Fireblocks

Securing Digital Assets in an Evolving Threat Landscape

The Fireblocks **Defense-in-Depth** Approach to Security

Version 1.0 | Feb 2026

Fireblocks

Table of contents:

1. Introduction: What it Takes to Secure Digital Assets 3

2. The Blockchain Threat Landscape 4

Threat Actors 5

State-Sponsored Actors: The Democratic People's Republic of Korea (DPRK) 5

Organized Cybercrime: Drainer-as-a-Service (DaaS) 6

Unorganized Cybercrime: Opportunistic Exploitation 8

Threat Vectors 9

Crypto Phishing and Wallet Drainers 9

Spear Phishing and Social Engineering 9

API Compromise 10

Privilege Abuse 11

Private Key Compromise 11

Blind Signing 11

Address Poisoning 11

3. Fireblocks' Defense-in-Depth Approach to Security 12

Zero-Trust Architecture 13

Multi-Device Approval 13

Distributed Wallet Infrastructure 14

Policy and Governance Engine 15

Secure Operations Environment 16

Transaction Scanning and DeFi Threat Detection 17

Backed by Certified Security Practices 18

4. How Fireblocks Defenses Are Designed to Mitigate Threats 19

Threat-to-Defense Mapping Matrix 19

Attack Scenario Walkthroughs 20

Scenario 1: Nation-State Attack on Asset Manager Treasury 20

Scenario 2: Wallet Drainer Targeting DeFi Operations 22

Scenario 3: Malicious Insider Threat 23

5. Conclusion: Staying Ahead of Evolving Threats 25

Appendix: Defense-in-Depth Readiness Assessment for Digital Asset Operations 27

Securing Digital Assets

1. What it Takes to Secure Digital Assets

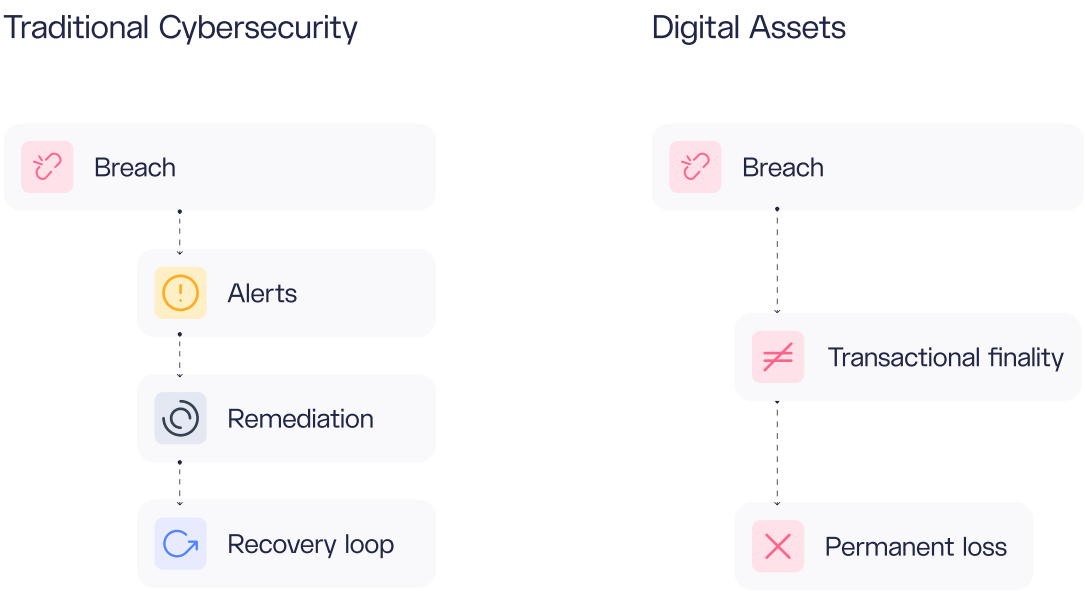
In digital assets, attackers need to succeed only once. When a malicious transaction reaches finality on the blockchain, there's rarely a way to recover funds.

This makes the domain fundamentally different from traditional cybersecurity, where breaches can mean system downtime, leaked sensitive data, or compliance penalties—problems that are serious but often recoverable. With digital assets, the risk is direct and irreversible loss of funds, posing an existential threat to the growing number of businesses managing these assets for their own operations or on behalf of customers.

> ***The primary objective of security in digital assets, therefore, is to protect assets from theft or loss.***

This can be daunting given the ever-evolving threats in the blockchain landscape. But with the right approach that takes into consideration multiple attack vectors and sophisticated actors, it is possible to effectively mitigate threats and better protect funds.

In this paper we outline a comprehensive framework for protecting digital assets against modern threats. We provide an overview of some of the most concerning attack vectors in the space, and show how Fireblocks' **defense-in-depth security** protects your operations even when individual components are compromised.



Threat Landscape

2.

The Blockchain Threat Landscape

$3.4B

Annual crypto losses (2025)

The blockchain ecosystem is one of the most lucrative targets for cyber adversaries, with cryptocurrency-related theft reaching unprecedented levels. In 2025 alone, hackers stole over $3.4 billion worth of cryptocurrency, with stolen totals since 2020 now surpassing $17 billion.[1] As digital asset adoption expands and valuations climb, the space becomes even more attractive to malicious actors.

The threat landscape spans from sophisticated nation-state operations to commoditized services that give even non-technical actors access to wallet-draining tools. It is crucial for institutions in the digital asset space to understand these adversaries and their methods.

1. *Chainalysis, "North Korea Drives Record $2 Billion Crypto Theft Year," December 2025*

Threat Actors

State-Sponsored Actors: The Democratic People's Republic of Korea (DPRK)

North Korea is the most significant nation-state threat to digital asset security, with state-sponsored hacking operations serving as a critical revenue stream for the regime's weapons programs. Since 2017, DPRK-linked actors have stolen approximately $6.75 billion in cryptocurrency. They account for three-quarters of all attacks on crypto platforms, with operations nearly five times larger on average than other threat actors.[2]

$6.75B

Estimated cryptocurrency stolen by DPRK-linked actors since 2017.

If you think your organization is too small to attract North Korean attention, think again. While headline-grabbing breaches like the $1.5 billion Bybit hack dominate the news, Fireblocks Security Research shows many DPRK attacks target smaller platforms for single-digit millions or less. The regime's cyber operations, conducted primarily through the Reconnaissance General Bureau (RGB), cast a wide net. Unlike traditional state-sponsored groups focused on espionage, DPRK operations blend intelligence gathering with direct financial theft, targeting institutions of all sizes.

The Lazarus Group: DPRK's Premier Cyber Weapon

The Lazarus Group, also known as APT38, is the most notorious and capable threat actor targeting the cryptocurrency ecosystem. Operating under DPRK's RGB, Lazarus has evolved from conducting disruptive attacks, such as the 2014 Sony Pictures breach and the 2017 WannaCry ransomware outbreak, to becoming the world's most prolific cryptocurrency thief. Their February 2025 breach of Bybit is currently the largest cryptocurrency heist in history.

2. *Chainalysis, "North Korea Drives Record $2 Billion Crypto Theft Year," December 2025*

Threat Actors

DPRK threat actors employ multiple attack vectors:

Operation Dream Job (active 2019-2020) and Contagious Interview (active since 2023) are North Korean social engineering campaigns targeting employees in defense, aerospace, and cryptocurrency sectors. Using fictitious LinkedIn profiles and fake job offers from prestigious companies, attackers deliver malware through trojanized documents, malicious coding tasks, and compromised repositories during fabricated application and interview processes. These campaigns have successfully compromised major cryptocurrency platforms and industry leaders.

Transaction Manipulation Attacks target private key infrastructure and signing processes. The Bybit attack involved compromising a developer machine to alter what signers saw in the multi-signature wallet UI, causing them to unknowingly approve malicious transactions. Similar transaction manipulation techniques have been documented to steal over $1 billion from industry leaders in recent years.

IT Worker Infiltration involves North Korean operatives using fraudulent identities to obtain remote technical positions at crypto and technology companies. In the best case scenario, the target organization unknowingly funds a sanctioned regime through salary payments. In the worst case, operatives exploit their privileged access for reconnaissance or future attacks.

Once an exploit has succeeded, DPRK actors maintain highly efficient laundering pipelines utilizing mixers, cross-chain bridges, no-KYC exchanges, and Chinese-speaking OTC networks, frequently laundering entire stolen amounts within 48 hours of theft, including operations exceeding hundreds of millions of dollars.

Organized Cybercrime: Drainer-as-a-Service (DaaS)

Organized cybercrime in the digital asset space has become increasingly professionalized and service-oriented. Drainer-as-a-Service (DaaS) operations represent the most visible example of this shift by packaging sophisticated theft tools for non-technical criminals. In this model, DaaS developers create wallet-draining kits and license them to affiliates—the criminals who deploy phishing campaigns and execute attacks—on a revenue-share basis. Affiliates retain the majority of stolen assets while developers collect a commission, enabling theft at scale by actors who lack specialized skills.

Inferno Drainer: A Case Study in DaaS Evolution

Inferno Drainer is the most notorious example of the DaaS model. It provided turnkey phishing infrastructure that enabled non-technical affiliates to deploy high-quality phishing pages impersonating over 100 legitimate cryptocurrency brands, including Seaport, WalletConnect, and Coinbase. When victims interacted with these fake interfaces, they unknowingly authorized fraudulent transactions that drained their wallets, with stolen funds split between affiliates and the drainer developers.

The revenue-share model proved extraordinarily successful, enabling theft from over 167,000 victims with total losses exceeding $250 million across more than 16,000 unique phishing domains. [3] [4]

The DaaS model has spawned numerous competitors and successors including Monkey Drainer, Venom Drainer, Angel Drainer, Pink Drainer, and MS Drainer, creating an ecosystem where service providers compete on features, reliability, and revenue terms. Product sophistication has become a key differentiator: leading services offer support for 30+ EVM-compatible networks, built-in anti-analysis measures to prevent researchers from viewing source code, and continuous innovation in attack techniques and evasion capabilities—operating like legitimate SaaS businesses competing for market share.

3. *Group-IB, "Crypto Wallet Drainers," 2024*
4. *Check Point Research, "Inferno Drainer Reloaded," January 2025*

Threat Actors

Unorganized Cybercrime: Opportunistic Exploitation

Beyond state-sponsored operations and organized DaaS ecosystems there is a substantial threat from individual actors and loosely-coordinated groups exploiting vulnerabilities opportunistically. These threat actors lack the resources and sophistication of nation-state operators but compensate through volume and adaptability.

This category encompasses individual hackers and small groups exploiting various vulnerabilities in the digital asset ecosystem. Indictments that illustrate the varied nature of these threats include:

2025 Feb | **Liquidity Pool Manipulation**: A hacker exploited DeFi protocols through manipulative trading, draining $65 million from liquidity pools. [5]

2024 Nov | **Multi-Year Phishing Campaign**: Five individuals ran a multi-year phishing campaign spoofing authentication portals, stealing millions in cryptocurrency and compromising confidential data from tech companies. [6]

2024 Aug | **Social Engineering Scheme**: Criminal associates impersonated tech support staff to steal authentication credentials and obtain 4,100 Bitcoin ($263 million) from a Washington DC victim. [7]

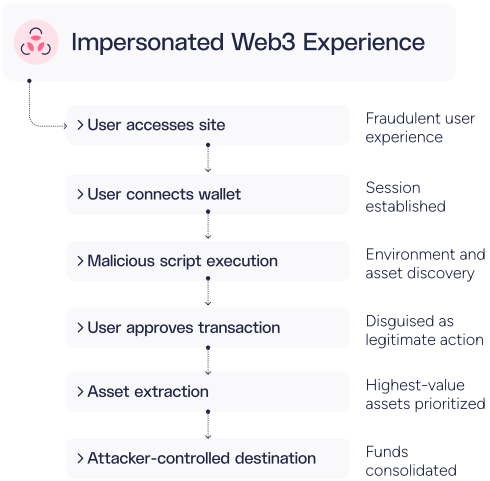
2024 May | **Transaction Validation Manipulation**: Two MIT-educated brothers exploited Ethereum transaction validation protocols to steal $25 million in 12 seconds. [8]

2022 July | **Flash Loan Exploit**: A former security engineer exploited DeFi smart contracts via flash loans, stealing over $12 million. [9]

Beyond external attackers, **malicious insiders** including employees, contractors, or partners with legitimate access can exploit their privileged positions to steal private keys, manipulate transaction signing processes, or disable security controls. The cryptocurrency industry's rapid growth and competitive hiring environment can result in insufficient background verification, while the high value of accessible assets creates significant temptation.

5. *U.S. Department of Justice, "Canadian National Charged with Stealing $65 Million in Cryptocurrency from DeFi Protocols," February 2025*
6. *U.S. Department of Justice, "5 Defendants Charged Federally with Running Scheme that Targeted Victim Companies via Phishing Text Messages," November 2024*
7. *TRM Labs, "DOJ Uses Organized Crime Statute in $263 Million Cryptocurrency Theft," May 2025*
8. *U.S. Department of Justice, "Two Brothers Arrested for Attacking Ethereum Blockchain and Stealing $25M in Cryptocurrency," May 2024*
9. *U.S. Department of Justice, "Former Security Engineer Sentenced to Three Years in Prison for Hacking Two Decentralized Cryptocurrency Exchanges," July 2024*

Threat Vectors



Crypto Phishing and Wallet Drainers

Wallet drainers are malicious scripts designed to siphon assets from cryptocurrency wallets by tricking users into authorizing fraudulent transactions. Attackers deploy phishing sites that impersonate legitimate web3 services, token airdrops, or NFT mints, prompting victims to connect their wallets and sign malicious transactions. Drainer scripts automatically identify and steal the wallet's most valuable assets. The proliferation of DaaS platforms has made these attacks accessible to low-skilled operators, establishing wallet drainers as one of the most pervasive and dangerous threat vectors in DeFi.

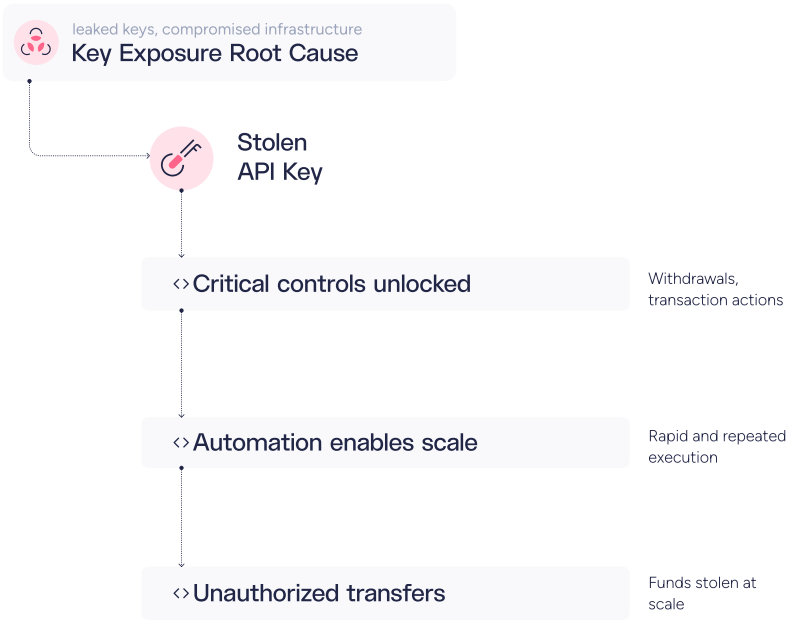
Spear Phishing and Social Engineering

Unlike broad phishing campaigns, spear phishing involves highly targeted attacks against specific individuals with privileged access. Attackers conduct extensive reconnaissance to craft personalized lures, such as fake job offers, investment opportunities, or partnership proposals tailored to the target's role and interests. These campaigns often involve prolonged engagement through LinkedIn, email, phone calls, videoconference calls, or messaging platforms to build trust before delivering malware or extracting credentials. Contagious Interview exemplifies this approach, successfully compromising employees at defense contractors and cryptocurrency firms through fabricated recruitment processes.

> *An evolution of this tactic involves fake videoconference meetings. Attackers compromise Telegram accounts to impersonate trusted contacts, then guide victims to fake Zoom or Teams calls using pre-recorded footage of real individuals from previous hacks or public sources. During the call, the attacker claims audio issues and sends a "patch file" or requests an SDK update to "fix" the problem, which actually installs Remote Access Trojan (RAT) malware granting complete system control. This method has proven highly effective, stealing over $300 million from cryptocurrency executives.*[10]

API Compromise

API compromise, particularly through stolen API keys, has become one of the most common attack vectors against cryptocurrency platforms. Attackers compromise developer machines, internal systems, or legitimate users to steal API credentials that control critical functions such as withdrawals, transaction signing, or account management. With these stolen keys, attackers can take over transaction management systems, identify vulnerabilities in authorization logic, and execute unauthorized fund transfers at scale, often bypassing user-facing security controls entirely. The automated nature of API access enables rapid, large-scale theft once credentials are obtained.



10. *Crypto.news, "North Korean 'Fake Zoom' Hustle Drains $300m from Crypto Execs' Wallets," December 2025*

Threat Vectors

Privilege Abuse

Privilege abuse occurs when individuals with legitimate system access such as employees, contractors, or compromised administrator accounts exploit their permissions for unauthorized purposes. This may include direct theft of funds or private keys, manipulation of transaction approval processes, disabling of security controls or monitoring systems, and theft of sensitive customer or operational data.

Private Key Compromise

Private key compromise remains the most consequential attack vector, as control of private keys grants complete authority over associated assets. Attackers target keys through malware, phishing for seed phrases, exploitation of insecure key storage, or compromise of key management infrastructure.

Blind Signing

Blind signing occurs when users or operators approve transactions without full visibility into what they are authorizing. Attackers exploit this by manipulating transaction data, compromising signing interfaces, or presenting malicious transactions through legitimate-appearing workflows.

Address Poisoning

Address poisoning exploits user reliance on transaction history for address verification. Attackers monitor the blockchain for active wallets then send negligible transactions from addresses that visually resemble the victim's frequently-used addresses, matching the first and last characters while differing in the middle. When victims copy addresses from their transaction history without careful verification, they inadvertently send funds to the attackers' lookalike address.

Defense-in-Depth

3. Fireblocks' Defense-in-Depth Approach to Security

Fireblocks was founded by cybersecurity veterans who spent years defending critical nation-state infrastructure. From day one, the platform was built with an adversarial mindset, anticipating sophisticated attackers would probe every layer and that no single security control could be considered unbreakable. This founding philosophy remains core to Fireblocks' solutions as the platform evolves to meet increasingly complex and diverse institutional requirements.

As our customers' activities have grown from securing static treasury holdings to managing high-frequency trading operations, complex DeFi strategies, tokenization programs, and cross-border payment flows, our architecture has evolved in parallel. What began as a focus on private key protection through multi-party computation (MPC) has expanded into a comprehensive framework. Each new capability is designed as an integrated layer within a holistic security system, where multiple independent controls work together to help prevent, detect, and contain threats.

The result is a multi-layered security architecture where the compromise of a single component should not by itself enable unauthorized access to funds. In an environment where nation-state actors, organized crime networks, and malicious opportunists pose persistent threats, organizations need security systems designed to withstand sophisticated, multi-vector attacks.

Defense-in-Depth

Zero-Trust Architecture

Fireblocks' platform is designed with zero-trust principles as a foundation: it denies implicit trust to internal systems, requiring explicit validation for all authentication attempts. From development and deployment, to data storage, to transaction verification and authorization, critical operations execute within hardware-isolated secure enclaves to help protect cryptographic material, algorithms, and sensitive code from both external attackers and malicious insiders.

Key capabilities:

Cryptographic enforcement
Sensitive operations (transaction initiation, approval, signing) are cryptographically enforced to originate only from the user's trusted (non-Fireblocks) environment

Hardware isolation
Critical operations execute in hardware-isolated environments (SGX/Nitro enclaves and other TEEs), protecting MPC key shares, policy enforcement, transaction serialization, and operation approval flows

Attestation
End-to-end attestation verifies that authorized code runs within secure enclaves (TEEs)

Inter-module authentication
Inter-module communication is cryptographically authenticate

Zero-trust DevOps
Zero-trust principles applied to privileged back-end operations, including code deployment, updates, ongoing system administration, and troubleshooting

Encrypted enclave-protected storage
Sensitive data (MPC shares, credentials, secrets) and integrity-critical data (whitelisted addresses, approval keys, policy rules) are stored in encrypted databases running inside secure enclaves, accessible only to attestated services

Distributed data verification
Co-signers independently verify cryptographically signed transaction data and state across the entire flow, helping to prevent single-point data manipulation and support transaction immutability

Zero-trust verification model
Signing callbacks enable customers to cryptographically verify that both the signing process and transaction data match their original requests, eliminating reliance on trust in Fireblocks' infrastructure alone

Multi-Device Approval

Sensitive operations require multi-device and user approval, separating transaction initiation from authorization and signing to help prevent single point of compromise. Approval and signing devices leverage trusted execution environments (TEEs), biometric authentication, hardware-backed authenticators such as YubiKey and PIN code, and system integrity checks to help ensure that transaction approvals and signing can be performed only by authorized users on uncompromised devices and that cryptographic signing operations are executed in a tamper-resistant environment.

Multiple device types are available for sensitive approval operations, providing device diversification to reduce platform-specific risks. Approval and signing interfaces clearly display human-readable transaction data, mitigating threats of blind signing, data manipulation, or address manipulation.

Defense-in-Depth

Multiple device types are available for sensitive approval operations, providing device diversification to reduce platform-specific risks. Approval and signing interfaces clearly display human-readable transaction data, mitigating threats of blind signing, data manipulation, or address manipulation.

Key capabilities:

Transaction initiation
Transactions can be initiated via web console, API, or integrations

Independent approval
Approval occurs independently via Fireblocks' mobile app with biometric authentication (FaceID/TouchID) and PIN code, or through customer-managed co-signer infrastructure that enables integration of custom authorization logic via secure callback mechanisms

Device integrity verification
System checks help verify that devices are not jailbroken or compromised before approval

Payload verification
Transaction payloads are cryptographically signed and verified end-to-end across devices and services

Approval and signing key isolation
Approval and signing keys are stored in hardware-isolated secure enclaves (TEEs)

Diversified approval device support
Support for iOS devices, Android devices, Intel SGX, AWS Nitro, GCP Confidential Spaces, and others, with diversified operating systems and TEE options

Transaction clarity
Transaction details are displayed to approvers in clear, contextual, human-readable language with built-in transaction simulation

Multi-factor authentication
PIN code, biometrics, and optional YubiKey help protect against unauthorized mobile app access

Push notification system
Real-time push notifications enable flexible approval and signing from any authorized location

Asymmetric API authentication
API authentication uses asymmetric cryptography, minimizing the risk of API secret exposure and unauthorized fund movements

Distributed Wallet Infrastructure

By default, MPC-based private keys are distributed across multiple isolated environments with cryptographic guarantees to mitigate single points of compromise while giving clients control of their assets. Key material should not exist whole at any point, whether during generation, storage, or signing, to help prevent any single component from resulting in key extraction or unauthorized transaction signing.

Key capabilities:

Distributed key generation
MPC-CMP protocol generates and stores private keys as shares that never combine

HD wallet architecture
Hierarchical Deterministic (HD) wallets mitigate seed phrase vulnerability

Defense-in-Depth

Flexible storage models
Support for hot, warm, and cold storage configurations

Key refresh and rotation
Optional automated key refresh and rotation capabilities

Blockchain agnostic
Universal signature generation supporting ECDSA and EdDSA algorithms

Peer-reviewed protocol
Open-source MPC-CMP protocol audited by leading security firms

Flexible deployment models
Customers can deploy MPC key shares across a hybrid SaaS model with shares distributed between Fireblocks and customer infrastructure, self-hosted private cloud (SGX), self-managed HSM environments, or custom in-house signing infrastructure

Policy and Governance Engine

The Fireblocks Policy Engine enforces granular transaction authorization rules across digital asset operations, creating separation of duties that helps prevent internal collusion and cryptographically enforces approval quorums. Policies themselves are designed to operate in a zero-trust manner, are protected within secure enclaves (TEEs), and require a quorum for modification. Maker-checker quorums are available for all sensitive operations.

Key capabilities:

Granular transaction controls
Rules based on source, destination, asset type, amount, and onchain operations including smart contract methods (EVM) and program calls (Solana)

Multi-user approval workflows
Customizable quorums and role-based permissions for transaction authorization

Custom callbacks for policy and verification
Customer-driven co-signer integrations for independent transaction verification and extended policy enforcement prior to signing

Risk-based automation
Automated and manual approval paths triggered by configurable risk thresholds

Admin quorum protection
Multi-administrator approval requirements for policy modifications and configuration changes

Address whitelisting
Interactions restricted to pre-approved addresses only

API permission scoping
Programmatic access limited to explicitly defined operations per API user

Compliance integration
Native integration with compliance screening for AML, KYT, and sanctions checks

Multi-layer enforcement
Policy validation occurs at initiation, validation, and signing stages

Comprehensive audit trail
Comprehensive logging of policy decisions, approvals, and authorization events with real-time streaming to external logging and analytics platforms

Defense-in-Depth

Policies are the first line of defense

As enterprises scale their digital asset operations, manual security reviews can struggle to keep pace with frequent updates to settings and workflows. Policy drift can accumulate unnoticed until an incident occurs.

Fireblocks' Security Research shows ***nearly all digital asset theft incidents*** *stem from misconfigured policies leading to unintentional authorizations. Policies are the most critical line of defense, and a single overlooked setting can expose an organization to significant risk.*

Fireblocks Security Posture Management (FSPM) is the first security posture management solution purpose-built for digital assets. FSPM acts as a dedicated security advisor, continuously monitoring configurations and providing clear guidance on how to remediate issues before they become incidents.

Secure Operations Environment

Fireblocks provides secure-by-design pathways for digital asset operations, enabling direct interactions with CeFi and DeFi applications without requiring external connections that increase attack surface. Native integrations for staking and token swaps allow users to access institutional validators and decentralized exchanges directly through the Fireblocks platform, protecting against risks associated with browser extensions and unvetted dApps.

The Fireblocks Network extends this secure interaction model to counterparty transfers, providing authenticated channels between institutional participants. Encrypted tunnels verify destination addresses at the source, protecting against man-in-the-middle attacks, address poisoning, and human error in address handling. The Network connects over 2,400 institutions, including exchanges, banks, and liquidity providers, facilitating secure, direct transfers without the need for manual address exchange.

Key capabilities:

Native financial operations
Integrated DeFi operations, including staking and swaps, reduce exposure to external connection risks and unvetted applications

Direct counterparty connections
Elimination of manual address exchange between network participants

End-to-end encryption
Encrypted tunnels protect deposit address queries throughout transmission

Hardware-terminated connections
Secure enclaves utilized on both sending and receiving endpoints

Cryptographic attestation
Source attestation cryptographically proves destination address authenticity

Automatic address rotation
Option to automatically rotate UTXO addresses to preserve transaction privacy

Defense-in-Depth

Fail-close architecture
Automated blocking of transfers upon detection of attacks or anomalies

Network profile management
Institutional counterparty management and deposit routing without direct address sharing

Attack vector mitigation
Protection against clipboard malware, phishing, and address poisoning attacks

Transaction Scanning and DeFi Threat Detection

Real-time threat intelligence and transaction simulation protect against malicious apps, compromised contracts, and hidden exploits. Transaction clarity decodes complex smart contract interactions into human-readable actions, enabling informed approval decisions rather than blind signing.

Key capabilities:

Transaction simulation
Pre-execution simulation displays exact transaction outcomes before approval

Smart contract analysis
Automated identification of suspicious or malicious contract behavior

Real-time threat intelligence
Integration with blockchain security vendors and OSINT for recent threat data

Malicious dApp protection
Real-time risk analysis and blocking of dangerous dApp connections

Transaction decoding
Raw call data translated into human-readable actions and operations

Typed message interpretation
Signature requests decoded to reveal true intent (permits, approvals, transfers)

Permission risk detection
Identification of unlimited token approvals and high-risk permission grants

DeFi-specific policies
Granular controls for contract interactions, asset approval operations, and dApp connectivity

Anomaly detection
Automated flagging of unusual transaction patterns or suspicious destinations

Defense-in-Depth

Backed by Certified Security Practices

Fireblocks' defense-in-depth architecture is independently validated through rigorous third-party certifications: SOC 2 Type II with zero material findings, ISO 27001/27017/27018/22301, and the industry's gold standard C4 CCSS Qualified Service Provider Level 3. Fireblocks was the world's first platform to achieve this certification.

These certifications are backed by continuous operational vigilance. We maintain a 24/7 Security Operations Center (SOC), with analysts distributed across the US, EMEA, and APAC, providing round-the-clock monitoring and real-time threat response.

This combination of rigorous third-party validation and continuous operational monitoring enables our defenses to evolve to address emerging threats.

Fireblocks Defenses

4. How Fireblocks Defenses Are Designed to Mitigate Threats

The following section maps Fireblocks' security controls to specific threat vectors through a comprehensive coverage matrix and detailed attack scenarios, demonstrating how our defense-in-depth architecture uses layers of controls to help defend against unauthorized fund movements.

Threat-to-Defense Mapping Matrix

The matrix below shows how these security layers work in concert to mitigate threat vectors. Crypto phishing attacks must overcome multi-device approval, policy engine restrictions, and Transaction scanning and DeFi threat detection. Private key compromise attacks, which are a primary vector for nation-state actors, face distributed key shares that never exist whole, hardware-isolated enclaves protecting those shares, and multi-device approval requirements. These identified threat vectors face at least three independent security layers, with most confronting four or more.

Threat Vector	Zero-Trust Architecture	Multi-Device Approval	Distributed Wallet Infrastructure	Policy & Governance Engine	Secure Operations Environment	Transaction Scanning & DeFi Threat Detection
Crypto Phishing & Wallet Drainers	▰	▲	▰	▲	▰	▲
Spear Phishing & Social Engineering	▲	▲	▰	▲	▲	▰
API Compromise	▰	▰	—	▲	—	—
Privilege Abuse	▲	▲	—	▲	—	▰
Private Key Compromise	▲	▰	▲	▰	—	—
Blind Signing	▰	▲	—	▲	▰	▲
Address Poisoning	—	▰	—	▲	▲	▰

▲ Primary Defense: Core control specifically designed to prevent threat

▰ Secondary Defense: Provides additional protection or detection capability

— Indirectly applicable to this vector

Fireblocks Defenses

Attack Scenario Walkthroughs

Sophisticated threat actors employ multi-stage attacks that target different aspects of digital asset operations. Fireblocks' defense-in-depth architecture is designed to protect against these complex attacks by significantly increasing the difficulty and cost for attackers at each stage.

Scenario 1: Nation-State Attack on Asset Manager Treasury

Threat Context

An asset management firm's treasury wallet is targeted by a sophisticated threat actor employing tactics consistent with DPRK-affiliated groups. The attackers aim to manipulate the transaction signing process to redirect funds to attacker-controlled addresses, mirroring the techniques used in the Bybit breach where attackers compromised signing infrastructure to display fraudulent transaction details to approvers.

Hypothetical Attack Flow:

Stage 1: Initial Compromise
Attackers conduct reconnaissance on the firm's employees, identifying operations staff with transaction approval responsibilities. Through a targeted spear phishing campaign, they deliver malware to a senior operator's workstation.

Stage 2: Infrastructure Manipulation
With persistent access to the compromised workstation, attackers monitor transaction workflows and identify the signing process. They deploy tools to intercept and modify transaction data displayed on the compromised machine, similar to the Bybit attack where the Safe{Wallet} interface was manipulated to show legitimate-appearing transaction details while the underlying payload contained malicious instructions.

Stage 3: Execution Attempt
The attackers wait for a routine large transfer and attempt to substitute the destination address with an attacker-controlled wallet. The compromised workstation displays the expected recipient address to the initiating operator while the actual transaction payload targets the malicious address.

Fireblocks Defenses

Fireblocks Defense Layers Activated

Layer	Control	Action
Multi-Device Approval	Independent mobile verification	Transaction details rendered on separate, uncompromised mobile devices show the true destination address, revealing discrepancy with workstation display
Transaction Scanning & DeFi Threat Detection	Transaction decoding	Mobile app displays decoded transaction details including actual recipient, amount, and contract interactions; approver sees true transaction intent
Policy & Governance Engine	Destination restrictions	Transaction to non-whitelisted address automatically blocked; requires address to be pre-approved through separate admin quorum process
Policy & Governance Engine	Amount thresholds	Large transfer triggers enhanced approval requirements; multiple independent approvers required, each viewing transaction on separate trusted devices
Zero-Trust Architecture	IP whitelisting	API calls and transaction initiations restricted to pre-approved IP ranges; attempts from unauthorized network locations blocked

Outcome Objective

The hypothetical attack should fail at multiple independent points, even with full control of an operator's workstation. The multi-device approval architecture is designed to render transaction details on trusted, hardware-secured devices outside attacker control. Policy controls provide additional layers to help block unauthorized destinations and enforce approval quorums, while IP restrictions reduce the attack surface. Implementing the defense-in-depth approach helps prevent a compromise of a single component, including an employee workstation, from resulting in unauthorized fund transfers.

Fireblocks Defenses

Scenario 2: Wallet Drainer Targeting DeFi Operations

Threat Context

A digital asset fund employs traders who interact with DeFi protocols as part of their investment strategy. A threat actor targets these operations by deploying a wallet drainer designed to harvest unlimited token approvals and permit signatures.

Hypothetical Attack Flow:

The attacker deploys a fraudulent site mimicking a legitimate DeFi protocol, embedding malicious JavaScript that manipulates the interface to deceive users into signing:

> Unlimited ERC-20 token approvals to attacker-controlled contracts

> Offchain permit signatures granting spending authority without onchain approval transactions

Fireblocks Defense Layers Activated

Layer	Control	Action
Transaction Scanning & DeFi Threat Detection	dApp scanning	dApp connection is scanned to identify compromised malicious indicators and known drainer infrastructure, alerting user before any transaction is initiated
Transaction Scanning & DeFi Threat Detection	Transaction scanning	Behavioral and technical heuristics evaluate pending transactions for anomalies including interaction with flagged addresses and deviation from expected contract behavior; user alerted to specific risks before signature
Policy & Governance Engine	Address whitelisting	Transaction to non-whitelisted addresses automatically blocked; attacker-controlled approval target must first pass through separate approval and whitelisting process
Policy & Governance Engine	dApp access policy	Access restricted to pre-approved dApps only, limiting the attack vector at the interface layer
Policy & Governance Engine	Typed message policies	Typed message initiation and signing restricted to specific users and conditions only

Fireblocks Defenses

Outcome Objective

The hypothetical attack should fail at multiple points before asset theft can occur. Real-time threat intelligence is intended to help identify the impersonating dApp upon connection attempt, alerting users before transaction initiation. Should a user proceed, transaction-level analysis should help detect the anomalous transaction patterns and flag interaction with suspicious destination addresses. Policy controls provide additional defensive layers: dApp access policy prevents connection to unapproved or fraudulent sites, address whitelisting blocks transactions to non-approved contracts, and typed message policies enforce additional approval requirements for offchain signatures. This defense-in-depth approach helps prevent a convincing protocol impersonation from resulting in unauthorized transactions and asset theft.

Scenario 3: Malicious Insider Threat

Threat Context

A senior employee authorized to approve transactions decides to exploit their position to steal funds. The insider holds legitimate credentials and is an authorized approver within the organization's transaction workflow. This scenario represents one of the most challenging threats: an adversary operating with valid authorization, institutional knowledge, and an understanding of operational procedures within the organization.

Hypothetical Attack Flow:

The insider attempts to execute an unauthorized withdrawal to an external wallet under their control. As a legitimate approver, they expect their authorization to be sufficient, or attempt to manipulate timing and procedures to complete a transaction without proper oversight.

Fireblocks Defenses

Fireblocks Defense Layers Activated

Layer	Control	Action
Policy & Governance Engine	Approval quorums	Transaction authorization requires multiple independent approvers; the insider's legitimate approver status provides only one of the required signatures
Policy & Governance Engine	Destination whitelisting	Transfers restricted to pre-approved addresses; adding new destinations requires separate admin quorum that prevents unilateral approval
Policy & Governance Engine	Role-based permissions	Approval authority scoped to specific transaction types and amounts; elevated withdrawals require additional independent approvers
Multi-Device Approval	Independent device verification	Each approver authorizes on their own hardware-secured device; insider cannot access or impersonate other approvers' sessions
Multi-Device Approval	Biometric authentication	Approval requires biometric verification on enrolled device; credential knowledge insufficient without physical device and biometric presentation

Outcome Objective

The hypothetical attack should fail despite the insider holding legitimate approval authority. Policy restrictions are intended to prevent a single individual from unilaterally authorizing transactions. The insider can initiate a transaction and provide one approval, but should not be able to complete the required quorum or add unauthorized destinations without collusion from independent parties outside their control.

Fireblocks' security layers provide overlapping controls that help protect against sophisticated threats, but their effectiveness depends on proper implementation and ongoing maintenance. Strong security posture requires continuous monitoring of configurations, regular review of policy settings, and prompt remediation of gaps or misalignments as operations evolve. For organizations seeking additional assurance that security controls remain properly configured at scale, Fireblocks Security Posture Management (FSPM) provides automated monitoring, helping to identify configuration drift, policy weaknesses, and potential vulnerabilities before they can be exploited.

Conclusion: Staying Ahead of Evolving Threats

The digital asset security landscape continues to evolve. Nation-state actors conduct sustained campaigns with advanced technical capabilities and long-term persistence as a strategic revenue generation engine. Organized crime has professionalized, with criminal networks operating like legitimate SaaS businesses. Meanwhile, artificial intelligence is amplifying threat sophistication, enabling unprecedented phishing personalization through deepfake impersonations, AI-assisted social engineering, and automated vulnerability discovery that evolves faster than traditional defenses can adapt.

In this environment, legacy security approaches designed for simpler times and more limited use cases will not suffice. What remains constant, however, are the fundamental principles that enable organizations to stay ahead:

Defense in depth is non-negotiable. No single security technology can protect against the full spectrum of modern threats. Comprehensive security requires multiple independent layers that maintain protection even when individual controls are compromised.

Assume breach, architect for resilience. Security systems must be designed with the assumption that attackers will gain some level of access. The question is not whether defenses will be tested, but whether the architecture contains and limits damage when they are.

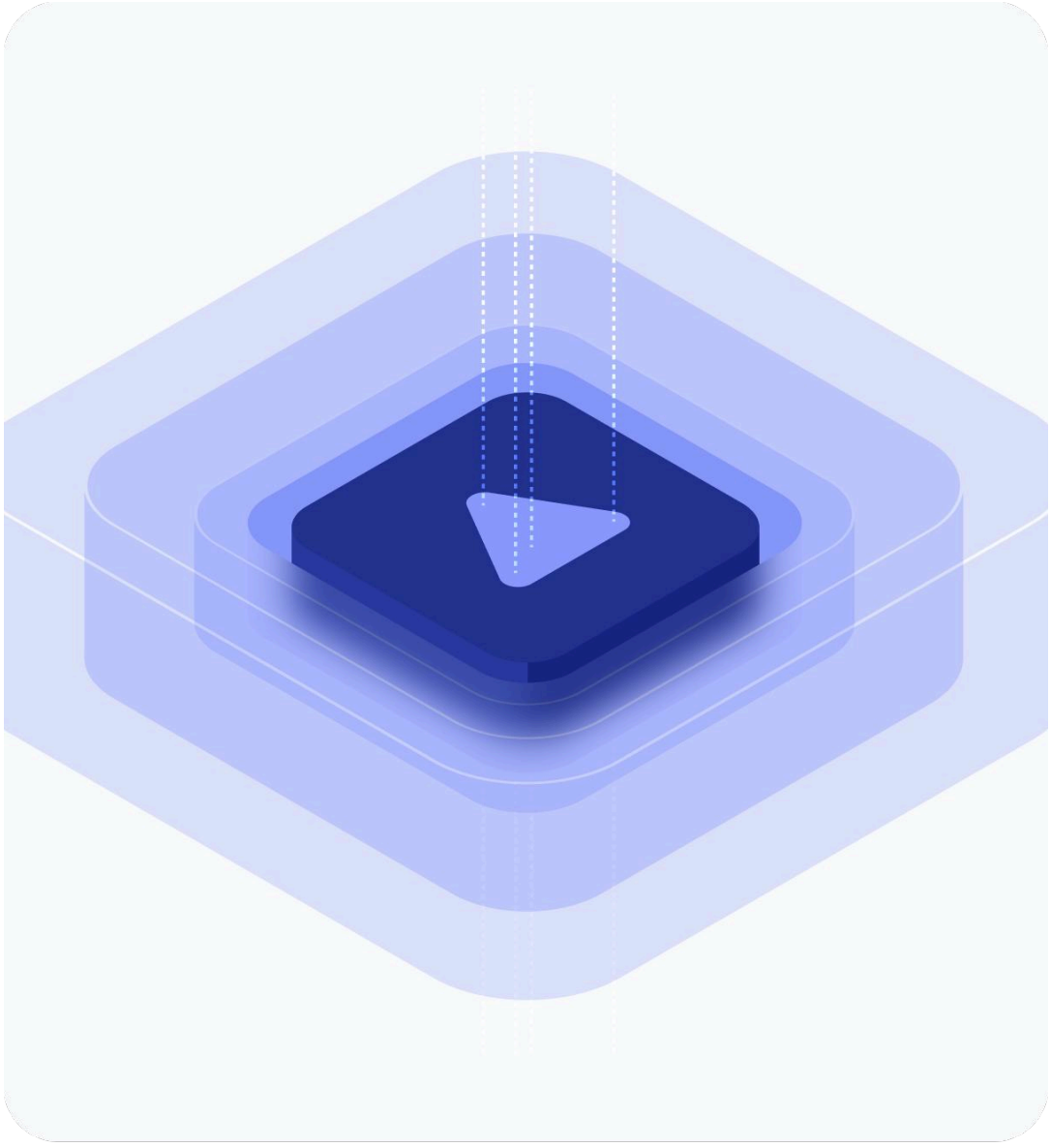
People, process, and technology must align. Technical controls are effective only when combined with sound operational security, including proper access management, personnel vetting, continuous monitoring, and incident response capabilities. The most sophisticated cryptography cannot protect against social engineering if approval processes lack adequate scrutiny.

Transparency enables trust and detection. In an environment where threats constantly evolve, the ability to audit, monitor, and verify security controls in real-time is essential. Transaction clarity prevents blind signing, comprehensive logging enables forensic analysis, and continuous anomaly detection identifies suspicious patterns before they result in losses.

Security must enable operations, not obstruct them. A secure system that cannot support business requirements will be circumvented. Effective security architectures balance protection with operational efficiency, enabling high-frequency trading, automated workflows, and complex DeFi strategies while maintaining strong controls.

Conclusion

Organizations entering or expanding digital asset operations face well-resourced attackers that continuously adapt their tactics. Success requires comprehensive defense-in-depth architectures that address threats from multiple angles through coordinated, overlapping controls. Organizations that implement robust security frameworks and maintain them as threats evolve can confidently leverage the transformative potential of digital assets while protecting against even the most sophisticated adversaries.



Appendix

Defense-in-Depth Readiness Assessment for Digital Asset Operations

Use this assessment to evaluate whether your security architecture can withstand modern digital asset threats.

This is not a compliance checklist. It is designed to test whether your controls could remain effective under realistic failure conditions, including compromised endpoints, stolen credentials, malicious insiders, and manipulated transaction interfaces.

Answer each question using an assume-breach mindset: if a single user device, admin account, cloud environment, or integration were compromised, would independent controls still prevent unauthorized fund movement? “Yes” should mean the control is enforced by architecture and cryptography, not dependent on procedure, trust, or manual review.

Any uncertainty is a signal to investigate, as attacks in digital assets typically exploit the gaps between people, process, and technology.

Transaction Intent & Initiation Controls

Protect against phishing, UI manipulation, API compromise

- [ ] Does your system require independent verification for transactions, even if a workstation, browser, or CI/CD pipeline is compromised?
- [ ] Are transaction initiation channels (UI, API, integrations) cryptographically authenticated to a trusted execution environment?
- [ ] Are API credentials:
  - Least-privileged by default?
  - Bound to IP ranges and environment context?
  - Unable to trigger withdrawals without additional independent approval?
- [ ] Does your system prevent display manipulation attacks that show operators different data than what is signed?

If transaction intent can be altered upstream of signing without independent verification, you are exposed to Bybit-style attacks.

Transaction Approval & Human Verification

Protect against blind signing, social engineering, insider coercion

- [ ] Do approvers view transaction details on independent, hardware-secured devices?
- [ ] Are approval devices isolated from:
  - The initiating workstation?
  - Corporate endpoint management systems?
- [ ] Is the transaction data:
  - Decoded into human-readable actions?
  - Simulated against the current blockchain state?
- [ ] Are unlimited approvals, hidden token transfers, and signature requests that grant unintended spending rights flagged before signing?

If approvers rely on a single UI or sign opaque payloads, this creates a blind-signing risk, even with multi-sig.

Private Key and Wallet Architecture

Protect against private key compromise, insider access, infrastructure attacks

- [ ] Are private keys prevented from ever existing in full at any point, including generation, storage, signing, and backup?
- [ ] Are key shares:
  - Distributed across independent fault domains?
  - Protected inside hardware-isolated secure enclaves?
- [ ] Are cloud administrators, DevOps staff, and SREs prevented from:
  - Accessing key material?
  - Influencing signing logic?
- [ ] Can key shares be rotated or refreshed without changing wallet addresses?

If any individual, machine, or cloud environment can reconstruct keys, you have a structural single point of failure.

Policy, Governance, and Change Management

Protect against insider threat, collusion, privilege abuse

- [ ] Are transaction policies:
  - Cryptographically enforced?
  - Protected from unilateral admin modification?
- [ ] Do policy changes require:
  - Multi-party admin quorum?
  - Independent device approval?
- [ ] Is there a complete, immutable audit trail for:
  - Policy changes?
  - Approval overrides?
  - Admin actions?
- [ ] Are administrators prevented from unilaterally reducing approval thresholds or whitelisting destinations?

Insider risk is not solved by trust. It is solved by enforced separation of authority.

Onchain Interaction and DeFi Risk Controls

Protect against wallet drainers, malicious contracts, and address poisoning

- [ ] Are smart contract interactions restricted to pre-approved, reviewed contracts?
- [ ] Can transactions be simulated and decoded before approval, not only after execution?
- [ ] Are high-risk patterns explicitly flagged, including unlimited token approvals and signature requests that grant unintended spending rights?
- [ ] Are all deposit addresses exchanged through authenticated channels, eliminating manual copy-paste?
- [ ] Are address poisoning and clipboard malware attacks prevented from routing funds to attacker-controlled wallets?

Allowing users to interact freely with arbitrary contracts or addresses keeps DaaS and address-poisoning viable.

Resilience, Detection, Blast-Radius Control

Protects against major damage if breach occurs

- [ ] If one security layer fails (endpoint, admin account, cloud region), do additional independent controls prevent unauthorized fund movement?
- [ ] Can you:
  - Detect anomalous behavior in real time?
  - Attribute actions to specific identities and devices?
- [ ] Are signing components geographically and operationally isolated?
- [ ] Can operations continue if a subset of infrastructure is disabled or compromised?

Resilience matters as much as prevention when attackers are persistent and well-resourced.

Organizational Readiness and Accountability

Aligns people and process with technology

- [ ] Are roles clearly separated between:
  - Initiators?
  - Approvers?
  - Policy administrators?
  - Infrastructure operators?
- [ ] Are high-risk actions:
  - Rare by design?
  - Observable by default?
- [ ] Do incident response plans explicitly cover onchain theft scenarios, not just IT breaches?

Security technology requires organizational discipline and accountability to be most effective.

If these questions reveal any uncertainty, ambiguity, or reliance on trust rather than enforced controls, your organization could be exposed to the same failure modes used in recent billion-dollar attacks.

Fireblocks' defense-in-depth architecture is designed to address these gaps—not through point solutions, but through layered, cryptographically-enforced controls.

This architecture is trusted in production by over 2,400 enterprises, has secured more than $10 trillion in digital asset transactions, and protects over 550 million wallets globally.

To learn more about protecting your digital assets, reach out to info@fireblocks.com, or visit Fireblocks.com.