Hardware Trojans – A Systemic Threat

John Shield, Bradley Hopkins, Mark Beaumont, Chris North
Defence Science Technology Organisation
PO Box 1500, Edinburgh 5111, South Australia

Abstract
Hardware trojans are a systemic threat that can impact the operations and infrastructure of corporations and government organisations. In this paper, we evaluate a credible and organisation-wide hardware trojan threat from compromised network cards. Our research examines the systemic threat of hardware trojans with an actual hardware trojan implementation to evaluate the impact. Our hardware trojan can degrade network services inside a corporate network, controllable from outside the network. An external activation mechanism is used to activate the trojan; the implementation bypasses data encryption, firewall packet inspection, and is agnostic to software protection and the operating system.

Keywords: Hardware Trojan, Networking, Systemic Effects, Threat Estimation

1 Introduction
Hardware Trojans are intentional, malicious modifications to electronic circuitry designed to disrupt operation or compromise security – including circuitry added into Integrated Circuits (ICs). These ICs underpin the information infrastructure of many critical sectors including the financial, military, and industrial sectors. Consequently, hardware trojans pose a security risk to organisations due to the broad attack surface and specific organisations’ reliance on ICT infrastructure. Hardware trojans can be difficult to prevent and even more difficult to detect (Beaumont, Hopkins and Newby 2011). Most of the current security protection mechanisms implicitly trust the hardware, allowing hardware trojans to bypass software or firmware security measures (Goertzel and Hamilton 2013). Hardware trojans inserted during fabrication or design stages can become widely dispersed within an organisation and pose a systemic threat.

There has been significant research exploring the threat of hardware trojans, particularly targeting the exploit of a single computer or electronic device (Rajendran and etc al. 2010, Chakraborty and etc al. 2009, Tehranipoor and Koushanfar 2010, Beaumont, Hopkins and Newby 2011). However, the research does not explore the broader effects that can be achieved through the systemic design of a hardware trojan attack. Such threat estimation requires additional considerations, such as trojan coordination, supply chain logistics, organisational processes, core-business, and security policies.

This paper contributes the construction, threat estimation, and analysis of a hardware trojan as a system-wide effector. Understanding the threat these hardware trojans pose to organisations paves the way for future security systems that will defend organisations against this threat. We present some of the key differences when a hardware trojan threat is scaled to an organisation level and highlight these differences with a network chip hardware trojan as an example. Describing our example trojan with a hardware trojan taxonomy (Tehranipoor and Koushanfar 2010) it is small, adds a functional change, externally activated, and modifies the bandwidth of the network card.

The remainder of this paper is outlined as follows: Section 2 describes the threat model; Section 3 discusses scaling considerations for the hardware trojan; Section 4 discusses related work in hardware trojans; Section 5 outlines our implementation of a network hardware trojan; Section 6 evaluates our hardware trojan and how it fits the threat model; Section 7 provides our conclusions and proposes future work.

2 Threat Model
Supply chain vulnerabilities are the key vector for hardware trojans to be placed into an organisation’s information infrastructure. A major vulnerability point in the supply chain is global manufacturing, which provides a pathway for hardware trojans to be placed in COTS (commercial off the shelf) information infrastructure. Whilst the designs for some ICs may be produced by trusted local engineers, the majority of IC and electronic component development and consequent manufacturing occurs in facilities outside of the control of the design vendor. These facilities are therefore considered untrusted and provide the opportunity for an adversary to add hardware trojans during manufacture, or further through the supply chain (Samuel 2008). The impact of this can be widespread due to the limited number of manufacturers. During 2013, the top foundry company supplied around 46% of the global market and the top 13 companies supplied 91% of the global market (IC Insights 2014). The problem of global manufacturing is exacerbated for countries such as Australia that lack the local industries and infrastructure needed for producing ICT hardware (Beaumont, Hopkins and Newby 2011). Furthermore, organisations usually have preferred suppliers and procurement procedures, which can assist an adversary in inserting hardware trojans into an organisation. In this paper, we are more interested in chip level hardware trojans; where the capability is inserted at, or prior to, chip masking. However, we don’t preclude
other hardware trojans that require the addition, or modification, of physical circuits performed later in the supply chain. Once insertion of hardware trojans is achieved through the supply chain, a viable pathway for compromising an organisation’s information infrastructure is created.

Through physical or logical disruption, a hardware trojan can affect the confidentiality of information, integrity of information and availability of services throughout a business or organisation – undermining the operations and even reputation of the business. The impact of a hardware trojan attack can be long term and far reaching, in part because there are few current security measures that can detect or counter the effects of malicious hardware (Beaumont, Hopkins and Newby 2011).

In this paper, we specifically explore the threat of a network hardware trojan to an organisation. Network chips are an ideal insertion point for a systemic hardware trojan, due to the supply and distribution characteristics. The number of suppliers for communication chips is limited. In 2013, Broadcom had a 40% market share in these chips (Wheeler and Bolaria 2013). Furthermore, network chips are ubiquitous in all critical infrastructure including, PCs, servers, switches, communications infrastructure and embedded devices. These factors significantly increase the likelihood that a network hardware trojan can gain widespread penetration within an organisation.

Network infrastructure forms a critical component of an organisation’s operations, even when the core business is not technology orientated. Examples include: email communication between internal staff and external stakeholders; accessing corporate information, such as client information, inventory and schedules; and software usage, which can either be in the cloud or require network access for licensing reasons. Network services are so pervasive in organisations that minor delays or outages can have cumulative impacts on all staff and external stakeholders, which can be crippling to an organisation’s core business.

3 Hardware Trojans at Scale

Previous research (Rajendran and et al. 2010, Chakraborty and et al. 2009, Tehranipoor and Koushanfar 2010, Beaumont, Hopkins and Newby 2011) looked at threats to individual electronic devices, but did not estimate the hardware trojan threat to an organisation’s processes and systems. When evaluating a hardware trojan at a systemic level instead of a device level there are three key differences that come from scaling:

- **Insertion** - The method of insertion should gain widespread penetration into an organisation. This widespread penetration and delivery of a hardware trojan ameliorates uncertainty of where and whether the trojan will be placed.

- **Activation** - Hardware trojan activation to achieve a systemic effect needs to consider infrastructure and security policies as hurdles to activation. The timing and reach of the activation mechanism also needs to be considered to achieve desired coordination and affect multiple disparate hardware trojans.

- **Effect** - Once activated, the hardware trojan needs to compromise organisational-wide processes rather than specific functionality on a single machine or device. The trojans need orchestration and coordination to have wide-reaching effects that cannot be achieved alone by a single hardware trojan.

**Insertion** vectors for achieving systemic effects need to achieve broad penetration to deal with the unpredictability of placement and provide scalability of the threat. Targeting specific machines or classes of components may not be possible, due to unpredictability of where the hardware trojan is eventually placed. To be effective, the compromised component needs to be generic and widespread within an organisation, although the trojan may only need a few instances to be effective. A widespread trojan improves the chances that it will be placed in a critical location.

Hardware trojan insertion vectors requiring physical interaction with individual machines will not scale to the desired penetration levels to achieve the kinds of systemic effects that are the focus of this paper. Methods for insertion of a widespread trojan need to occur before or during mass production, or during supply chain logistics. Insertion would ideally occur through compromised IP cores, chip designs, or added as part of the manufacturing process.

**Activation** of a systemic hardware trojan to achieve coordinated effects needs to account for the infrastructure and security policies of the organisation. Organisational security infrastructure can hamper specific types of activations signals, such as network data being blocked by firewalls and gateways. Organisational security policies can also block many side channel activation signals that require physical access or software access to the machine.

Consequently, the activation signal needs to be resilient, widespread and easily propagated to overcome the uncertainty of the hardware trojan’s placement and for coordinating the activation for multiple instances.

**Effect** of a systemic hardware trojan is most severe when it impacts the organisation as a whole, namely its core-business and processes. Factors that contribute to severity include sublety and enduring nature of effect, time cost incurred to discover, criticality of affected equipment and ability to remediate affected equipment with compatible hardware trojan free replacements. A traditional hardware trojan defines success as compromising very specific functionality of a machine or process. However, the compromised machine and the functionality may be unused within an organisation, or its effects may not reach beyond a single individual machine or person.

4 Related Work

There has been significant research work into describing and classifying hardware trojans (Rajendran and et al. 2010, Chakraborty and et al. 2009, Tehranipoor and Koushanfar 2010, Beaumont, Hopkins and Newby 2011). However, the literature is heavily weighted towards
hardware trojans that are designed and described as a threat to a single computer or device. In the literature, there has been minimal implementation or technical analysis for hardware trojans operating at a larger scale.

Some research implementations of hardware trojans are: CPU based hardware trojans (King and et al. 2008, Wang and et al. 2012) that can steal passwords, or break privilege protections; encryption hardware trojans (Lin, Burleson, and Paar 2009, Agrawal and et al. 2007, Jin and Makris 2009) to extract secret keys; methods of DoS (Denial of Service) on general circuitry (Shiyanovskii and et al. 2009, Wei and Potkonjak 2013); adding communication channels using USB peripheral trojans (Clark, Leblanc, and Knight 2009); and adding communication channels using a network card trojan (Farag, Lerner, and Patterson 2012). These previous research implementations focus on how the hardware trojan can impact individual computers or devices, while our paper explores the wider effects of a hardware trojan implementation on an organisation.

Previous research into defence mechanisms against hardware trojans focuses on security of the hardware design and in-built detection methods (Tehranipoor and et al. 2011). Some of these systems use in-built delay monitoring and power monitoring of the design against pre-calculated values (Wei, Kai, and Potkonjak 2012, Narasimhan and et al. 2012). Data guards can also be used to prevent trojan activation by scrambling input data (Waksman and Sethumadhavan 2011). Additional software tools can be used to verify that the EDA tools create correct designs (Potkonjak 2010).

These defence mechanisms against hardware trojans will eventually need to be evaluated in the context of systemic defence. Differences in implementation and operation of systemic hardware trojans could significantly impact effectiveness of the defence. Our work outlines some of the possible differences.

Previous research into hardware trojan implementations and defence mechanisms, only considers the security measures and impact for the immediate computer or device. It fails to consider defence mechanisms that can be implemented through procedures or guard electronics on an organisational scale. The lack of scale in the defence mechanisms is a side effect of hardware trojan implementations being narrowly focused on individual devices and machines. We hope to address this deficiency by exploring the design characteristics of large scale trojan threats.

5 Implementation

Our network hardware trojan performs remotely activated degradation of service, targeted at a RTL8111E Realtek Ethernet Controller chip.

5.1 Design Goals

Our design goals for the network hardware trojan were: easy insertion into the supply chain, simplicity in the design, small footprint to increase difficulty of detection, and broad and decentralised activation mechanisms.

Supply Chain: The trojan was designed to be easily inserted post-design during the chip manufacturing, or through another supply chain vulnerability. This is achieved by requiring no modifications to the original logic design and only accessing external signals on the IC for the design and implementation of our trojan. This fits our described threat model whereby manufacturing provides the opportunity for the broadest dissemination of a hardware trojan threat.

Simplicity: Although the chip handles gigabit ethernet, the trojan mechanisms used are simple and low frequency. This simplicity reduces the size and improves the robustness of the trojan system, and assists in making it easier to add to a manufactured chip.

Small: A smaller design creates a smaller footprint in the silicon, which is more difficult to detect. This in turn increases the chance that the trojan hardware activates and is used, and also increases the length of time it is present before potential detection and removal.

Systemic Activation: The activation signal needs to enable decentralised and widespread activation. This signal also needs to overcome common security measures such as firewalls and gateways.

5.2 Design

Hardware trojans are usually composed of a trigger and a payload (Chakraborty and et al. 2009). The trigger is the activation mechanism and the payload generates the effect. Prior to triggering, a hardware trojan lies dormant without interfering with the operation of any electronics. The trigger mechanism for our network hardware trojan is based on a communication channel in network packet timing, while the payload is an adjustable degradation level of the ethernet channel through noise injection into the ethernet controller’s clock.

5.2.1 Trigger

The trigger mechanism, for the network hardware trojan, uses the Ethernet controller chip’s activity LED light as a method to access the packet timing. The activity LED is used to give a very broad indication to a user of the current network traffic. For the RTL8111E chip, network activity causes the LED light to cycle on then off over a 160 millisecond period. There is a delay between these 160 millisecond cycles when there is no immediate network activity. The behaviour of the Ethernet Controller is shown in Figure 1.

![Figure 1: Ethernet controller chip LED behaviour](image)
hardware trojan. Figure 2 shows how different network activity is able to modulate the period of the activity LED.

This simple timing channel contains noise from normal network traffic. To overcome the impact of noise, a sufficiently long activation code can prevent false positives. Repeated signalling can also overcome noise in receiving the signal. In a few cases, the LED timing channel is absent, due to continuous network activity. However, most systems do not continuously communicate and a signal can be received during any breaks in normal communication, provided a sufficiently robust protocol is utilised.

For the RTL8111E chip, the period of the activity LED can be obtained with a coarse sampling of the signal, as shown in Figure 3.

The samples are matched against two pre-defined trigger sequences implemented in our hardware trojan. There is an activation sequence that increments a counter, which controls the network degradation level, and a deactivation sequence that resets the trojan.

The left half of Figure 4 shows logic for sampling. This generates a pulse every 21 milliseconds, based on the counter size and the 25MHz input clock from the RTL8111E chip (CLK). The sampling is also synchronised to the rising edge of the LED output of the RTL8111E chip.

The right half of Figure 4 shows logic for sample matching. The sample rate pulses are used to clock in the LED state into a shift register. To determine whether the sequence matches the trigger or reset signals, comparators are used to check for the rising edges for pre-defined sequences in timing.

5.2.2 Payload

The network hardware trojan payload performs a degradation of network services. It uses noise injection into the ethernet controller chip’s clock circuitry in the form of a bias voltage. This voltage slightly changes the resonant frequency on the external crystal. The change in frequency desynchronises the clock of the ethernet controller chip with the ethernet channel. This causes bit errors in the ethernet channel. Figure 5 shows simple bias voltage circuitry that can be directly fed into the crystal. Our demonstrator (described later) uses a pulse-width modulation (PWM) source where the pulse width sets the bias voltage. Figure 6 shows how an adjustable PWM can be generated using a small number of gates. Figure 7 shows where the hardware trojan injects the bias voltage into the standard crystal clock circuit. The adjustable bias voltage allows for variable degradation of the ethernet channel.
The hardware trojan is designed to minimise the size needed for the implementation, facilitating easy hardware modification and making it more difficult to detect. For our demonstrator, we implemented the hardware trojan in functionally similar firmware instead of circuitry.

### 5.3 Demonstaror

For our demonstrator of the hardware trojan, we used the ENW02A-1-BC01 Gigabit Ethernet PCI-Express Card. The RTL8111E controller chip is part of this card. We implemented our hardware trojan externally on a PIC16F690 Development Board and attached it to the pins of the controller chip. The experimental setup can be seen in Figure 8. An actual hardware trojan would be added inside the ethernet controller chip, most likely during manufacture.

In the demonstrator, firmware on the PIC microcontroller is used to implement the hardware trojan functionality – however, the fundamental operation is the same. This functionality would be implemented directly in logic if the hardware trojan was inserted during manufacturing.

The trigger detection sequence utilises the PIC timers to measure the period between rising edges. These periods are compared using coarse values against an expected sequence of delays.

The payload for the implemented version of the hardware trojan is achieved via an adjustable PWM output signal generated by the PIC. This is injected through a resistor onto one of the clock crystal inputs (Figure 7).

### 6 Evaluation

#### 6.1 Network Performance Adjustability

Figure 9 shows the range of degradation effects the hardware trojan can implement. The range of settings allows for a spectrum of disruption, from minor network slowdown to complete disabling of access.

The effects of our network hardware trojan are extremely difficult for users and IT support staff to debug even while activated. During our testing there was no error reporting to the user of a problem (under the Ubuntu OS), until the hardware trojan was set to completely disable the network card. Furthermore, the amount of reported TCP/IP packet loss was minimal even under high network degradation. This is attributed to the operation of the Transport Control Protocol (TCP), whereby packet losses are treated as being caused by congestion and this reduces bandwidth. Academic studies (Kumar 1998) have shown that dropping 1 in every 70 packets causes degradation of bandwidth performance in the order of 50%. These characteristics make our trojan difficult to isolate, even whilst operating.

#### 6.2 Threat Effectiveness

The network trojan is well suited to implement a systemic effect. It has the desired insertion, activation, and effect characteristics to provide a coordinated and disruptive organisation-wide attack.

**Insertion:** The network hardware trojan can be easily inserted into operation within an organisation. Firstly, there are a limited number of suppliers of ethernet controller chips making it easier to compromise the supply chain. One company supplies 40% of the communication chip market (Wheeler and Bolaaria 2013). Consequently, a potential compromise within one company could result in significant hardware trojan penetration within an organisation’s infrastructure. Such a compromise may be from a dissatisfied employee, organised group, or state sponsored actor. Secondly, network enabled devices are ubiquitous in today’s
organisational environments, providing numerous and widely spread device locations that can be compromised.

**Activation:** The common security policies within organisations will not limit our network hardware trojan and the activation method scales up well. Firstly, the activation method of the hardware trojan occurs before any software protections and is not reliant on software or other hardware. This allows the trojan to work irrespective of where the ethernet controller chip is located. Secondly, the signal is also based purely on packet timing and ignores data content, allowing it to bypass the data encryption, packet inspection and port blocking commonly found on firewalls and gateways. It will also work in the presence of packet encryption. Finally, the activation signal scales easily using the existing network hierarchy. The signal is simple enough to easily replicate and can activate all intermediate network card trojans as it propagates through the network. The signal can be blocked by noise in the form of other network traffic. However, the activation can still be received, provided a sufficiently robust protocol is utilised.

**Effect:** Our network hardware trojan is targeted at disrupting organisational operations. Firstly, a degradation of networked services can significantly reduce organisational efficiency by slowing down communication, information access and limiting software usage. Networked devices can include: servers, desktops, gateways, routers, faxes, phones and printers. Secondly, multiple network cards are chained together in connecting any service and it only takes a single network card within the chain to adversely affect the service. Thirdly, the variable effect of the trojan with network degradation and possibly intermittent behaviour encourages temporary workarounds rather than directly addressing the problem. These workarounds may prove more costly for an organisation in the long run, taking up significantly more time and encouraging departure from standard procedures. Departure from standard procedures may create new security vulnerabilities, such as using a personal email for business activities due to the corporate email being too slow. Finally, the trojan can remain unidentified for a long period. Degradation of networked services can come from any number of factors and in most cases, denial of service attacks aside, these are typically the result of a single hardware, or software failure. Having a wide attack surface, coupled with an intermittent effect will severely retard technical support within the organisation identifying the nature of attack and consequently isolating it.

### 7 Conclusions & Future Work

In this paper, we have outlined a natural development for hardware trojan research into systemic effects. Currently, there is a narrow focus on the threat estimation to individual machines, with little analysis on systemic effects of hardware trojans targeted at an organisational level. For our contribution, we have designed and implemented a network hardware trojan to better characterise the threat posed by organisation-wide hardware trojans. The trojan we designed was small, easy to implement and can be leveraged to provide coordinated and variable disruption to most organisations.

Our research has demonstrated that there are key differences, in insertion, activation, and effect, when scaling hardware trojan effects from an individual computer to an organisation. We have outlined the key differences as a basis upon which later work can build upon in researching and developing appropriate defence mechanisms.

Future work will look at additional system-wide threats and recommended policies and protection mechanisms that could be implemented by organisations.

### 8 References


