**Why YARA? (pt.2)** Matthew Jang

저번주에 YARA발표 하면서 제가 실재 test case 와 result 를 포함을 안해서 명학한 결론이 안나왔던 것 같습니다. 그래서TeslaCrypt 라는 Ransomware 를 실재로 실험해서 결과 랑 제 의견들을 보내 드립니다.

GitHub에 벌써 만들어진 YARA rule set들을 사용해서 우리가 갖고있는 TeslaCrypt 로그 데이터를 YARA가 탐지 할수있는지 없는지 를 먼저 알아 봤습니다.

예를들면 GitHub 에 있는 NioGuard\_TeslaCrypt.yar 랑 Reglar\_TeslaCrypt.yar rule 을 사용했습니다.

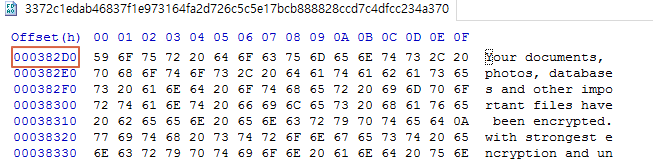
로그 는 TeslaCrypt 라는 file에 있습니다.

그래서 YARA 한테 보내봤습니다:

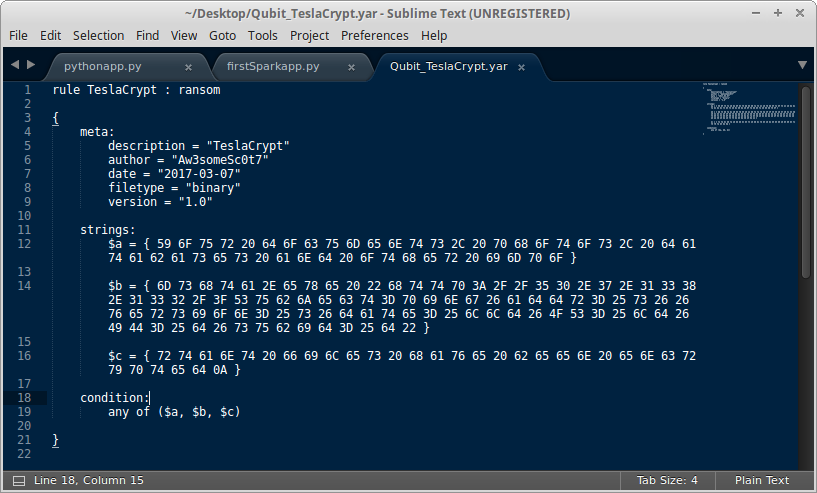
root@ubuntu:~/yara-3.5.0/yara\_test# yara -s -g NioGuard\_TeslaCrypt.yar TeslaCrypt

root@ubuntu:~/yara-3.5.0/yara\_test# yara -s -g Reglar\_TeslaCrypt.yar TeslaCrypt

하지만 보시다시피 아무 결과 가 나오지 안았습니다. GitHub에서 갖고온 rule들은 다른 회사 에서도 사용하는 rule들 인데 왜 우리 경우 에는 탐지가 안됬을가요? 답은 다음 step에 나왔습니다.

그래서 로그 파일을 직접 들어가서 hex strings들을 찾아서 manually YARA rule 을 만들어 봤습니다. (이런 method 는 static analysis 라고 합니다).   


Log file converted to hex strings

Qubit\_TeslaCrypt.yar

그리고 YARA 한테 보내봤습니다:

root@ubuntu:~/yara-3.5.0/yara\_test# yara -s -g Qubit\_TeslaCrypt.yar TeslaCrypt

TeslaCrypt [ransom] TeslaCrypt3

0x382d0:$a: 59 6F 75 72 20 64 6F 63 75 6D 65 6E 74 73 2C 20 70 68 6F 74 6F 73 2C 20 64 61 74 61 62 61 73 65 …

0x380f0:$b: 6D 73 68 74 61 2E 65 78 65 20 22 68 74 74 70 3A 2F 2F 35 30 2E 37 2E 31 33 38 2E 31 33 32 2F 3F ...

0x38300:$c: 72 74 61 6E 74 20 66 69 6C 65 73 20 68 61 76 65 20 62 65 65 6E 20 65 6E 63 72 79 70 74 65 64 0A

네, 결과 가 나왔습니다. 그럼 우리가 만든 rule 이랑 GitHub rule들 이랑 왜 다른 결과가 나왔을가요?

이유는 log data 마다 string 들이 다양하고 달라서 수많은 log data 마다 YARA rule을 만들어야 됩니다. 또는 TeslaCrypt 라는 Ransomware를 YARA를 써서 탐지하고싶다면 우리 data set 에 수많은 YARA rule들이 필요하겠죠. 하지만 TeslaCrypt 말고 다른 Ransomware 들도 엄청 많습니다. 그 많은 Ransomware 들 마다 다른 rule set 이 또 필요하겠죠.

Ransomware 들의 행동이나 strings들이 변하면 전에 있던 YARA rule 들은 효과 가 없어 집니다. Plura 입장에서 위에 처럼 malware가 있는 파일이 YARA 가 아무 이상없다고 agent한테 실수로 말하면 큰일이 나겠죠.

그리고 YARA rule 들은 log data가 발생한후에만 작성을 할수 있습니다. 그 말은 zero-day 인 경우에는 그 새로은 malware를 막을수 가 없습니다.

그래서 static analysis 는 accurate 하지가 않고 눈과손이 많이 가는 method입니다.

Dynamic analysis (e.g.행위분석) 도 있고 Machine learning 도 더 강한 alternative입니다. 이유는 machine learning 은 algorithm을 통해 prediction 을 할수있어서 malware의 행동이 변화되도 찾을수 있겠죠.