHS S&T/HSARPA/CSD), LRBAA 12-07 via contract number HSHQDC-13-C-B0052. All material in this paper represents the position of the authors and not necessarily that of DHS.

* Corresponding author, chettyv@byu.edu.

outlined above, deception attacks can be classified as attacks which are executed with the intent to provide false information to authorized sources [5].

C. Denial of Service Attacks

Denial of service attacks are characterized as attacks where disruption resources available to the attacker include the ability to disrupt communication between states in the system. These types of attacks may serve several purposes such as the destabilization of the system or a degradation the system's performance [6].

D. Destabilization Attacks

The primary focus of this work is on a system's vulnerability to destabilization attacks. The purpose of destabilization attacks is to create local disturbances on a network system in order to cause global cascading failures[7].

One specific attack that has recently received wide attention is Stuxnet, in large part due to the fact that, unlike many of the attacks and attack models studied in the past, the purpose of Stuxnet was not to steal, manipulate, or erase information; rather, it was intended to destroy physical systems, which it achieves by changing the input seen by the controllers on these physical systems [8]. Since Stuxnet causes global failure to these systems by making small changes to the controllers on these systems, it can be modeled as a destabilization attack.

For the purposes of this paper, destabilization attacks do not involve stealthiness constraints; however they will assert that the system knowledge as well as the disclosure resources available to the attacker are constrained to manifest or exposed states. Further, we only consider destabilization attacks on causal, linear time invariant (LTI) systems.

II. Background

We begin by discussing the methodology used throughout the paper for calculating the vulnerability of a link to a given attack model. The main results of this paper are developed using the dynamical structure function (DSF) representation, which is a partial structure representation of the system. We then motivate the use of the DSF representation for vulnerability analysis of destabilizing attacks and discuss previous work already conducted in this area on single link destabilization attacks. Note that a link in the dynamical structure function represents a causal relationship between either two measured states or an input and a measured state in the state space representation of the system, possibly through hidden states.

A. Methodology for Vulnerability Analysis

Vulnerability analysis begins by deciding which system variables are potentially exposed to attack. The DSF, described in detail in Section II-B, is then used to model the system in terms of these exposed variables.

An attack model specifies which exposed variables the attacker can measure and which they can affect. The system vulnerability with respect to this attack is calculated via the small gain theorem to determine how the attacker has to work to destabilize the system from its current position. The small gain theorem is applied by determining the transfer function matrix M that is in feedback with the attack Δ (see Figure 1).

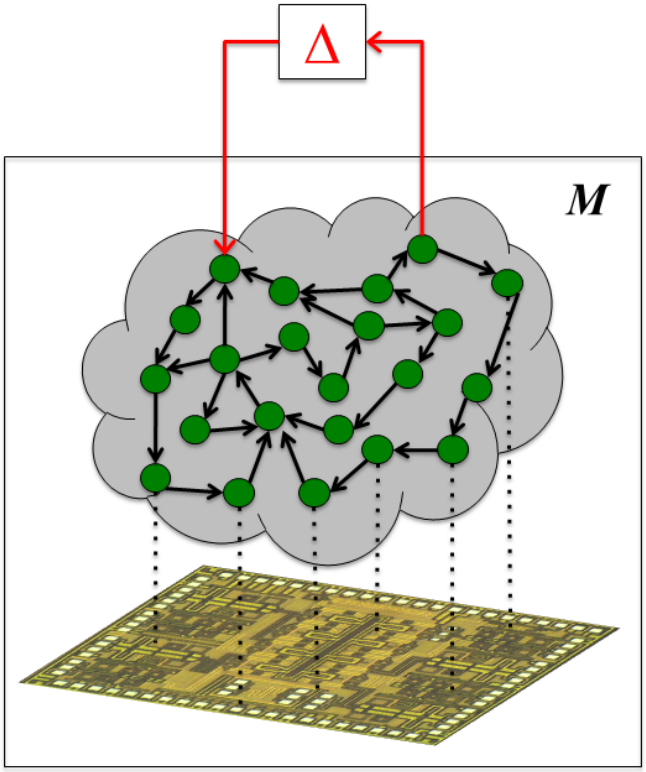

Fig. 1: Modeling attacks as a destabilizing perturbation to a system model built around exposed variables enables standard robustness analysis to define a meaningful notion of vulnerability. The figure shows a single-link attack, resulting in a Single Input Single Output Δ block, but the methodology is similar if the attacker has access to more exposed variables, making Δ Multi Input Multi Output (in general). The structure of Δ then characterizes decentralized attacks (diagonal structure) from coordinated attacks (full-block structure).

B. Dynamical Structure Functions

The DSF representation was first developed in [9] and represents the relationship among exposed variables in a system. Given a state space representation of the form:

$$\begin{aligned} \begin{bmatrix} \dot{y} \\ \dot{z} \end{bmatrix} &= \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} u \\ y &= \begin{bmatrix} I & 0 \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} \end{aligned}$$

where we note that $C = \begin{bmatrix} I & 0 \end{bmatrix}$ because we partition the states into the exposed states, y, and hidden states, z.

The associated DSF (Q, P), as shown in [9], is:

$$Y = QY + PU \qquad (1)$$

where Y and U are the Laplace transforms of y and u. Also, Q is a hollow matrix that describes how exposed states affect other exposed states, while P is a matrix that describes how inputs affect exposed states.

The relationship given in (1) details the structure between the exposed variables, which is the structure of the system that is most likely to be visible to an external attacker and highlights which variables an attacker can affect in the system. Therefore, the DSF is a useful representation for calculating the vulnerability of links to destabilization attacks.

C. Single Link Attack

From [7] we learn that a stable additive perturbation Δ on a link Q_{ij} or P_{ij} is able to destabilize the system if and only if the transfer function, M_{ij}, seen by Δ is nonzero. This means that the link Q_{ij} or P_{ij} is in feedback with some series of links in Q or P. Note that M_{ij} is the transfer function from ΔY_j to Y_j. From this result we can conclude that, for open-loop systems, links in the matrix P are never vulnerable to this type of an attack, so only links in Q are potentially vulnerable.

The vulnerability of a single link in Q is then defined to be the inverse of the smallest perturbation required to destabilize the system. So a large perturbation means the vulnerability is low and vice versa. By application of the small gain theorem, we note that the system will remain stable as long as $||\Delta||_\infty ||M_{ij}||_\infty < 1$.

Therefore, the smallest perturbation necessary to destabilize the system occurs when

$$||\Delta||_\infty ||M_{ij}||_\infty = 1$$

which means

$$||\Delta||_\infty = \frac{1}{||M_{ij}||_\infty}$$

Thus, the vulnerability of a link Q_{ij} is

$$v_{ij} = ||M_{ij}||_\infty$$

The overall vulnerability, V, of the entire system to single link attacks can then be defined as the largest vulnerability of any single link, looking across all links in the system; i.e.

$$V = \max_{Q_{ij} \neq 0 \in Q} v_{ij}$$

III. State Space Attack Model

The analysis of the vulnerability of a single link to a destabilizing attack can then be extended to a more general model incorporating simultaneous attacks on multiple links. We begin by formulating this model using the state space representation of LTI systems.

We model an attack on a network system as an external disturbance $F\psi$ to the system to get:

$$\begin{array}{rcl} \dot{x} & = & Ax + Bu + F\psi \\ y & = & \begin{bmatrix} I & 0 \end{bmatrix} x \end{array} \tag{2}$$

As noted in Section I, the purpose of the attack is to destabilize the system using a well-defined set of disclosure and disruption resources. Restricting ourselves to the class of stable systems we can see that a bounded input will create a bounded output, so the system can not be destabilized by a stable external disturbance, so we restrict ourselves to the case where $\psi = x$.

With $\psi = x$ an attacker is allowed to artificially create a link connecting state x_i to state x_j; moreover, if $F = B$ we can artificially create a link from a state x_i to an input u_j. Allowing the attacker to create links that have feedback with existing links has negative implications for the overall vulnerability of the system. This is discussed in more detail in Section IV-A.

For many attack models, a reasonable restriction may be that the attacker is only allowed to use the existing system infrastructure. Therefore, we can restrict this attack model such that if $a_{ij} = 0$, then $\delta_{ij} = 0$. We can restrict the attacker further to only allow attacks on exposed states. For this restricted class of attacks, a more useful representation of a system is the DSF. Therefore, our next step is to formulate a generalized attack model in the DSF framework before adding these restrictions.

IV. A Generalized Attack Model using the Dynamical Structure Function Representation

In order to define a generalized attack model in the DSF domain, we first need to define the DSF of a system when external disturbances are present. We rewrite (2) to be a state space system of the form

$$\begin{array}{rcl} \begin{bmatrix} \dot{y} \\ \dot{z} \end{bmatrix} & = & \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} u + \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} \psi \\ y & = & \begin{bmatrix} I & 0 \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} \end{array} \tag{3}$$

Taking the Laplace transform of (3) yields

$$\begin{bmatrix} sY \\ sZ \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} Y \\ Z \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} U + \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} \Psi \tag{4}$$

where Z and Ψ are the Laplace transforms of z and ψ, respectively.

Solving for Z in the second equation of (4) gives us

$$\begin{aligned} Z = (sI - A_{22})^{-1} A_{21} Y + \\ (sI - A_{22})^{-1} B_2 U + (sI - A_{22})^{-1} F_2 \Psi \end{aligned} \tag{5}$$

Plugging (5) into the first equation in (4) then gives us

$$sY = WY + VU + N\Psi \tag{6}$$

with $W = A_{11} + A_{12}(sI - A_{22})^{-1} A_{21}$, $V = B_1 + A_{12}(sI - A_{22})^{-1} B_2$, $N = F_1 + A_{12}(sI - A_{22})^{-1} F_2$

Finally, defining $D = diag(W_{11}, ..., W_{pp})$ subtracting DY from both sides of (6) we get

$$Y = QY + PU + \Delta\Psi \tag{7}$$

where $Q = (sI - D)^{-1}(W - D)$, $P = (sI - D)^{-1} V$, and $\Delta = (sI - D)^{-1} N$.

Equation (7) is then a generalized attack model in the DSF domain.

A. Calculating Vulnerability from a Generalized Attack Model

The class of attacks we focus on in this paper is the class of destabilizing attacks, which in turn can be split into two classes of attacks: the first in which the attacker can create communication links in a network and the second in which the attacker must use the existing communication links.

If we allow an attacker to create links in a system then it is possible for an attacker to create a feedback loop in any system. And as shown in Section IV-B, this feedback loop will create a vulnerability within the system. Therefore, under this assumption, no completely secure architecture can exist. For this reason and from this point forward, we only consider attack models that use the existing communication structure to conduct an attack. This is not an unreasonable assumption since creating new links within a system may be a difficult or expensive task for an attacker.

1) Vulnerability of a Single Link Attack: Starting with (7) and solving for Y in terms of U and Ψ we get

$$Y = (I-Q)^{-1}PU + (I-Q)^{-1}\Delta\Psi \tag{8}$$

where the input-output relationship is given by $G = (I - Q)^{-1}P$ and the transfer function describing how Ψ affects the exposed states, Y, is $(I - Q)^{-1}\Delta$. Given that we are only considering stable systems, we know that no bounded input can destabilize the system. Therefore, we consider the case when $\Psi = Y$, which means that an attacker is using some combination of additive perturbations on existing links to destabilize the system.

The transfer function seen by a perturbation ΔY is then given in (8) as $(I - Q)^{-1}$. In particular, if we want to determine the vulnerability of a single link attack on a link Q_{ij}, we know this can be modeled as $\Delta_{ij} = (sI-D_{ii})^{-1}N_{ij}$ with the rest of the entries in Δ equal to zero. Then, the transfer function seen by the perturbation on the link Q_{ij} is found from

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_{j-1} \\ Y_j \\ Y_{j+1} \\ \vdots \\ Y_p \end{bmatrix} = H \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \Delta_{ij}Y_j \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{9}$$

where $H = (I - Q)^{-1}$. From (9), we can see that $Y_j = H_{ji}\Delta_{ij}Y_j$ since $\Delta_{ij}Y_j$ is in the i^{th} row of the vector given in (9). Therefore,

$$v_{ij} = ||H_{ji}||_\infty$$

which means that

$$V = \max_{Q_{ij}\neq 0 \in Q} ||H_{ji}||_\infty \tag{10}$$

2) Vulnerability of a Multiple Link Distributed Attack: We now consider an attack in which multiple attackers are simultaneously performing unique single link attacks in the system and are not sharing information. This is modeled by the concatenation of several single link attacks on the system and by application of the small gain theorem, the vulnerability, $v_{ij,...,kl}$ of this type of an attack is the structured singular value, $\mu_{ij,...,kl}$, of the matrix

$$R_{ij,...,kl} = \begin{bmatrix} H_{ji} & 0 & ... & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & ... & 0 & H_{lk} \end{bmatrix} \tag{11}$$

That is,

$$v_{ij,...,kl} = \mu(R_{ij,...,kl}, \Pi)$$

The overall vulnerability of the system of a distributed attack is

$$V = \max_{R_{links}\in\mathscr{R}} \mu_{links}$$

where $\mathscr{R}$ is the set of matrices of the form (11) over the set of all possible combinations of links, $\mathscr{L}$, and μ_{links} is the structured singular value of R_{links}.

3) Vulnerability of a Multiple Link Co-ordinated Attack: A multiple link coordinated attack is a generalizaation of a single link attack and is similar to a distributed attack, except that it models either communication between multiple attackers or a single attacker targeting multiple links. The transfer function seen by a perturbation on multiple links when allowing for communication in the attack is then given by

$$T_{ij,...,kl} = \begin{bmatrix} H_{ji} & ... & H_{li} \\ \vdots & \ddots & \vdots \\ H_{jk} & ... & H_{lk} \end{bmatrix} \tag{12}$$

In this case, we find the vulnerability to be

$$v_{ij,...,kl} = ||T_{ij,...,kl}||_\infty$$

Thus the overall vulnerability of the system of a co-ordinated attack is then given by

$$V = \max_{links\in\mathscr{L}} ||T_{links}||_\infty$$

B. Reducing Vulnerability in Open-Loop Systems

Since the vulnerability of any given link in a system is the transfer function seen by a perturbation on that link the vulnerability of the system is nonzero if and only if feedback exists within the system. Therefore, one completely secure architecture is one in which no links in Q exist.

Note that since $G = (I - Q)^{-1}P$, when $Q = 0$, then $P = G$. Since links in P are never in feedback for open-loop systems in which attackers cannot create links, they are never vulnerable. Thus the overall vulnerability of a system with $Q = 0$ is $V = 0$, meaning there does not exist a finite additive perturbation on a link in the system that can destabilize the system under the assumption that the attacker can only use the existing communication network of the system [7].

However, if an attacker is allowed to create arbitrary links within the system, any system with at least one link will be vulnerable since the attacker can create a link in feedback with the existing link.

V. Feedback Systems

Although completely secure systems can be created by removing all feedback from the system, in many situations, feedback is a necessary component of the design of the system and cannot be removed. For example, an unstable system can be stabilized through a controller connected in feedback.

Throughout this section, we will refer to feedback systems as an autonomous system containing a plant with behavior G such that $Y = GU$ and a controller with behavior K such that $U = KY$, and where $G \neq K^{-1}$. Note that the plant and the controller may each be stable or unstable; however, we only consider the cases where the combined feedback systems are BIBO stable.

A. Feedback Systems as a Dynamical Structure Functions

For a feedback system, we consider an attack Ψ which may attack either the plant or the controller. Again, we are only considering stable systems and since no bounded input can destabilize a stable system, we only consider the case where $\Psi = \begin{bmatrix} Y & U \end{bmatrix}^T$.

Let the structure of G be given by (P_g, Q_g) and consider a general attack $\Delta_g \Psi = \begin{bmatrix} \Delta_{11} & \Delta_{12} \end{bmatrix} \Psi$ on the plant. Then, from (7), we get

$$Y = Q_g Y + P_g U + \begin{bmatrix} \Delta_{11} & \Delta_{12} \end{bmatrix} \Psi.$$

Similarly, let the structure of K be given by (P_k, Q_k) and consider a general attack $\Delta_k \Psi = \begin{bmatrix} \Delta_{21} & \Delta_{22} \end{bmatrix} \Psi$ on K. Then

$$U = Q_k U + P_k Y + \begin{bmatrix} \Delta_{21} & \Delta_{22} \end{bmatrix} \Psi.$$

Combining these equations, we get

$$\begin{aligned} \begin{bmatrix} U \\ Y \end{bmatrix} &= \begin{bmatrix} Q_k & P_k \\ P_g & Q_g \end{bmatrix} \begin{bmatrix} U \\ Y \end{bmatrix} + \begin{bmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{bmatrix} \Psi \\ &\triangleq Q_c \begin{bmatrix} U \\ Y \end{bmatrix} + \Delta_c \Psi \end{aligned} \tag{13}$$

The combined system in (13) is a DSF; therefore we can perform a vulnerability analysis on Q_c in the same manner as before. Note that the links in P_k and P_g are now in the Q of the combined system; therefore, it is possible for these links to have a nonzero vulnerability. We can also define Δ_{11} as an attack on Q_k, Δ_{12} as an attack on P_k, and so forth.

B. Separability in Feedback Systems

One property of vulnerability within feedback systems is the notion of separability. We define separability as follows:

Definition 1. *Consider all potential multiple link coordinated attacks on a controller (of which single link attacks are a subset). The vulnerability analysis on attacks in the controller are considered* separable *from the plant if the computation of the vulnerability of any and all attacks in the controller depends only on the behavior G of the plant and not on the structure (P_g, Q_g) of the plant.*

Note that we can make a similar definition for the separability of vulnerability of the links in the plant from the controller. Equipped with this definition, we can then show the conditions under which multiple link coordinated attacks allow for separability.

Theorem 1. *The vulnerability analysis of the controller resulting from any arbitrary multiple link coordinated attack is separable from the plant if and only if the attack does not include an attack on the plant. In other words, the vulnerability analysis of the controller is separable from the plant if and only if Δ_{21} and Δ_{22} are zero.*

Proof. Let Q_k and Δ_{11} be size $p \times p$. Also let P_k, K, and Δ_{12} be size $p \times m$. Consider also $H_c = (I - Q_c)^{-1} = \begin{bmatrix} \Gamma_1 & \Gamma_2 \\ \Gamma_3 & \Gamma_4 \end{bmatrix}$. Then we have

$$\begin{bmatrix} (I - Q_k) & -P_k \\ -P_g & (I - Q_g) \end{bmatrix} \begin{bmatrix} \Gamma_1 & \Gamma_2 \\ \Gamma_3 & \Gamma_4 \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}.$$

Thus

$$\begin{aligned} (I - Q_g)\Gamma_3 - P_g \Gamma_1 &= 0, \\ \Gamma_3 &= (I - Q_g)^{-1} P_g \Gamma_1 = G\Gamma_1, \\ (I - Q_k)\Gamma_1 - P_k \Gamma_3 &= I, \\ (I - Q_k)\Gamma_1 - P_k G \Gamma_1 &= I, \\ \Gamma_1 &= (I - Q_k - P_k G)^{-1} \\ &= ((I - Q_k) - (I - Q_k)KG)^{-1} \\ &= ((I - Q_k)(I - KG))^{-1} \\ &= (I - KG)^{-1}(I - Q_k)^{-1}, \\ \Gamma_3 &= G\Gamma_1 = G(I - KG)^{-1}(I - Q_k)^{-1} \end{aligned}$$

By similar logic, we can find Γ_2 and Γ_4, resulting in

$$\begin{gathered} H_c = (I - Q_c)^{-1} = \\ \begin{bmatrix} (I - KG)^{-1}(I - Q_k)^{-1} & K(I - GK)^{-1}(I - Q_g)^{-1} \\ G(I - KG)^{-1}(I - Q_k)^{-1} & (I - GK)^{-1}(I - Q_g)^{-1} \end{bmatrix} \end{gathered} \tag{14}$$

It can be shown that Γ_1 is size $p \times p$ and Γ_3 is size $m \times p$.

From (12), we have that the vulnerability of this attack will be $V = \max_{links \in \mathscr{L}} \|T_{links}\|_\infty$, where each entry with label $\{H_c\}_{ji}$ corresponds to a nonzero attack Δ_{ij}.

Assume that an arbitrary attack occurs only on the controller. Then we have that every entry in Δ_{21} and Δ_{22} is zero and all nonzero entries $\delta_{ij} \in \Delta$ will exist where $i \leq p$. Thus all entries with label $\{H\}_{ji}$ in T_{links} are taken from the first p columns of H_c. Since the width of Γ_1 and Γ_3 is p, we have that V is dependent only on Γ_1 and Γ_3. And since Γ_1 and Γ_3 are dependent only on G, K, and Q_k, we have that the vulnerability of this attack is independent from the structure of the plant.

Now assume that the computation of vulnerability depends on Q_g. Then, from (14), at least one of the entries of T_{links} must be in Γ_2 or Γ_4. Also, because $(I - Q_g)^{-1}$ is full rank, at least one entry in each of Γ_2 and Γ_4 must be nonzero. Without loss of generality, assume that this entry is $\{\Gamma_2\}_{11}$. Since Γ_1 has width p, $\{\Gamma_2\}_{11} = \{H_c\}_{1,(p+1)}$. The corresponding attack that would lead to this requires that $\Delta_{(p+1),1} \neq 0$. And since Δ_{11} has height p, $\Delta_{(p+1),1} =$

$\{\Delta_{21}\} \neq 0$; therefore, there exists a nonzero attack on the plant that depends on Q_g.

□

Corollary 1. *The vulnerability analysis of the plant resulting from any arbitrary multiple link coordinated attack is separable from the controller if and only if the attack does not include an attack on the controller.*

Proof. This follows by switching the roles of the plant and the controller in Theorem 1. □

Recall that single link attacks are a special case of multiple link coordinated attacks; therefore, when considering single link attacks on the controller, by Theorem 1, the vulnerability analysis on links in the controller are guaranteed to be separable (and a similar statement can be said about the plant).

C. Design Questions for Minimizing Vulnerability in Feedback Systems

As with open-loop systems, we consider the problem of minimizing the vulnerability of feedback systems. In particular, we hold the input-output behavior G and K of our plant and controller constant and then seek to change the structures (P_g, Q_g) and (P_k, Q_k) of the plant and the controller in order to minimize the vulnerability of the combined feedback system. To date, this problem remains unsolved. However, separability may provide important insights into the solution for this problem.

For simplicity through this discussion, we consider only single link attacks; therefore, separability is guaranteed in all cases. Note, however, that the implications analyzed within this discussion will extend to multiple link attacks so long as each attack is restricted to only attacking the plant or only attacking the controller.

The vulnerability of the combined system is then defined as the most vulnerable link within the combined system. We also define the vulnerability of the controller as the vulnerability of the most vulnerable link within the controller after it is connected in feedback to the plant (and we define the vulnerability of the plant in a similar manner).

Given these definitions, separability shows the following:

- The structure (P_k, Q_k) of the controller can be freely changed without inadvertently increasing the vulnerability of the plant.
- The structure (P_k, Q_k) of the controller cannot be changed in order to reduce the vulnerability that exists within the plant.

Consider now a sub-problem to the problem above where we hold (P_g, Q_g) constant. Separability shows that, when minimizing vulnerability, the following scenarios may occur:

- We find a structure (P_k, Q_k) of the controller such that the vulnerability of the controller is the vulnerability of the plant. Then there is nothing more we can do to minimize the vulnerability of the combined system; the vulnerability is locked at the vulnerability of the plant. In this case, a better problem may be to minimize the vulnerability of the controller rather than minimizing the vulnerability of the combined system.
- We find that there exists no structure (P_k, Q_k) of the controller such that the vulnerability of the controller is less than the vulnerability of the plant. Then the vulnerability of the least vulnerable controller will be the minimum vulnerability of the combined system.

VI. Conclusion

In conclusion, we have developed a generalized attack model in the dynamical structure function domain that can handle a multitude of attack vectors including accidental failures in a system. Using this model we showed how to calculate the vulnerability of a system for destabilizing attacks on a networked system. In future work we plan to determine the vulnerability of systems in the face of other attack models, including state hijacking, which looks at how bounded inputs can change state trajectory without destabilizing the system, and inference threats, where an attacker does not affect the system, but can listen to traffic along communication links.

Furthermore, we also noted that to reduce the vulnerability of a system in the face of destabilization it is necessary to remove all feedback from the system, thus the only secure architecture across all models for attacks that use existing communication links occurs when $Q = 0$. Finally, we noted that for cases when removing feedback is impossible, there is a principle of separability in which the computation of vulnerability for attacks on the controller is separable from the structure of a plant when the plant or controller are not attacked simultaneously. This result will hopefully lead to a method for minimizing vulnerability in systems where feedback is essential, which will be explored in future work.

References

[1] Siddharth Sridhar, Adam Hahn, and Manimaran Govindarasu. Cyber–physical system security for the electric power grid. *Proceedings of the IEEE*, 100(1):210–224, 2012.

[2] G Andersson, P Donalek, R Farmer, N Hatziargyriou, I Kamwa, P Kundur, N Martins, J Paserba, P Pourbeik, J Sanchez-Gasca, et al. Causes of the 2003 major grid blackouts in north america and europe, and recommended means to improve system dynamic performance. *Power Systems, IEEE Transactions on*, 20(4):1922–1928, 2005.

[3] J. D. Sachs. Blackouts and cascading failures of the global markets. *Scientific American*, December 2008.

[4] André Teixeira, Daniel Pérez, Henrik Sandberg, and Karl Henrik Johansson. Attack models and scenarios for networked control systems. In *Proceedings of the 1st international conference on High Confidence Networked Systems*, pages 55–64. ACM, 2012.

[5] André Teixeira, Saurabh Amin, Henrik Sandberg, Karl Henrik Johansson, and Shankar S Sastry. Cyber security analysis of state estimators in electric power systems. In *Decision and Control (CDC), 2010 49th IEEE Conference on*, pages 5991–5998. IEEE, 2010.

[6] A. Wood and J.A. Stankovic. Denial of service in sensor networks. *Computer*, 35(10):54–62, Oct 2002.

[7] A. Rai, D. Ward, S. Roy, and S. Warnick. Vulnerable links and secure architectures in the stabilization of networks of controlled dynamical systems. In *American Control Conference*, pages 1248–1253, Montreal, Canada, 2012.

[8] Ralph Langner. Stuxnet: Dissecting a cyberwarfare weapon. *Security & Privacy, IEEE*, 9(3):49–51, 2011.

[9] J. Goncalves and S. Warnick. Necessary and sufficient conditions for dynamical structure reconstruction of lti networks. *IEEE Transactions on Automatic Control*, August 2008.