Canary Wallets: Early Warning of a Leaked Seed

November 26, 2026 · Document #440 · bip39-phrase.com

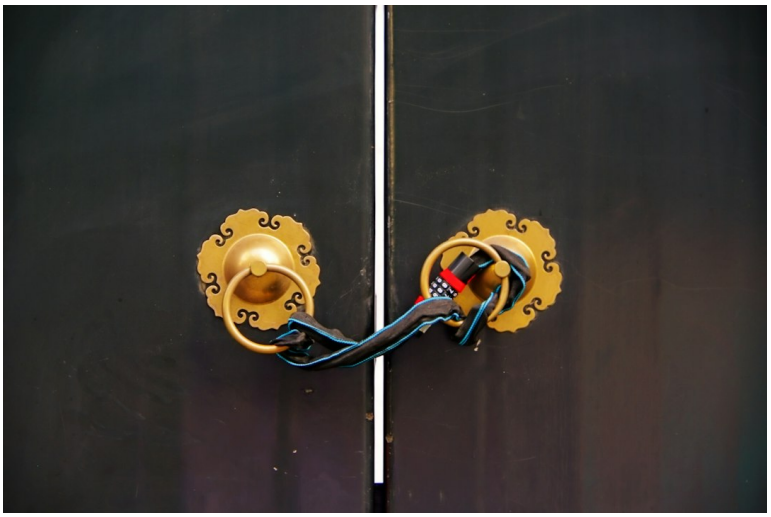

Photo by Chris DUNN on Unsplash

A small canary balance can expose unauthorized spending after a seed leak, but cannot guarantee warning before protected funds are targeted.

Contents

How the canary works	2
Set up a useful tripwire	2
What the warning does not prove	3
Respond to an unexpected spend	3
FAQ	3
Sources	4
Example sequences from the BIP39 wordlist	4
Reading a BIP39 phrase	5
See also	6

A small balance in the wallet derived from your BIP39 words with an empty passphrase can act as a tripwire: an unexpected withdrawal may reveal that someone obtained the words. Funds derived using a separate BIP39 passphrase might remain inaccessible, but the warning is neither guaranteed nor necessarily early enough to help.

How the canary works

BIP39 derives a 512-bit seed using PBKDF2-HMAC-SHA512 with 2,048 iterations. The mnemonic is the password input; the salt is `mnemonic` followed by the passphrase. Both the mnemonic and passphrase are encoded as UTF-8 after NFKD normalization. An empty passphrase is valid; adding a passphrase derives a separate wallet. Passphrase strings that normalize to the same text derive the same seed.

The **base wallet** here means the wallet derived with an empty BIP39 passphrase. It is not a parent wallet containing all your passphrase-protected accounts. Distinct normalized passphrases derive separate seeds, from which wallet software derives keys and addresses.

Someone who steals only your mnemonic can check the empty-passphrase wallet immediately. If they spend its small balance, your monitoring system can alert you. They may also try passphrase guesses, checking derived addresses for activity. A strong passphrase can make that search impractical; a weak one may fall quickly.

A BIP39 passphrase is not an application's unlock password or PIN. Those usually protect local access or encrypted storage, rather than deriving a separate wallet.

Set up a useful tripwire

1. Confirm that your wallet supports BIP39 passphrases and that its recovery procedure distinguishes an empty passphrase from your protected wallet's passphrase. Never enter the mnemonic into a website to create or test a canary.
2. Record the relevant network, account and derivation path. For example, Bitcoin mainnet native SegWit accounts following BIP84 use account paths of the form `m/84'/0'/account'`, where `account` is a hardened account index. The mnemonic alone does not specify which address types or accounts to search.
3. Fund one receiving address in the empty-passphrase wallet with an amount you can lose completely. Make it economically spendable: Bitcoin dust relay policies and transaction fees matter, while account-based networks generally require native currency for fees unless another mechanism pays them.
4. Put protected savings in the wallet derived with your separate passphrase. Back up that passphrase accurately, separately from the mnemonic. Losing either component can make those funds unrecoverable.
5. Monitor the canary without keeping signing keys online. Use a watch-only address, wallet descriptor containing only public information or appropriately scoped extended public key. An address watches one destination; an extended public key can expose a wider account history and future addresses.
6. Test notifications with a deliberate transaction. Monitor outgoing activity, not merely incoming payments or a displayed balance. Where supported, pending-transaction alerts can precede confirmation, but pending transactions can be replaced or dropped.

A single watched address is simplest if you never move the canary yourself. If you spend from it, change may go to another address. Update monitoring or use a watch-only account that follows receiving and change addresses.

What the warning does not prove

An unauthorized canary spend indicates that the funds' spending conditions were satisfied, not necessarily that the entire mnemonic leaked. A stolen individual private key, compromised signing device or previously signed transaction may also explain it. A pending transaction alone does not establish that the spend will confirm.

Conversely, no movement proves nothing. An attacker can recognize the canary, ignore it, search other accounts first or wait. If they obtained your passphrase too, they can target protected funds immediately. There is no guaranteed interval between the canary moving and other funds disappearing.

A watched address in the protected wallet can also reveal unauthorized spending, but that is detection of access to the protected funds—not advance warning from the empty-passphrase wallet.

Other limits are operational. Notifications can fail, an indexer can lag, and your device can be offline. A transaction you forgot about can cause a false alarm. Unexpected incoming tokens or tiny deposits are not, by themselves, evidence that your keys leaked.

Monitoring also has privacy costs. Giving a service an address reveals your interest in it; giving it an extended public key reveals more. Funding the canary directly from protected savings may create an observable connection between them.

Respond to an unexpected spend

Treat an unexplained outgoing transaction as a potential key compromise. Check its inputs or sending account, destination and your own transaction records using an independent view of the network. Do not type the mnemonic into a recovery service advertised beside an alert.

If compromise is plausible, create a new wallet with fresh entropy on a trusted device and move remaining funds promptly. Merely changing the BIP39 passphrase leaves the exposed words in use and does not move existing funds; moving them to a wallet with fresh recovery material replaces the compromised backup. Preserve transaction details for investigation, but prioritize protecting assets still under your control.

FAQ

How much should a canary hold?

Only an amount you can lose, but enough to remain spendable under the network's fees and any applicable minimum-balance or output-relay policies.

Does a canary make a weak passphrase safe?

No. It adds a chance of detection, not resistance to guessing. An attacker can ignore it while testing passphrases.

Should monitoring software receive my seed words?

No. Monitoring needs public addresses or public derivation information, never the mnemonic, passphrase or private keys.

Sources

- BIP39: Mnemonic code for generating deterministic keys (github.com)
- BIP32: Hierarchical Deterministic Wallets (github.com)
- BIP84: Derivation scheme for P2WPKH based accounts (github.com)

Continue reading on bip39-phrase.com: How to Protect Your Crypto Wallet from Phishing
More on this topic: Security

Example sequences from the BIP39 wordlist

Randomly generated sequences of words from the BIP39 English wordlist. None of them passes the BIP39 checksum, so no wallet will accept them: they show the format only. Never store funds with a phrase you found in a document.

Example 1			
1 dinosaur	2 couch	3 planet	4 design
5 spice	6 magnet	7 lemon	8 produce
9 resemble	10 screen	11 nation	12 lecture

Example 2			
1 escape	2 black	3 grief	4 corn
5 arrest	6 ripple	7 into	8 zone
9 top	10 cup	11 trumpet	12 stadium

Example 3			
1 box	2 sketch	3 three	4 quit
5 green	6 clip	7 olive	8 aware
9 case	10 sell	11 journey	12 tooth

Example 4			
1 bench	2 slot	3 inspire	4 candy
5 symptom	6 mansion	7 grid	8 match
9 harvest	10 bench	11 arctic	12 rich

Example 5			
1 topple	2 melody	3 rack	4 page
5 bachelor	6 tunnel	7 wise	8 world
9 stove	10 chronic	11 swift	12 boat

Example 6			
1 iron	2 ivory	3 decorate	4 supreme
5 reflect	6 cycle	7 burst	8 volume
9 filter	10 cushion	11 learn	12 column

Example 7			
1 valid	2 goose	3 jacket	4 error
5 hour	6 sun	7 cherry	8 cash
9 remember	10 love	11 organ	12 account

Example 8			
1 eight	2 survey	3 client	4 provide
5 arrow	6 idle	7 feed	8 pair
9 luxury	10 spoil	11 force	12 loud

Reading a BIP39 phrase

Every word stands for an 11-bit number: its position in the 2048-word list, counted from 0. Here is Example 1 decoded word by word. Each word links to its page on bip39-phrase.com, where the list is numbered from 1.

#	Word	Index (0-2047)	Word no.	11 bits
1	dinosaur	499	500	00111110011
2	couch	391	392	00110000111
3	planet	1327	1328	10100101111
4	design	478	479	00111011110
5	spice	1676	1677	11010001100
6	magnet	1071	1072	10000101111
7	lemon	1022	1023	01111111110
8	produce	1373	1374	10101011101
9	resemble	1466	1467	10110111010
10	screen	1548	1549	11000001100
11	nation	1178	1179	10010011010
12	lecture	1016	1017	01111111000

The 12 words carry 132 bits: 128 bits of entropy followed by a 4-bit checksum. The checksum must equal the first 4 bits of the SHA-256 hash of the entropy.

```
Entropy (hex): 3e661e979ded190bdff55db758324d3f
```

This sequence ends in the checksum bits 1000, but SHA-256 of its entropy starts with 1101, so wallets reject it as an invalid phrase. That is the point of the checksum: about 15 of every 16 random or mistyped sequences fail it.

See also

- Why Hot and Cold Wallets Must Never Share a Seed · October 24, 2025
- Watch-Only Wallets: Tracking Funds Without Private Keys · October 18, 2026

Educational document. The example sequences are random and fail the BIP39 checksum: they are not wallets and must never hold funds. Security notes: bip39-phrase.com/security/