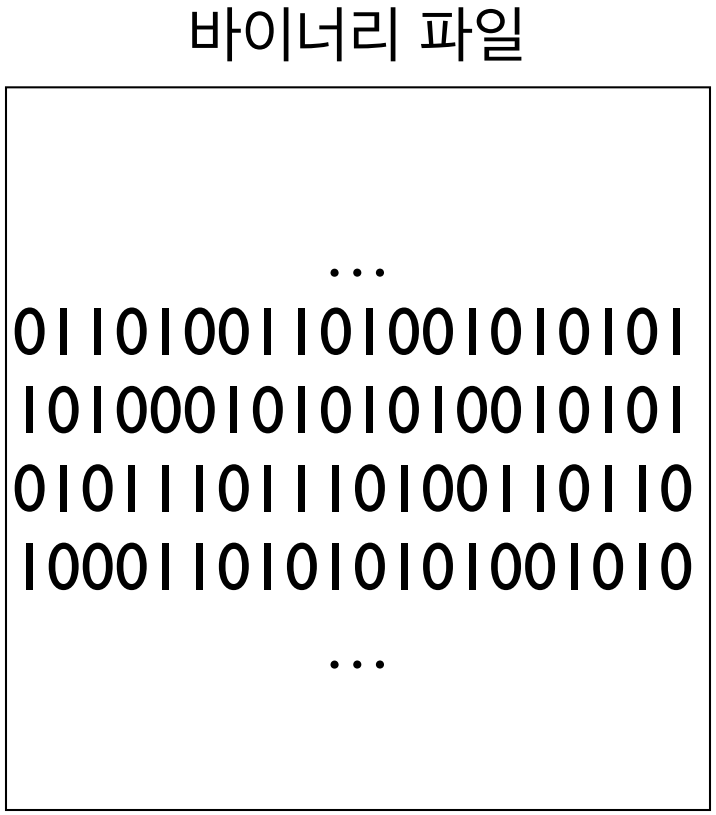
바이너리에서 코드영역을 찾아내는 정적분석

김솔
서울대학교 ROPAS
ROSAEC Workshop

바이너리 파일에서 코드 영역을 찾아 내야하는 이유

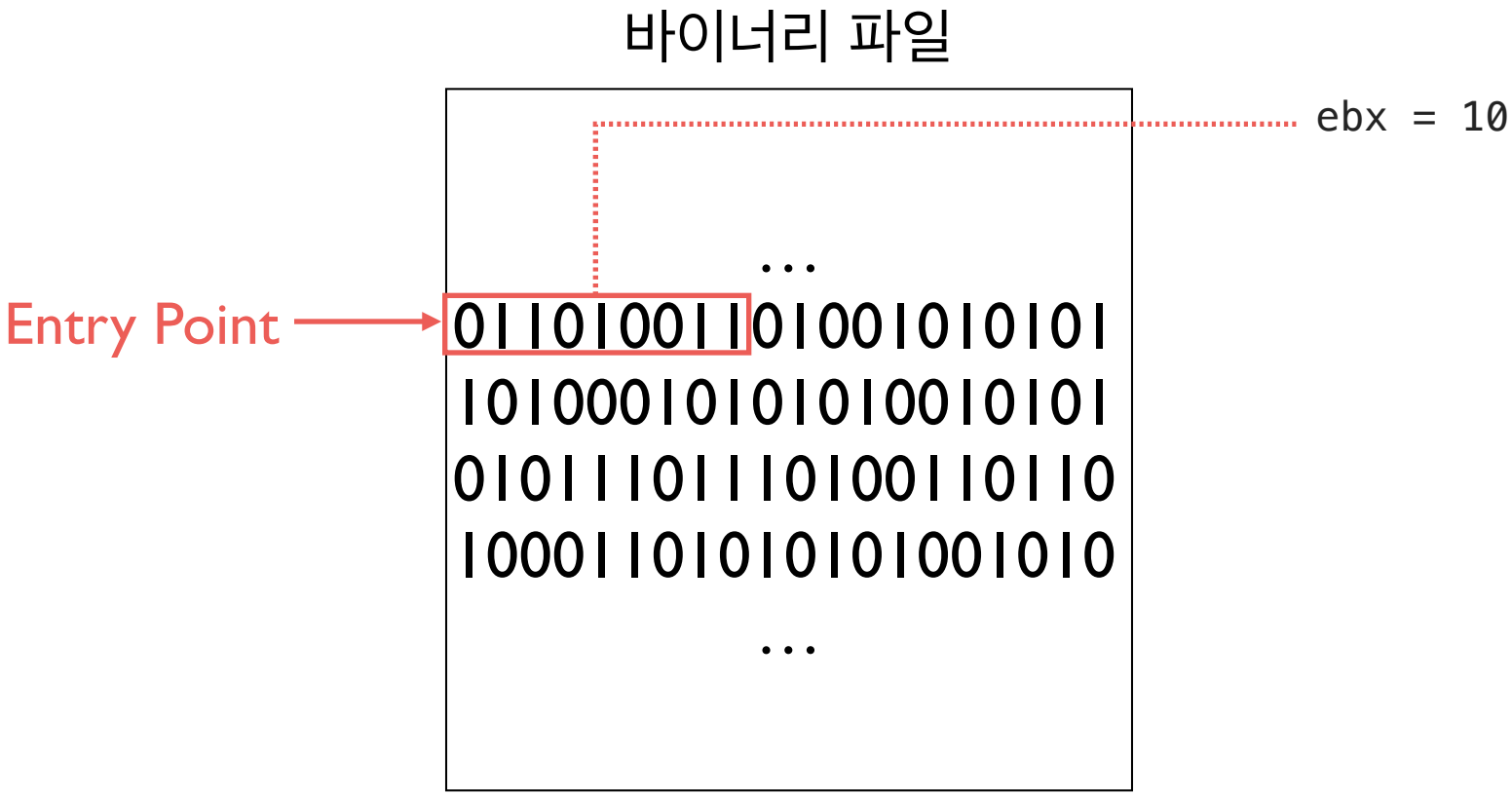


대체 어디부터 어디까지가 코드일까요?

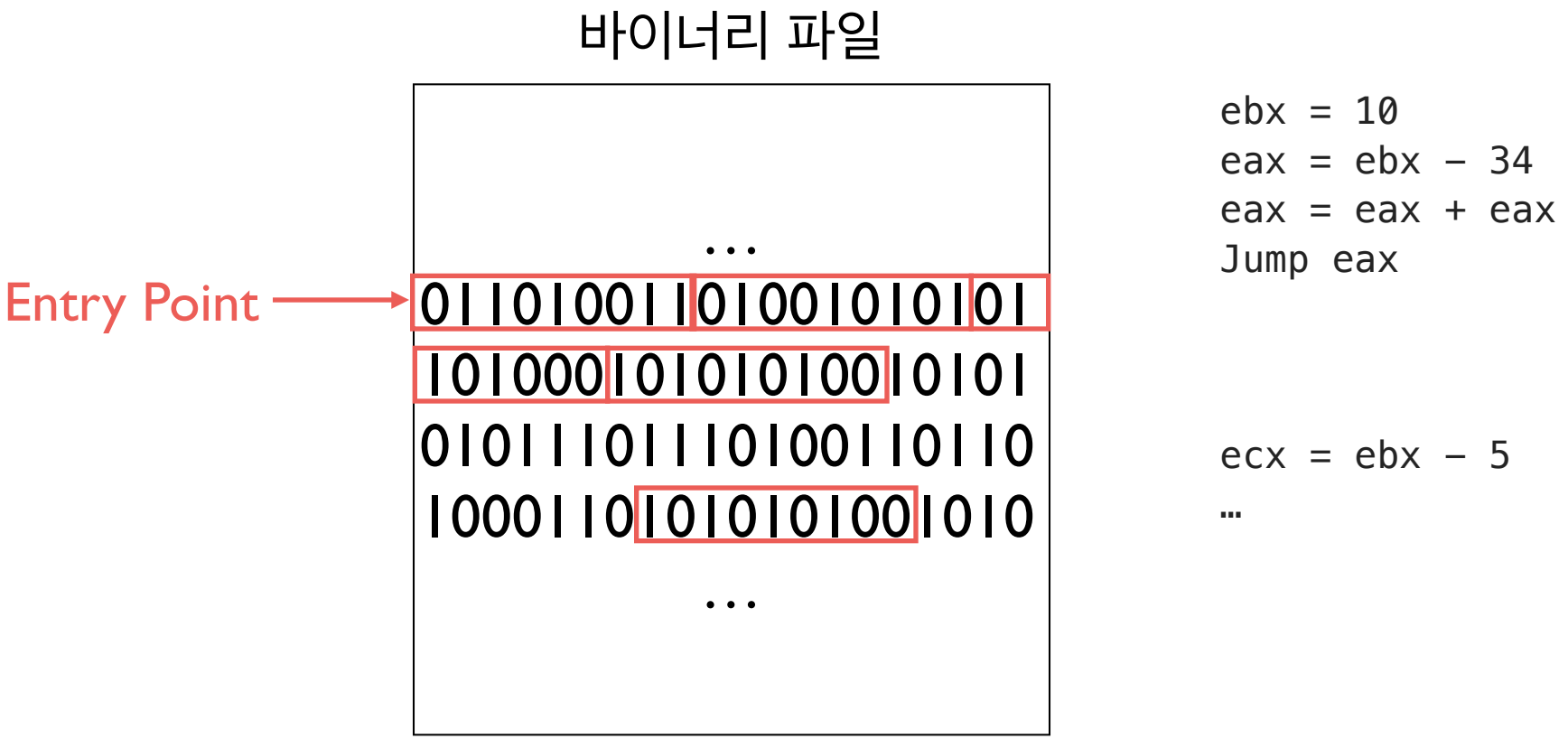
바이너리 파일이 우리에게 알려주는 정보들

- 제일 처음 실행할 코드의 주소
- 바이너리에 존재하는 각 Section들의 시작 주소와 사이즈
- Timestamp
- Checksum
- …

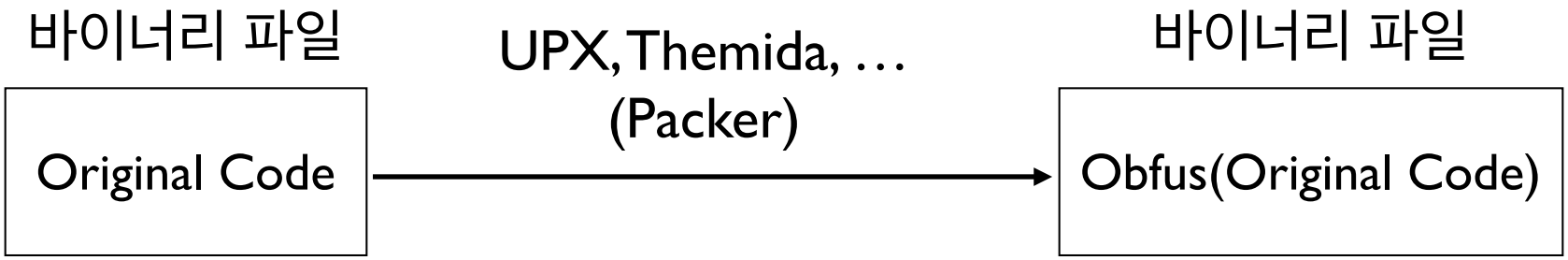
CPU는 그냥 믿고 따라갑니다



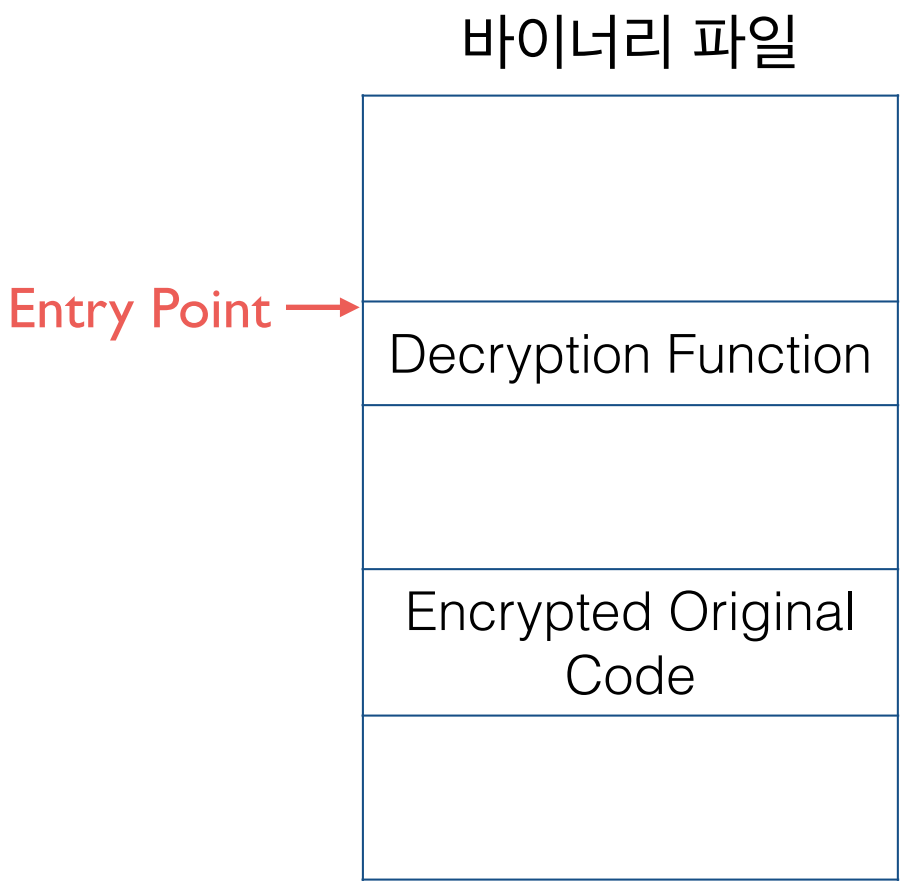
CPU는 그냥 믿고 따라갑니다



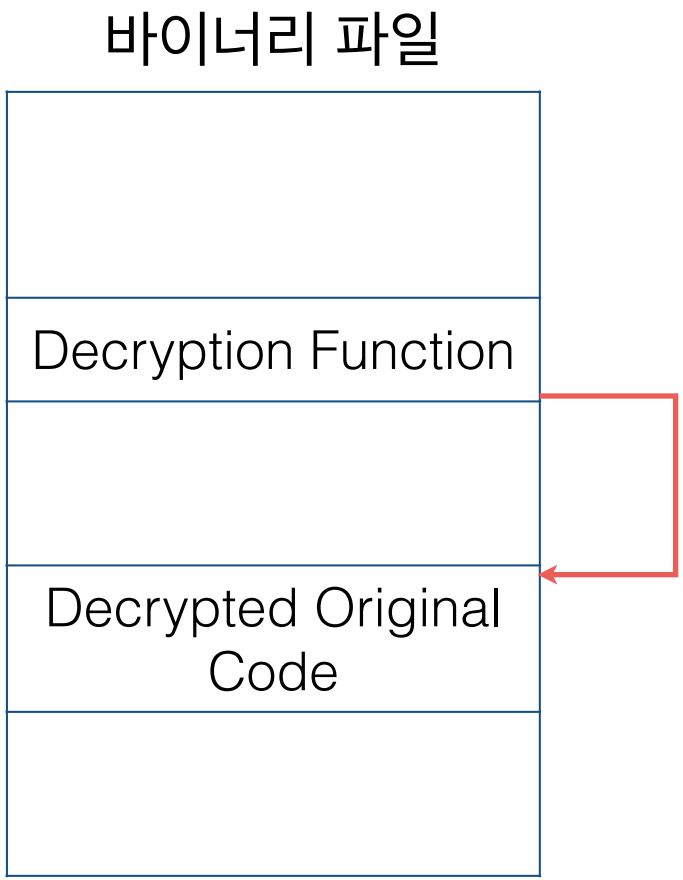
코드를 숨기기 위해 하는 복잡한 동작들



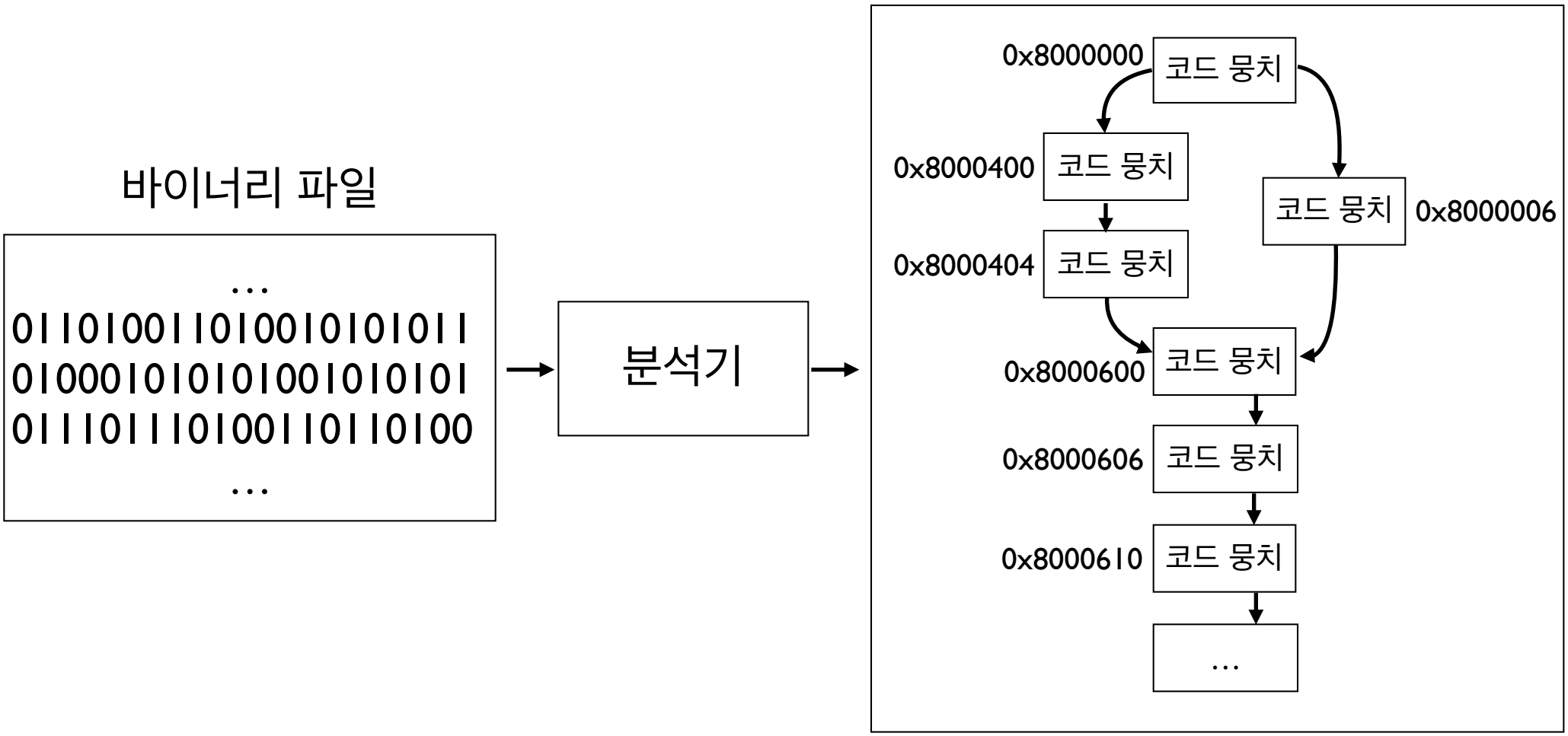
코드를 숨기기 위해 하는 복잡한 동작들



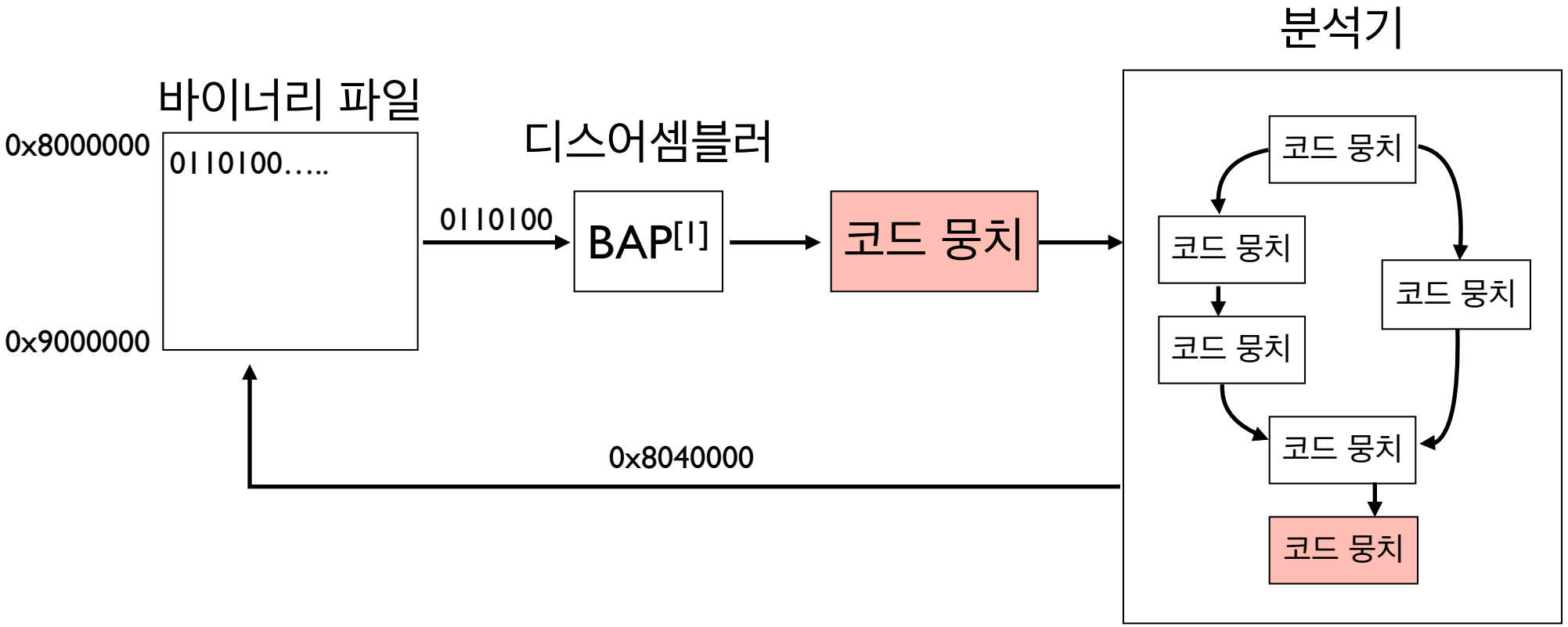
코드를 숨기기 위해 하는 복잡한 동작들



분석기가 하는 일

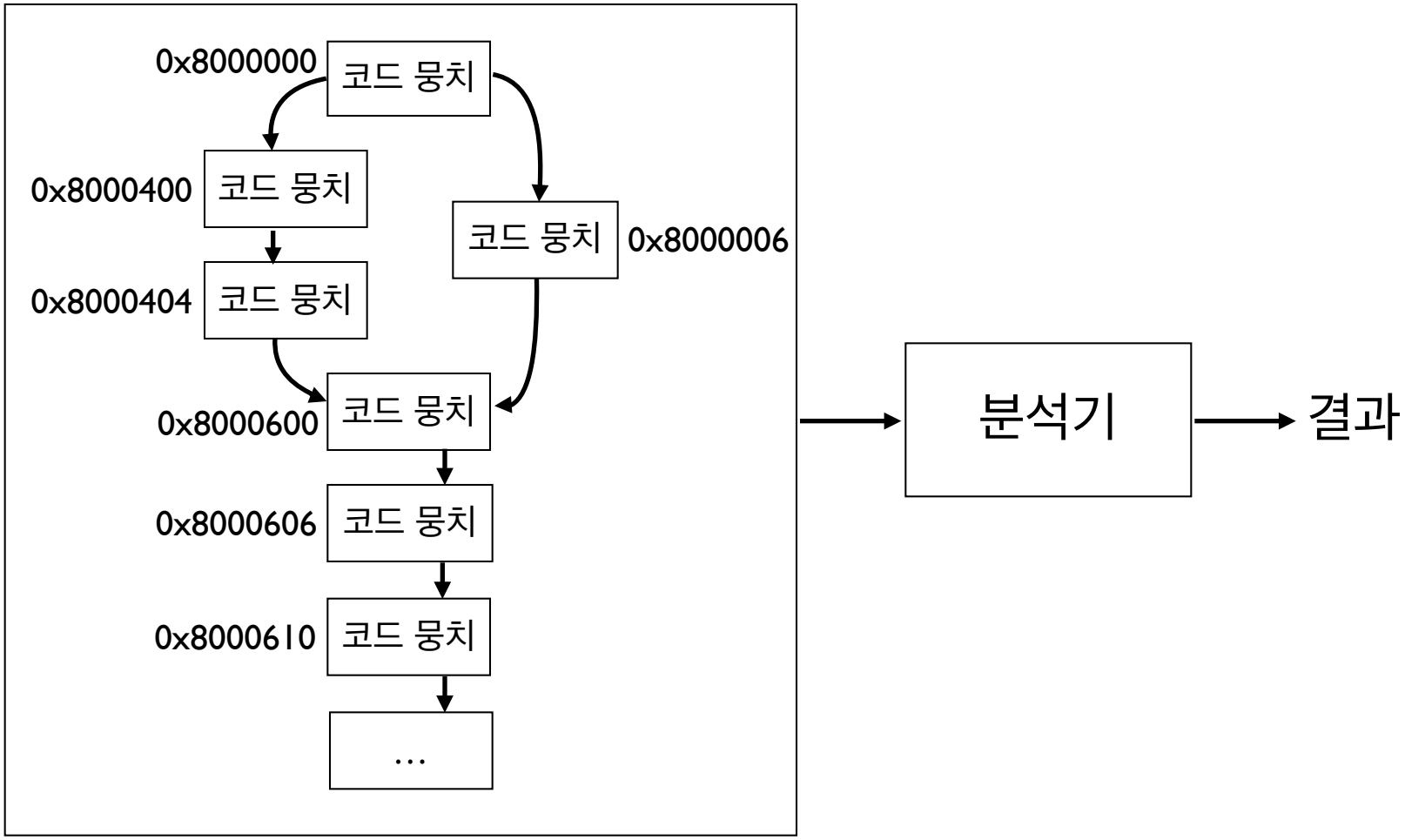

바이너리 파일
…
0110100110100101010011
0100010101010010101010
0111011101001101101000
…
분석기
0x8000000
코드 뭉치
0x8000400
코드 뭉치
0x8000404
코드 뭉치
코드 뭉치
0x8000006
0x8000600
코드 뭉치
0x8000606
코드 뭉치
0x8000610
코드 뭉치
…

분석기 구조



[1] BAP : Binary Analysis Platform.
David Brumley, Ivan Jager, Thanassis Avgerinos, Edward J. Schwartz. 2011

분석 결과를 다시 분석에 이용



감사합니다.