BitTrap White Paper

Abstract

Endpoint intrusion detection has traditionally been ineffective against hackers who leverage new techniques and develop new and unique attack vectors. BitTrap employs a new endpoint intrusion detection technique effective against arbitrary attack vectors or methods. This new technique works by placing valuables, e.g., a bitcoin wallet, in the endpoints and monitoring the blockchain for transactions involving this wallet. When a hacker is after financial gain and compromises an endpoint, he is bound to take the valuables. Then, this same action (e.g., the transfer) is broadcasted in the blockchain and works as a signal of endpoint compromise. We analyzed the incentives for the defender and attacker in this scenario showing the benefits of our approach.

Index

Introduction

Computer assets include computer systems or endpoints (e.g., desktops, mobile phones, tablets, embedded devices) and other information systems, including databases, storage systems, email, and social network accounts. A computer asset may be compromised, and its availability, confidentiality, and integrity are affected in the process.

The intentional compromise of an asset occurs when there is intent, including but not limited to monetization, damaging the availability of the asset, or retrieving sensitive information stored within. Nowadays, most hacks occur as part of a hacker's business: the hacker compromises an asset and looks to monetize the hack by stealing credit card or personal information, encrypting the disk, asking for ransomware, etc. However, the hacker may hold an amount as the expected return for his efforts. Remember, he is a businessman.

Even more, a hacker selects his targets (e.g., IP ranges, domains) and the actions performed against these targets, iteratively based on the information he obtains from his actions and external signaling.

In particular, during an attack, the hacker may decide to perform an action that harms an asset he has compromised. (e.g., encrypting data and calling for ransomware, joining the computer in a botnet, gathering and possibly selling secret information in the black market, pivoting to other assets in the network/organization, installing and persisting keyloggers or other information-gathering tools).

Organizations implement defenses to ensure availability, and more generally, minimize losses. Yet, organizations' efforts to minimize losses are (at least partially) blind to what hackers can do.

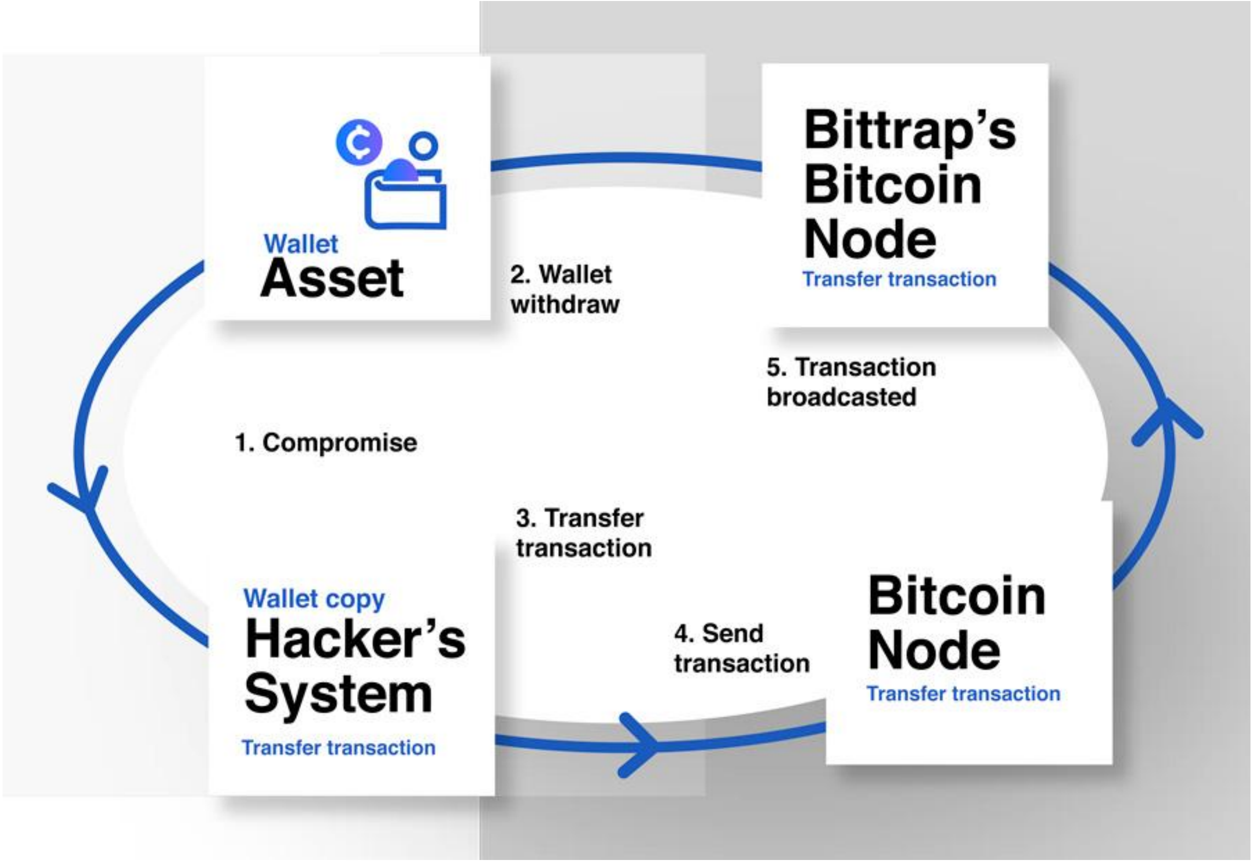
Moreover, it has been seen repeatedly that the cost suffered by organizations outweigh the hacker's financial gains. For example, in the recent cases of Colonial Pipeline or JBS ([CP], [JBS]) where resources for these organizations became unavailable for several days and the bare financial losses of interrupting the businesses was larger than the ransoms paid. This also does not consider the recovery expenses, fines, or attorney expenses, to name a few additional items.

The standard security approach detects and diagnoses a threat and reduces risk by eliminating, mitigating, or transferring threats. Threats that are undetected, accepted, or whose risk is misdiagnosed remain as residual risk.

Assume that once a hacker compromises an asset, he is asked if he is willing to accept a reward and walk away. BitTrap's game-changing approach allows a hacker who has compromised an asset to opt out of causing any harm and obtain a reward instead.

The BitTrap system and service

In short, the BitTrap service places a Bitcoin wallet, including an unspent transaction and the credentials required to spend the transaction. A hacker who has compromised a BitTrap-protected asset, who is aware of the wallet, may decide to spend the transaction. To do this, the hacker must publish a transaction transferring the wallet's bitcoins to an address he controls within a block's transaction list. A BitTrap service, running a bitcoin full node, receives the block, including the transaction (which is broadcasted to all nodes), and immediately triggers an alert for the asset's owner.



1. Negotiation - The Offer: Can we mediate with the hacker to prevent any harm?

BitTrap offers the hacker a way out, where he can be rewarded for spotting a threat and secure a payback. The act of transferring the bitcoin is enough to alert the asset owner that the asset has been compromised and, therefore, that the asset reveals a vulnerability used by the hacker. If the hacker transfers the reward and leaves without damaging the asset, it is a win-win situation.

In that sense, BitTrap offers the hacker a reward for spotting a vulnerability and moving away. A hacker compromising the asset may not be aware of the wallet, and hence, look for other ways to monetize. That is why the BitTrap service may include signs to advertise the presence of the wallet and the fact that this is an asset protected with BitTrap. We briefly discuss why this is an effective strategy in Section 4.

The amount of bitcoin placed in the wallet needs to be enough to entice the hacker to accept. This amount can be estimated empirically. For example, one may collect public information about ransomware attacks, including the amounts asked and paid, to determine a compelling value. For example, picking the 90th percentile would mean that 9 out of 10 hackers are comfortable with this value (see Section 4). Of course, when setting rewards for an endpoints class (e.g., desktops in a financial institution), the sampled ransomware attacks may be filtered to include only incidents for endpoints of the same class.

On top of public information, BitTrap has conducted experiments using vulnerable and BitTrap-protected endpoints in the cloud with different reward values. Again, this information may be used to compute a reward value, e.g., the 90th percentile of the rewards cashed in the experiment.

2. Negotiation - The Answer: Why would a hacker accept?

Once a hacker has noticed the offer, he must decide if he wants to take it or not. He might guess that taking the offer implies that the compromise will be noticed, and the detection may interfere with other activities which he can use to monetize the hack. The hacker may decide not to transfer the unspent transaction, thereby rejecting the BitTrap offer. This could be because he is not interested in monetary profit or thinks he may obtain a more significant profit by other means.

In the case the hacker accepts a monetary reward to walk away, the question is, what is the best offer BitTrap can make so that the hacker accepts the BitTrap option. For example, if the hacker holds an amount as the expected return for his efforts, and the wallet he finds does hold an amount at least that size, the hacker's expectations would be fulfilled.

A business-minded hacker may weigh his options: accept the offer or attempt to capture more monetary profit through additional steps. These additional steps will cost the hacker his time, and he might also reveal part of his toolset or techniques.

Even after these investments, the attack may fail or may end up with less value. For example, because the hacker attempted to compromise a second asset and failed, or because he expected to find 5,000 credit card numbers and found far less, or he was detected and kicked out by an endpoint detection system.

Hence, the hacker is blind to the difficulty of obtaining an alternative reward and the value of the reward. BitTrap's offer, on the other hand, is transparent and risk-free.

Moreover, since the nature of Bitcoin transactions is semi-private, the hacker accepting the reward can hide his identity within the blockchain and remain anonymous.

3. Why is the deal beneficial to the asset owner?

BitTrap complements the endpoint protection and detection ecosystem by bringing an additional layer of security. By nature, BitTrap-led detection is agnostic to the attack entry point and the attack vector. BitTrap does not require updates, and this is a great property. Moreover, not only does the BitTrap service not require updates, but it also has the additional property that protection does not decrease when the hacker is aware of the solution. Therefore, we can argue that security increases when the BitTrap technology is present.

BitTrap’s protection benefits need to outweigh the cost for the solution to be beneficial for the owner. When a hacker compromises an asset and transfers the bitcoin, this translates to a loss for the endpoint owner and his organization. Yet, there is an intrinsic asymmetry between how the owner, or the organization, values an asset and the value for a hacker. (This is the cost asymmetry described in the Introduction.) Therefore, there is an opportunity in those cases where the owner's valuation on his endpoint outweighs the hacker's expected return. When BitTrap sets the reward over the hacker's expectations, the chances are that he accepts the reward, refrains from harming the endpoint, and, finally, that the owner's investment (in the reward) paid off.

Nonetheless, the hacker who has compromised a BitTrap-protected asset can perform actions against the will of the asset's owner, such as copying files from the endpoint's storage, pivoting to other networked assets, and other actions which were possible before BitTrap. However, some possibilities are removed when the endpoint is protected with BitTrap. For example, if the hacker were to leave a backdoor or an agent running and at the same time cash the BitTrap reward, then it is possible that the security team will remove these–as we explain below.

The security team for the organization owning the asset is then notified of the compromise by the BitTrap service and can take action. Possible reactions include:

- Isolating the asset to prevent new attack steps, pivoting, or other forms of escalation.
- Triaging the attack in search for the attack path and a way to fix this for other endpoints.
- Investigating what has the hacker done after compromise
- Replacing passwords and credentials within the asset
- Reinstalling or redeploying the asset via a recovery procedure, free of whatever damage the hacker may have done.

Therefore, the organization learns with each hack, suffering minimal damage as described in the previous bullets, which represent a considerably reduced impact than what happens with modern-day ransomware.

4. Foreword: What happens in the long run?

Engaging hackers in the BitTrap negotiation affects the landscape positively. A quick argument would be that bringing BitTrap protection is positive because less harm is done every time a hacker accepts a reward instead of damaging the asset. This is positive.

On the other hand, one could argue that the collected pot of all the BitTrap rewards may incentivize hackers, and more harm could be done in the long run. Examining this argument, we notice that it assumes that BitTrap increases the incentive collectively. This is not necessarily true. BitTrap could calibrate the rewards so that the hackers obtain (statistically) the same reward as they obtain without BitTrap. Hence, the incentive does not change. Moreover, since BitTrap reduces the hackers' efforts (because they do need to sell any goods in the black market, mount and maintain a botnet, etc.), this could also be factored into the reward amount so that the incentive remains unchanged.

One possible outcome of BitTrap-protection at large would be that the hackers spend all their time looking for endpoint vulnerabilities, not in mounting spam botnets, ransomware rings, growing the black market, and other activities. According to that outcome, hackers would compete in a world of collaboration with endpoint maintainers, and eventually, a fair price for their efforts would be set.

However, even if compromises would grow as a result of hackers accepting rewards rather than harming assets (say, this happened temporarily), this would then result in the following benefits:

- Whatever additional hacks that happened to BitTrap-protected endpoints would be harmless
- A portion of the old harmful hacks would be converted to harmless hacks against BitTrap-protected devices
- Defenders learning faster from their vulnerabilities

One could further argue that a hacker who decides to accept a BitTrap reward and do no harm will continue to do this. This 'converted' hacker is a win. He is now part of a community of security researchers that collaborate in finding flaws in BitTrap-protected endpoints.

Conclusion

We have framed the asset-security problem as a problem of balances. Firstly, allowing organizations to balance their losses with hackers’ gains. Also, for the hacker to balance accepting the BitTrap offer instead of their other choices. And finally, a balance between the global repercussions of a BitTrap option that rewards hackers, showing an improvement on the overall situation.