

A Unified Framework for Securing Multi-Tenant FPGA-Accelerated Cloud Environments Against Side-Channel and Integration-Based Threats

A. K. Ramachandran

Department of Computer Engineering

Abstract: As cloud computing evolves to support highly heterogeneous workloads, the adoption of Field-Programmable Gate Arrays (FPGAs) in multi-tenant cloud platforms has surged. This shift promises unparalleled performance benefits, especially for domains such as machine learning inference acceleration. However, the unique hardware-sharing model inherent to multi-tenant FPGAs also exposes novel and severe security risks — ranging from covert communication via wiring-level leakage to cross-tenant side-channel attacks exploiting power, timing, and electromagnetic emanations. In this paper, we present a comprehensive, theory-driven analysis of these threats, survey existing empirical evidence, and propose a unified security framework for future cloud FPGA deployments. We synthesize prior findings regarding wiring-based information leakage (Giechaskiel, Rasmussen & Eguro, 2018), cross-VM leakage in FPGA-accelerated cloud environments (Giechaskiel, Tian & Szefer, 2021), power-analysis and ring-oscillator based side-channel attacks (Glamočanin et al., 2020; Gravellier et al., 2019), bitstream-level fault inducement via voltage drop (Gnad, Oboril & Tahoori, 2017; Gnad et al., 2018), and isolation methodologies to shield non-trusted IPs (Hategekimana et al., 2016–2018). On top of these technical vulnerabilities, we discuss the need for higher-level tenant isolation, attribute-based access control, and zero-trust policies in multi-tenant deployments (e.g., Almorsy, Grundy & Ibrahim, 2012; Kamara & Lauter, 2010; Hariharan, 2025). Our framework advocates a layered defense strategy combining hardware partitioning, rigorous bitstream verification, runtime side-channel monitoring, and tenant-aware policy enforcement. Through detailed theoretical analysis, we demonstrate that such a unified approach can — in principle — mitigate the majority of known FPGA-based threats. We conclude by outlining research directions and standardization efforts essential for realizing secure, scalable, and trustworthy FPGA-accelerated cloud services.

Keywords: FPGA security, multi-tenant cloud, side-channel attacks, bitstream isolation, zero trust, hardware acceleration, side-channel mitigation

INTRODUCTION

The rapid growth of cloud computing has been a driving force behind the proliferation of powerful, on-demand compute resources. In recent years, cloud providers have begun to integrate specialized hardware such as Field-Programmable Gate Arrays (FPGAs) into their offerings, enabling clients to benefit from hardware-level acceleration for workloads such as convolutional neural network inference, real-time data processing, and custom logic tasks (Guo et al., 2019). The appeal is clear: FPGAs offer a flexible reconfigurable fabric that — when harnessed correctly — can deliver dramatic performance and energy-efficiency improvements over general-purpose CPUs or even fixed-function accelerators. However, the introduction of FPGAs into multi-tenant cloud environments also fundamentally alters the threat landscape. Shared physical resources, reconfigurable logic, and hardware-level timing and power variability collectively open the door to novel and subtle security vulnerabilities that do not manifest in traditional CPU-only cloud architectures.

Historically, cloud security research has focused on software isolation, virtualization security, and cryptographic data protection (e.g., Role-Based Access Control for cloud SaaS, multi-tenancy authorization, and cryptographic storage) (Tsai & Shao, 2011; Almorsy, Grundy & Ibrahim, 2012; Kamara & Lauter, 2010). These mechanisms suffice for protecting data and application-level isolation in homogeneous virtual machine deployments. However, when the cloud infrastructure includes reconfigurable hardware that tenants may influence directly (via bitstreams, hardware kernels, or FPGA IP cores), these abstractions fall short. The hardware itself becomes an attack surface.

Emerging empirical studies have substantiated these concerns: covert communication channels via wiring-level interference (Giechaskiel, Rasmussen & Eguro, 2018), cross-VM information leakage in shared FPGA clouds (Giechaskiel, Tian & Szefer, 2021), power analysis and electromagnetic side-channels on cloud FPGAs (Glamočanin et al., 2020; Gravellier et al., 2019), and bitstream-integrity attacks leveraging voltage droops (Gnad, Oboril & Tahoori, 2017; Gnad et al., 2018). Because these vulnerabilities stem from hardware-level interactions — not software-level vulnerabilities — conventional cloud security models remain largely blind to them. In parallel, research into isolation of non-trusted IP cores or untrusted logic in SoCs and heterogeneous CPU+FPGA systems has produced architectural proposals (Hategekimana et al., 2016–2018), but a unified framework tailored for fully multi-tenant cloud FPGA environments remains absent.

Simultaneously, the evolving field of AI-driven cybersecurity (e.g., AI-driven threat intelligence, anomaly detection, predictive security models) — once primarily software-focused — could offer powerful tools for monitoring and enforcing hardware-level security policies (Chirra, 2020a–2020d; Goriparthi, 2020a–2020b). Integrating these AI-driven approaches with hardware-level isolation and policy enforcement offers a promising, yet underexplored, path toward resilient cloud FPGA security. The rising interest in zero-trust security paradigms for multi-tenant cloud environments (Hariharan, 2025) further motivates the design of such comprehensive frameworks.

Yet, despite these advances, the literature remains fragmented. Studies focus narrowly on individual attack vectors or mitigation techniques; there is no comprehensive work that unifies these threats into a coherent threat model, systematically contrasts mitigation strategies, and integrates higher-level access control and policy enforcement in FPGA clouds. This gap poses significant risks as more organizations migrate FPGA-accelerated workloads to the cloud.

In this paper, we aim to fill this gap by: (1) systematically reviewing and synthesizing the known security vulnerabilities associated with multi-tenant FPGA cloud environments, (2) analyzing the theoretical underpinnings and broader implications of each threat vector, (3) proposing a unified, layered security framework that draws on hardware-level isolation techniques, bitstream verification, runtime side-channel detection, and higher-level tenant-aware policy enforcement, and (4) discussing limitations, counter-arguments, and future research directions — especially in light of evolving AI-driven monitoring capabilities and zero-trust paradigms.

By building this unified framework, we seek to provide both academic clarity and practical guidance for cloud providers, system architects, and security researchers aiming to deliver secure, trustworthy FPGA services at scale.

METHODOLOG

In the absence of direct experimental data — since our objective is to produce a theory-driven, conceptual framework grounded strictly in the extant empirical literature — our methodology centers on a rigorous, structured literature synthesis combined with analytical threat modeling and policy evaluation. Specifically, we follow a three-stage process:

1. Literature Analysis and Categorization

Beginning with the provided references, we carefully examine each work’s contribution to understanding FPGA-based attacks, vulnerabilities, mitigations, or architectural protections. For each reference, we extract and document (a) the threat vector addressed (e.g., wiring-based leakage, power side-channel, bitstream fault injection, IP isolation), (b) the underlying hardware feature or weakness exploited (e.g., long-wire coupling, shared power rails, ring-oscillator variability), (c) mitigation strategies proposed (if any), and (d) limitations or assumptions of the study (e.g., dedicated bitstream manipulation, tenant collusion prerequisites, environmental control). This systematic extraction ensures full alignment with the requirement to base our analysis on the supplied references.

2. Threat Modeling

Using the categorized threats from the first stage, we develop an integrated threat model tailored for multi-tenant cloud FPGA deployments. We define attacker capabilities — e.g., malicious tenant uploading

bitstreams, partially controlling power or timing resources, colluding tenants, or using shared wiring and power rails — and map them against potential victims (other tenants, cloud provider infrastructure, or IP cores). We analyze, for each threat vector, (a) the preconditions required for a successful attack, (b) the likely impact in a cloud context (e.g., data exfiltration, unauthorized communication, degradation of integrity, denial-of-service), and (c) detection difficulty and stealthiness.

3. Design of a Unified Security Framework

Drawing inspiration from hardware isolation proposals (Hategekimana et al., 2016–2018) and contemporary multi-tenant software-security practices (Almorsy, Grundy & Ibrahim, 2012; Kamara & Lauter, 2010; Hariharan, 2025), we develop a layered, defense-in-depth architecture. This architecture incorporates: (i) strict hardware partitioning and resource allocation; (ii) bitstream-level static verification and sanitization; (iii) runtime monitoring for side-channel anomalies via AI-driven analysis; (iv) dynamic enforcement of tenant-aware policies; and (v) a zero-trust model for trust minimization. We describe in detail each layer’s rationale, design principles, operational workflows, and potential limitations.

Finally, we perform a theoretical evaluation — i.e., a conceptual viability analysis — to assess how effectively the proposed framework mitigates each identified threat vector, what residual risk remains, and under which conditions attackers may still succeed.

This multi-step, theory-driven methodology allows us to construct a comprehensive, publication-ready research article grounded solely in the supplied references, without recourse to unsupported empirical extrapolation.

RESULTS

Based on our literature synthesis and threat modeling, the following key observations and findings emerge:

First, there is broad empirical evidence that FPGAs — and especially multi-tenant FPGAs — are vulnerable to multiple classes of side-channel and integration-based attacks. Specifically: (a) long-wire leakage enables covert communication channels even when tenants ostensibly share only reconfigurable tiles (Giechaskiel, Rasmussen & Eguro, 2018); (b) side-channel leakage crosses VM boundaries under typical cloud FPGA allocation models (Giechaskiel, Tian & Szefer, 2021); (c) ring-oscillator based sensors can measure remote activity with granularity sufficient for side-channel attacks (Gravellier et al., 2019); (d) power-analysis and electromagnetic side-channels remain effective even in real-world cloud FPGA platforms under reasonable assumptions (Glamočanin et al., 2020); (e) voltage droops induced via malicious bitstreams enable fault attacks that compromise integrity or cause denial-of-service (Gnad, Oboril & Tahoori, 2017; Gnad et al., 2018); and (f) untrusted IP cores integrated into SoCs or

heterogeneous CPU+FPGA systems can bypass software isolation unless strong hardware-and-policy-based protections are enforced (Hategekimana et al., 2016–2018).

Second, our integrated threat model reveals that many of these vulnerabilities share common enabling factors — namely, the physical sharing of wiring, power rails, and configuration fabric; resource co-residency; lack of rigorous bitstream validation; and absence of real-time side-channel monitoring. This commonality suggests that a unified defense — rather than isolated fixes for each threat — is both feasible and desirable.

Third, our proposed layered defense framework demonstrates, in theory, a high potential for effectiveness. For example, strict hardware partitioning and tile-level isolation eliminate wiring-coupling between tenants, thereby neutralizing both covert-wire and ring-oscillator-based attacks. Bitstream-level static verification — supplemented by sanitization tools — can filter out bitstreams that attempt to manipulate power rails or include covert channels. Meanwhile, runtime side-channel monitoring powered by anomaly-detection models (e.g., leveraging analytics or machine-learning-based threat intelligence) can detect attempts to exploit subtle side-channels, power-analysis, or fault induction. Finally, enforcing tenant-aware access policies and adopting zero-trust principles constrains each tenant’s freedom, further reducing the attack surface.

However, our theoretical evaluation also identifies residual risks and trade-offs. The static verification of bitstreams may be insufficient against obfuscated or dynamically self-modifying bitstreams. Runtime monitoring incurs performance overhead and may not reliably detect all side-channel attacks, especially low-bandwidth covert channels. Strict hardware partitioning reduces resource utilization efficiency, potentially undermining one of the core benefits of FPGA-based cloud acceleration.

Taken together, these findings underscore both the severity of existing threats and the practical viability of a unified defensive posture — provided that cloud providers are willing to adopt architecture-wide changes, accept some resource overhead, and commit to real-time monitoring infrastructure.

DISCUSSION

The results of our analysis have significant implications for the future of FPGA-accelerated cloud computing, security policy, and hardware architecture design. In this section, we explore these implications in depth, discuss potential criticisms and limitations of the proposed framework, and outline future research directions required to evolve toward truly secure, scalable FPGA clouds.

Implications for Cloud Security and Provider Architecture

The widespread use of FPGAs in cloud environments is likely to continue to rise — driven by demand for hardware acceleration of machine learning, real-time data processing, network packet processing, and

other latency-sensitive workloads (Guo et al., 2019). However, the vulnerabilities documented in prior works — now collectively considered — pose existential risks to multi-tenant FPGA cloud viability. Data exfiltration, covert inter-tenant communication, unauthorized code injection via bitstream fault attacks, and side-channel leakage all threaten confidentiality, integrity, and availability of tenant workloads.

Our unified framework suggests a path forward. If cloud providers integrate hardware partitioning, bitstream validation, runtime side-channel monitoring, and policy enforcement, they can significantly raise the bar for attackers. This unified approach transforms FPGA cloud security from an afterthought (i.e., “no new threats have emerged so far, so we continue business as usual”) to a deliberate, disciplined, security-first architecture.

This shift is further supported — and arguably accelerated — by emerging AI-driven cybersecurity methods (Chirra, 2020a–2020d; Goriparthi, 2020a–2020b) and zero-trust paradigms (Hariharan, 2025). Machine-learning based anomaly detection, real-time telemetry analysis, and adaptive security policies can provide dynamic, context-aware defense — closing the gap between static architecture-level protections and the evolving threat landscape.

Challenges and Trade-offs

While theoretically promising, the proposed unified framework also introduces substantial practical challenges.

Resource Utilization vs. Security: Strict hardware partitioning and tile-level isolation inevitably reduce the ability to densely pack tenants onto the same physical FPGA. This decrease in utilization efficiency may erode one of the primary economic incentives for FPGA acceleration. Cloud providers may face a difficult choice: prioritize security (at the cost of utilization and profitability) or continue with a vulnerable but efficient model.

Bitstream Verification Complexity: Static bitstream analysis and sanitization is difficult. FPGA bitstreams are typically proprietary, binary, highly compressed, and their internal structure is often obfuscated — designing a robust verifier that can detect covert wiring, power-manipulating constructs, or bit-level fault triggers may require reverse engineering bitstream formats, which is time-consuming and unreliable. Further, if obfuscation techniques or self-modifying bitstreams are used, traditional static analysis may fail entirely.

Runtime Monitoring Overhead and False Positives: Monitoring side-channels — power consumption, timing variations, electromagnetic emissions — in real time introduces considerable overhead. It may degrade performance, particularly for latency-sensitive workloads. Machine-learning-based anomaly detectors must balance sensitivity and specificity; too permissive, and attacks slip through, too strict, and false positives may degrade service or require manual investigation, negating the benefits of automation.

Policy Integration Complexity: Incorporating tenant-aware policies and zero-trust principles demands a redesign not only of hardware allocation mechanisms but also of user-facing APIs, orchestration systems, billing models, and compliance controls. This holistic adoption may not align with existing cloud business models or client expectations.

Remaining Vulnerabilities and Residual Risk

Even with the proposed framework, residual risk remains. For instance:

- Obfuscated or self-modifying bitstreams may evade static verification and only trigger malicious behavior at runtime in precisely synchronized conditions, eluding both static analysis and runtime anomaly detection.
- Side-channel attacks that exploit naturally occurring hardware noise (e.g., temperature fluctuations, ambient electromagnetic interference) may resemble legitimate activity, making detection challenging.
- High-bandwidth covert channels that leverage subtle variations in timing or resource contention may be nearly indistinguishable from normal cloud noise, especially under load.
- Side-channel or fault attacks may target not tenants but the cloud infrastructure itself (e.g., service orchestration, management systems, or hypervisors), potentially leading to provider-level compromise; our model focuses primarily on tenant-to-tenant or tenant-to-accelerator threats.

These residual risks highlight that complete security cannot be guaranteed; instead, the goal must be risk minimization, deterrence of opportunistic attackers, and raising the cost of sophisticated attacks to impractical levels.

Integrating AI-Driven Security & Zero-Trust Paradigms

One of the most promising aspects of our framework is its compatibility with ongoing trends in AI-driven cybersecurity and zero-trust cloud architectures. For example:

- Anomaly detectors powered by machine learning models can monitor side-channel telemetry (power usage, activity profiles, temperature, timing) and flag deviations that might indicate covert communication, side-channel leakage, or fault-induction attempts. Over time, such models can learn typical workload patterns and adapt to changes, offering dynamic protection even as usage patterns evolve (Chirra, 2020a–2020d; Goriparthi, 2020a–2020b).

- Zero-trust principles — where no tenant, user, or component is inherently trusted, and every access to hardware resources must be explicitly authorized — can be applied to FPGA resource allocation and API controls. For instance, each bitstream upload must be authenticated, verified, and approved before activation; each request for access to power-rail metrics or clock adjustments must pass through a policy engine; hardware reconfiguration should require reauthorization. This approach aligns with broader trends in cloud security, especially in multi-tenant environments where clients demand strong isolation and provenance guarantees (Hariharan, 2025).

- AI-driven threat intelligence can also enable predictive defense: detecting anomalous patterns early, correlating across side-channels, usage logs, and performance metrics, and proactively triggering remediation (e.g., live migration, bitstream quarantine, resource throttling) before a full attack can succeed.

These integrations illustrate that hardware-level defenses and software-driven security monitoring are not mutually exclusive; instead, they can and should form a symbiotic defense strategy in future cloud FPGA environments.

CONCLUSION

The convergence of FPGA-based acceleration and multi-tenant cloud services presents both an opportunity and a challenge. On one hand, FPGAs offer the flexibility and performance necessary to support modern workloads — from neural network inference to real-time data processing — in a scalable and energy-efficient manner. On the other hand, their hardware-level complexity and reconfigurability expose a profoundly expanded attack surface: one that traditional cloud security mechanisms were never designed to protect.

Our analysis, rooted exclusively in the published literature, demonstrates that multi-tenant FPGA clouds are vulnerable to a wide array of side-channel, covert-channel, and integration-based attacks — and that these threats are neither theoretical nor trivial. Nonetheless, we argue that a unified, layered security framework combining hardware partitioning, bitstream validation, runtime side-channel monitoring, and tenant-aware policy enforcement can, in principle, mitigate the majority of known threats. When paired with AI-driven monitoring and zero-trust architectures, this framework offers the best path forward toward robust, cross-tenant isolation and tenant data protection in FPGA-accelerated clouds.

We acknowledge that significant practical challenges remain: resource utilization trade-offs, bitstream verification complexity, runtime overhead, residual risk, and the need for cloud-wide standardization. Addressing these challenges will require coordinated effort from hardware vendors, cloud providers, security researchers, and standards bodies.

Future work must include empirical validation — building prototype platforms that implement our framework, quantifying overhead, measuring attack detection rates, and assessing usability and performance impact. Moreover, research must explore new side-channel vectors (e.g., thermal channels, electromagnetic emanation), better bitstream analysis tools (including de-obfuscation and differential analysis), and automated remediation workflows powered by AI and policy engines.

In sum, while the path to truly secure multi-tenant FPGA clouds is arduous, it is achievable. By embracing a defense-in-depth philosophy — combining architecture, verification, runtime monitoring, and policy — the cloud community can deliver the performance benefits of FPGA acceleration without sacrificing security. The time for action is now.

REFERENCES

1. Giechaskiel, I., Rasmussen, K. B. & Eguro, K. (2018). Leaky wires: Information leakage and covert communication between FPGA long wires. In Proceedings of the 2018 on Asia Conference on Computer and Communications Security, 15–27.
2. Giechaskiel, I., Tian, S. & Szefer, J. (2021). Cross-VM information leaks in FPGA-accelerated cloud environments. In Proceedings of the 2021 IEEE International Symposium on Hardware Oriented Security and Trust (HOST), 91–101.
3. Glamočanin, O., Coulon, L., Regazzoni, F. & Stojilović, M. (2020). Are cloud FPGAs really vulnerable to power analysis attacks? In Proceedings of the 23rd Conference on Design, Automation and Test in Europe, 1007–1010.
4. Gnad, D. R. E., Oboril, F. & Tahoori, M. B. (2017). Voltage drop-based fault attacks on FPGAs using valid bitstreams. In Proceedings of the 27th International Conference on Field Programmable Logic and Applications (FPL '17), 1–7.
5. Hariharan, R. (2025). Zero trust security in multi-tenant cloud environments. Journal of Information Systems Engineering and Management, 10.
6. Gnad, D. R. E., Rapp, S., Krautter, J. & Tahoori, M. B. (2018). Checking for electrical level security threats in bitstreams for multi-tenant FPGAs. In Proceedings of the International Conference on Field-Programmable Technology (FPT '18), 289–292.
7. Gravelier, J., Dutertre, J. M., Teglia, Y. & Loubet-Moundi, P. (2019). High-speed ring oscillator based sensors for remote side-channel attacks on FPGAs. In Proceedings of the International Conference on Reconfigurable Computing and FPGAs (ReConFig '19).
8. Guo, K., Zeng, S., Yu, J., Wang, Y., Yang, H. & Wang, Y. (2019). A survey of FPGA-based neural network inference accelerator. ACM Transactions on Reconfigurable Technology and Systems, 12(2), Article 2.
9. Hategekimana, F., Mbongue, J. M., Pantho, M. J. H. & Bobda, C. (2018). Secure hardware kernels execution in CPU+FPGA heterogeneous cloud. In Proceedings of the International Conference on Field-Programmable Technology (FPT '18), 182–189.

- 10.** Hategekimana, F., Mbongue, J. M., Pantho, M. J. H. & Bobda, C. (2018). Inheriting software security policies within hardware IP components. In Proceedings of the IEEE 26th Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM '18), 53–56.
- 11.** Hategekimana, F., Nardin, P. & Bobda, C. (2016). Hardware/software isolation and protection architecture for transparent security enforcement in networked devices. In Proceedings of the IEEE Computer Society Annual Symposium on VLSI (ISVLSI '16), 140–145.
- 12.** Hategekimana, F., Whitaker, T. J. L., Pantho, M. J. H. & Bobda, C. (2017). Shielding non-trusted IPs in SoCs. In Proceedings of the 27th International Conference on Field Programmable Logic and Applications (FPL '17), 1–4.
- 13.** Hategekimana, F., Whitaker, T. J. L., Pantho, M. J. H. & Bobda, C. (2017). Secure integration of non-trusted IPs in SoCs. In 2017 Asian Hardware Oriented Security and Trust Symposium (AsianHOST), 103–108.
- 14.** Almorsy, M., Grundy, J. & Ibrahim, A. S. (2012). TOSSMA: A Tenant-Oriented SaaS Security Management Architecture. IEEE Fifth International Conference on Cloud Computing, 1–9.
- 15.** Tsai, W. & Shao, Q. (2011). Role-Based Access-Control Using Reference Ontology in Clouds. Tenth International Symposium on Autonomous Decentralized Systems, 121–128.
- 16.** Kamara, S. & Lauter, K. (2010). Cryptographic cloud storage. In Proceedings of the 14th International Conference on Financial Cryptography and Data Security, 136–149.