Attacker Incentive Analysis

Ariel Futoransky

www.bittrap.com

Abstract

This is a short article discussing the relationship between a traditional attack lifecycle and an attack where the BitTrap incentive is present. We introduce a model to describe layered attacks and analyze the impact of applying this type of incentives to transform the attacker behavior.

We show that after adding the BitTrap incentive there is at least a 10x pressure to switch to the more benevolent path. Furthermore, if the hacker chooses to ignore and proceed with the traditional monetization strategy, the presence of the BitTrap incentive is neglectable in terms of the reward obtained (0.25% increase).

Our model

We are considering a setting with one attacker and an organization with multiple devices.

The attacker has a set of tools he can use to try to breach into each device. These tools can either be exploits that profit from software vulnerabilities or system misconfigurations, or social engineering or phishing tricks that make use of unsafe user behavior.

The attacker is trying to gain an economic advantage by hacking into the organization and accessing valuable information from one or more devices. This information can be sold on the dark web for a meaningful amount of money. In real life there are other possible monetization strategies but on this model we are focused on data exfiltration. Not all the devices contain this kind of valuable data, so the attacker will try to traverse through different nodes to access the ones he is interested in. To progress in his plans, the attacker will use his tools, or take advantage of trust relationships present among the nodes.

The organization has a layered approach to security. With only some devices on the perimeter that can be directly attacked. Security components patrol the infrastructure, trying to identify the attacker tools, suspicious behavior or attempts to violate access policy. If an event is detected, an incident response is started, kicking the hacker out of the compromised devices.

In order to analyze this setting we create a simplified model with two stages, X1: the perimeter exposed device, and X2: the target monetizable device. We represent the attacker tools as two numbers per device. The probability of successfully breaking into the device without being detected by trying each available tool, and the time it will take on average to complete the whole task.

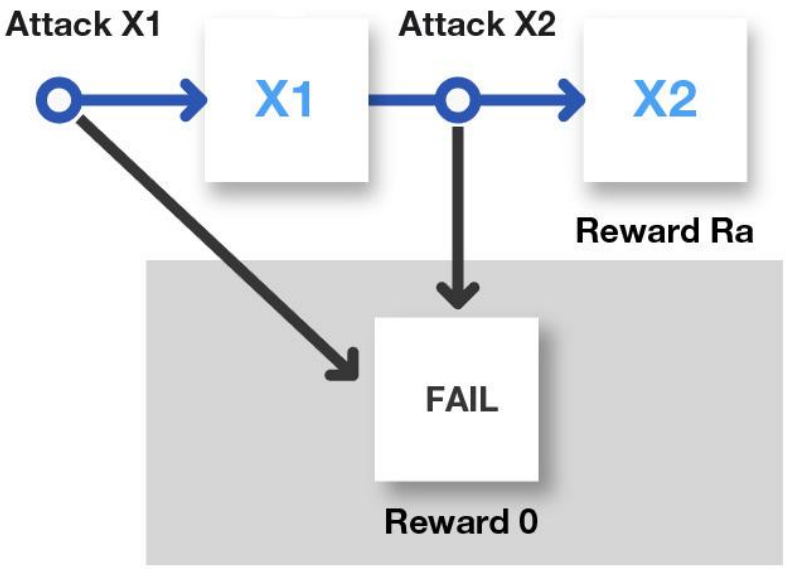
This is a good representation of our conceptual framework because the probabilities and times associated with breaking into X1 or X2 can summarize several intermediate steps, or alternative targets in a succinct way.

Scenarios

With this model in mind we describe two Scenarios:

Scenario A (SA)

This is the traditional scenario. There is a single path for the attacker, he needs to first break into the surface asset (X1), and then complete a second step to breach the target asset (X2)



If successful, he will collect the standard reward Ra.

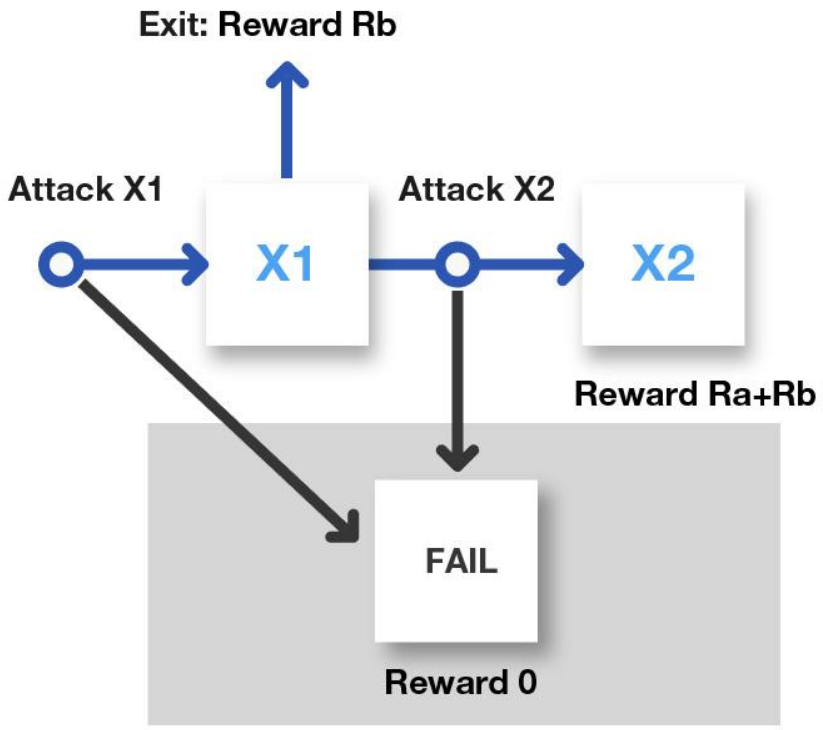
We model the first stage of the attack, breaking into X1, using two parameters P_1 and T_1. P_1 represents the probability of success for the attacker. T_1 represents the average amount of time invested in the attack for this stage, independent of the result.

We model the second stage of the attack likewise using P_2 as the probability of breaking into X2, and T_2 the time required given that X1 was breached successfully.

Scenario B (SB)

This is the scenario augmented with BitTrap. A new reward is added to X1 (Rb), Here the attacker has two choices, either going for the original target by breaking into X1 and then X2, where if successful he will be able to collect Ra + Rb (If the attack fails he is detected and collects 0 reward). Or retire as soon as breaking into X1 and collecting Rb.

The parameters of the model are equivalent to SA.



Analysis

For both scenarios, the Probability of breaking into X1 is

$$\mathbf{P(X1) = P_1}$$

And the probability of breaking into X2, which requires breaking into X1 first is:

$$\mathbf{P(X2) = P_1 \,.\, P_2}$$

(Here $P_2 = P(X2|X1)$ is the probability of the attacker breaking into X2 given that he has already broken into X1.)

The expected time for trying to break into X2 is:

$$\mathbf{ET_2 = T_1 + P_1 \,.\, T_2}$$

The expected reward for SA is:

$$\mathbf{Ea = P(X2) \,.\, Ra = P_1 \,.\, P_2 \,.\, Ra}$$

For scenario SB there are two possible strategies. Choosing exit on X1 yields the following expected reward:

$$\mathbf{Eb_1 = P_1 . Rb}$$

While choosing to attack X2 yields[1]:

$$\mathbf{Eb_2 = P_1 . P_2 . (Ra + Rb)}$$

The Return On Investment (*ROI*) for SA (expected reward per unit time invested) is:

$$\mathbf{\textit{ROI}\text{-}SA = Ea / ET_2}$$

Return on investment for SB choice 1 (exit) is:

$$\mathbf{\textit{ROI}\text{-}SB\text{-}1 = Eb_1 / T_1}$$

Return on investment for SB choice 2 (Attack X2) is:

$$\mathbf{\textit{ROI}\text{-}SB\text{-}2 = Eb_2 / ET_2}$$

Selecting some reasonable parameters:

P_1	0.1	The attacker breaks into 1/10 of the chosen initial targets
P_2	0.001	Breaking into the final target is as hard as breaking into three surface assets
T_1	10	10 days to try to break into X1
T_2	300	10 months average attack lifecycle
Ra	1,000,000	The stolen information from X2 can be sold for 1 Million dollars
Rb	2,500	We set $2500 as BitTrap's incentive wallet

[1] We assume that if the attacker fails with X2, he is detected and has 0 reward. Thus, he either gets a reward of Ra+Rb with probability $P_1.P_2$, or he gets nothing.

Yields the following results[2]:

Ea	100	Expected reward $100 per target
ET_2	40	Expected average time spent on targets looking for X2 is 40 days
Eb_1	250	Expected reward for choosing bittrap is $250 per target
Eb_2	100.25	Expected reward for choosing X2 on a BitTrap protected site
ROI-A	2.500	For SA, the expected reward per day is $2.5
ROI-B-1	25.000	For SB choosing exit, the expect reward per day is $25
ROI-B-2	2.506	For SB ignoring BitTrap and going for the original target yields $2.506 per day

[2] There are other possible extensions to the model that can be studied:

- Sometimes X1 is also monetizable. Like for example under ransomware monetization tactics. We could add an Ra-1 parameter representing this reward, and calculate the expected rewards accordingly.
- We could represent different sets of exploits with dissimilar characteristics, each with their own P,T pairs.
- We could represent independently the exploits characteristic with three possible outcomes: Probability of successfully breaking into the device, probability of failing but remain undetected, and probability of being detected.
- Another option is considering that each attack Xn has a cost to the attacker.

Conclusions

From this simple analysis we can conclude:

1. *ROI*-B-1 is 10 times higher than *ROI*-B-2, so there is a significant incentive for the attacker to choose the incentive on companies protected with BitTrap.

2. *ROI*-B-2 is only 0.25% higher than *ROI*-A, so there is no noticeable increase in the incentives to attack X2 after installing BitTrap. X2 is where the valuable information is stored.

This is a simplified model, but it can illustrate how the deployment of a simple strategic incentive can radically transform the behavior and perceived preferences of the attacker in a beneficial way. Considering that incentives can be deployed with almost no resource consumption (network, storage, computation) and that maintenance costs are negligible, the upside of investing in this type of strategy can be huge.