Learning-Based Denial-of-Service Attack Mitigation in Network-on-Chip Architectures

Undergraduate

ABSTRACT

Recent advances in chip manufacturing technologies have enabled computer architects to integrate an increasing number of processor cores and other heterogeneous components on a System-on-Chip (SoC). State-of-the-art SoC designs consist of many Intellectual Property (IP) cores that interact using a Network-on-Chip (NoC) architecture. NoC is widely employed by multi-core architectures to cater to their communication requirements. NoC has allowed computer architects to fully utilize the computational power in an SoC by facilitating low-latency and high-throughput. With the increased complexity of SoCs, manufacturers have favored IP licensing and outsourcing, where only a subset of IPs are manufactured in house and the rest is sourced from third-party vendors. Hence, the reliance on global supply chains for obtaining third-party IP cores, the distributed nature of the NoC and its increased usage has made the NoC a focal point for potential security threats making the NoC IP cores as the ideal candidates to insert hardware Trojans. An implanted hardware Trojan can launch a wide variety of attacks ranging from denial-of-service attacks, eavesdropping attacks, data integrity attacks, buffer overflow attacks, and side-channel attacks. Specifically, Denial-of-Service (DoS) attacks pose a serious threat in degrading the SoC performance by flooding the NoC with unnecessary packets. The primary objective of a DoS attack is to prevent legitimate users from accessing services and information. In the context of NoC, malicious IPs sending unnecessary requests to IPs can delay legitimate requests leading to delay of service or denial of service. Such "flooding" type of DoS attacks can cause congestion in the network, further degrading performance and energy efficiency [**?** **?**].

Previous work that explored defenses against DoS attacks proposed traffic latency comparison [**?**] and security verification techniques [**?**]. However, these approaches give sub-optimal results due to inherent drawbacks in their methodologies such as injection of additional packets that can further congest the network [**?**] and the inability to detect if an attack happens [**?**]. Fiorin et al. introduced a countermeasure against DoS attacks that has an architecture similar to our work [**?**]. However, their method is fundamentally different form ours since they monitor the bandwidth considering the data loaded/stored by an initiator from/to a specific memory block or range of addresses. Charles et al. proposed to statically profile the normal behavior of the SoC and detect DoS attacks during runtime [**?** **?**]. Their approaches made an unrealistic assumption, highly predictable NoC traffic patterns, which allowed the construction of linear statistical bounds to detect DoS attacks. Kulkarni et al. proposed a SVM-based approach for hardware Trojan detection for many-core platforms. But, their method includes highly subjective features such as destination IP and source IP [**?** **?**]. While such methods are efficient when the applications and application mappings are fixed, they are not suitable in many real-world scenarios where NoC traffic behavior can be altered due to task migration, task preemption, and input variations, etc.

As a potential solution to address such runtime variations, in our work, we explore the viability of using a learning-based approach for DoS attack detection and localization. While machine learning (ML) has shown promising results for optimizing NoC power consumption [**?**], to the best of our knowledge, there is no comprehensive method that secure NoC-based SoCs from DoS attacks using a learning-based approach. Following the realistic architecture model proposed in [**?**], we use the "GARNET2.0" framework [**?**], that is integrated with the gem5 [**?**] cycle-accurate full-system simulator, to build the architecture model for experimental evaluation. DoS attack scenarios are modeled by utilizing IPs that did not run instances to inject memory request packets to the NoC increasing the overall network traffic. The NoC traffic data is gathered at each router by using probes attached to routers and the collected data is sent to dedicated IP to make speculation about the state of the NoC. The overall probability for an attack is calculated using the weighted average of probabilities from the results of models dedicated to each router. Major contributions of our work are as follows:

- We outline features that can be extracted from NoC traffic as well as engineered features, and experimentally evaluate the suitable features utilizing a genetic algorithm based approach.
- We propose an ML-based DoS attack detection and localization method that trains ML models during design time and uses the trained models to classify network traffic behavior as normal or attack during runtime, to detect and localize flooding type of DoS attacks.
- We perform a comprehensive exploration of different ML models to select the best fit for the given architecture and threat models.
- We evaluate the effectiveness of our approach against different NoC topologies with varying number of IP cores.
- Our approach achieves high accuracy in DoS attack detection and localization across different NoC traffic patterns caused by various applications and application mappings.
- Our approach does not rely on subject features such as destination IP and source IP as opposed to previous work [**?** **?**].
- Our observations reveal that the key to achieving high accuracy is to carefully craft features out of the data extracted from NoC traffic.
- Our approach can detect and localize DoS attacks in real-time with detection times comparable to previous work [**?** **?**] without requiring highly predictable traffic patterns.