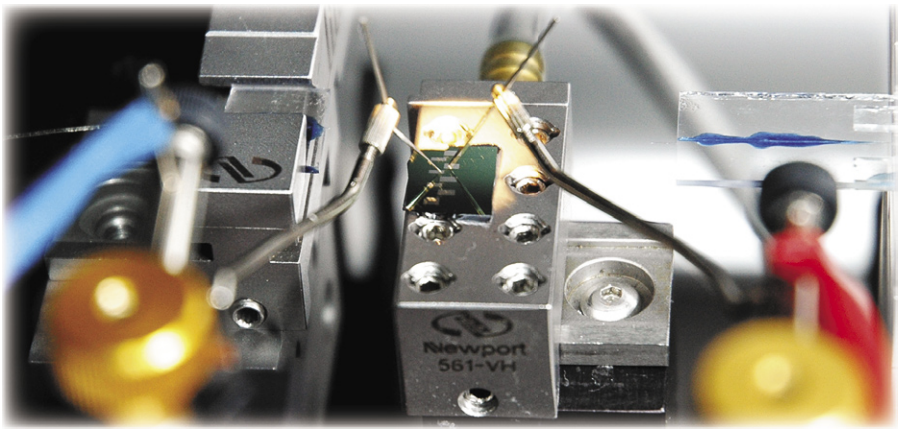
Atak na mikrokontrolery!

Już kiedyś na łamach Elektroniki Praktycznej (EP 8…10/2003) zajmowaliśmy się zagadnieniami bezpieczeństwa danych zapisanych w pamięci mikrokontrolera. Wówczas opisywaliśmy metody i techniki dostępne praktycznie tylko osobom dysponującym odpowiednim kapitałem niezbędnym na wynajęcie potrzebnego osprzętu. A jak jest dziś? Czy postęp technologiczny, który teoretycznie powinien poprawić poziom zabezpieczeń danych naprawdę poszedł w parze z technologią? Opowiada o tym niniejszy artykuł napisany na podstawie materiałów dostępnych z badań przeprowadzonych na uniwersytecie Cambridge [1].



Bezpieczne mikrokontrolery i karty *smartcard* opracowano w celu ochrony poufności i integralności ważnych informacji. W związku z tym, że częstokroć są to dane ściśle poufne, związane z dostępem do pilnie strzeżonych zasobów (np. konta bankowe klientów, klucze kodujące transmisję GSM), to wielokroć różne osoby próbują je odczytać i powielić z rozmaitych, niekoniecznie uczciwych, pobudek. Zabezpieczenie nie pozwalające włamywaczowi na znalezienie wartości zapisanego klucza kryptograficznego nie jest wystarczające. Nie wolno mu również pozwolić na ustalenie wartości którejkolwiek części klucza, albo spowodować takiego uszkodzenia, które pozwoli na odczytanie poufnej informacji (przypomnijmy, że większość kluczy szyfrowania składa się z co najmniej dwóch części: prywatnej i publicznej). Dotyczy to zarówno pamięci danych, jak również kodu programu. W dzisiejszych czasach najpowszechniej używaną metodą powodowania tego typu uszkodzeń i błędów jest *glitching*. Jest to metoda polegająca na powodowaniu krótkotrwałych zmian napięcia na liniach zasilania lub zegarowych. Tak naprawdę skutki wywoływane przez stosowanie tej metody są trudne do przewidzenia i często zależą od przypadku. Znane są jednak „recepty" na odczytanie zawartości pamięci mikrokontrolera po zwarciu bitów bezpieczników, wykorzystujące rozmaite formy *glitchingu*. Częstokroć bezpieczniki są wykonywane w technologii Flash, po to aby układ można było wielokrotnie przeprogramowywać. Jest to cecha większości popularnych mikrokontrolerów.

Metody używane do odczytania zawartości *smartcard* lub mikrokontrolera dzielą się na dwie grupy: inwazyjne – przeprowadzane z wykorzystaniem wyposażenia służącego normalnie do testowania układów (stacja z mikrosondami) oraz aparatury umożliwiającej operowanie skupioną wiązką jonów i ich implantację w strukturze układu, lub nieinwazyjne – wykorzystujące emisję elektromagnetyczną układu, protokół komunikacyjny, lub inne możliwości (np. dostęp przez zewnętrzne interfejsy układu, błędy w oprogramowaniu firmowym). Atak może być wykonywany metodą aktywną lub pasywną. Ogólnie mówiąc, atak pasywny polega na „podsłuchiwaniu" komunikacji z układem podczas realizacji procedury komunikacyjnej lub programu. Atak aktywny to już próba przejęcia kontroli nad układem przez wymuszenie stanów sygnałów zewnętrznych. Klasycznym przykładem jest dołączenie do linii sygnału zegarowego licznika rozkazów mikrosondy połączonej z masą tak, aby zabronić realizacji instrukcji skoku.

Aż do teraz przeprowadzenie ataku inwazyjnego wymagało zaangażowania sporego kapitału w zakup wyposażenia laboratoryjnego plus bliżej nieokreślonego w wysiłek poświęcony każdemu z układów. Ataki nieinwazyjne takie, jak analiza pola elektromagnetycznego w najbliższym sąsiedztwie układu, wymagają raczej niewielkiego kapitału niezbędnego na zakup i konstrukcję sprzętu, i jak w poprzednim przypadku – nieokreślonego na analizę każdego z układów. W związku z takimi cechami, ta metoda ataku jest bardzo atrakcyjna i bardzo często wykonywana przez osoby, które owszem, potrzebują kupić i skonstruować niezbędny osprzęt, ale nie liczą własnego czasu spędzonego na opracowywaniu metod włamania oczekując przyszłych zysków. Z tego też powodu metoda nieinwazyjnej analizy promieniowania elektromagnetycznego jest często stosowana.

Na nieszczęście włamywaczy, producenci układów scalonych już zabezpieczają je przed oczywistymi metodami ataku, wprowadzając na przykład losowy *jitter* do zegara systemowego układu, co sprawia, że analiza realizowanych sekwencji poleceń jest bardzo trudna. Z drugiej strony układ reaguje na szybkie zmiany napięcia restartem, przerywając realizację realizowanego kodu. W międzyczasie wykonywanie ataków inwazyjnych również staje się coraz trudniejsze, ponieważ wzrasta po pierwsze skala integracji, a po drugie struktury układów zaczynają być budowane w postaci wielowarstwowej i przez to dostęp do ich ważnych elementów staje się coraz trudniejszy. Jako odpowiedź na tego typu zabezpieczenia pojawiła się nowa metoda odczytu zawartości pamięci łącząca w sobie obie metody i przez to nazywana *semi-inwazyjną*.

W tej metodzie struktura układu musi zostać wyeksponowana, ale nie ma potrzeby zdejmowania warstwy pasywacyjnej chroniącej strukturę, ponieważ nie jest wymagany kontakt elektryczny ze strukturą układu.

Atak semi-inwazyjny nie jest metodą całkowicie nową. Owszem, rozwój elektroniki umożliwił wykorzystanie pewnych nowych technik ataku, jednak od lat znana jest metoda odczytu danych z pamięci EPROM polegająca na naświetleniu bitu *write protect* światłem ultrafioletowym. Wymaga ona co prawda wyeksponowania struktury układu, ale nie jest potrzebny

żaden kontakt elektryczny, jak również nie jest ona w żaden sposób uszkadzana, czy naruszana. Oczywiście taki „atak" pokazuje tylko pewne teoretyczne możliwości selektywnego oddziaływania na strukturę układu, ponieważ pamięć EPROM w żaden sposób nie była zabezpieczona przed odczytem i można było bez problemu skopiować jej zawartość. Bit miał jedynie zabezpieczać przed przypadkowym zaprogramowaniem pamięci przez nieuwagę, a funkcję swoją pełnił tylko do najbliższego kasowania zawartości pamięci światłem ultrafioletowym.

Teoretycznie, atak semi-inwazyjny może być przeprowadzany z wykorzystaniem narzędzi, takich jak: odczynniki chemiczne i światło ultrafioletowe, promieniowanie rentgenowskie, światło laserowe, pole elektromagnetyczne i lokalne podgrzewania struktury układu. Wszystkie wymienione czynniki mogą być użyte indywidualnie, lub w pewnych kombinacjach wzmagających skutki ich działania. Niestety, ostatnie lata pokazały, jak niebezpieczna i łatwa w użyciu stała się ta metoda.

Złącze półprzewodnikowe jest bardziej wrażliwe na promieniowanie jonizujące, niż stosowane w początkach elektroniki lampy. W latach sześćdziesiątych, podczas eksperymentów przeprowadzanych z laserem pulsacyjnym stwierdzono, że intensywne światło jest w stanie powodować w strukturze półprzewodnika zjawisko podobne do wywoływanego przez promieniowanie jonizujące. I wówczas zaczęto używać światła laserowego zamiast promieniowania, jako znacznie tańszego i wygodniejszego w użyciu. Z biegiem czasu lasery gazowe zostały zastąpione przez znacznie tańsze lasery półprzewodnikowe i w rezultacie technologię zaczęto stosować również poza laboratoriami naukowymi.

Światło może zjonizować obszar półprzewodnika, jeśli energia jego fotonów przekracza energię bariery potencjału w półprzewodniku. Światło laserowe o długości fali 1060 nm, o energii fotonów 1,17 eV, penetruje krzem na głębokość około 700 μm i zapewnia dobrą, jednorodną, przestrzenną jonizację obszaru. Światło laserowe o kolorze czerwonym lub zielonym jest bardziej absorbowane przez półprzewodnik i nie jest w stanie jonizować go na tak dużą głębokość, ale trzeba mieć na uwadze, że współcześnie wykonywane układy scalone są coraz cieńsze. W praktyce więc światło widzialne jest zupełnie wystarczające, a co za tym idzie, tego typu laser można nabyć w praktycznie każdym sklepie elektronicznym lub sklepie z zabawkami czy materiałami biurowymi. Możliwości tej technologii są ograniczone, ponieważ skupienie światła jest ograniczone przez dyspersję praktycznie do obszaru kilku mikrometrów i nie jest odpowiednie do nowoczesnych układów scalonych.

Oczywiście każdy może użyć opisanej techniki, ale nie każdy ma wystarczającą wiedzę. Mimo, że w Internecie umieszczonych jest wiele publikacji na temat jonizującego wpływu światła lasera na półprzewodnik i możliwych zastosowań, to jednak brak jest konkretnych i sprawdzonych informacji. Nikt nie podaje w jaki sposób światło lasera mogłoby prowadzić interakcję z funkcjonującym układem. Czy można na przykład zmienić zawartość komórki pamięci CMOS i jak łatwe, bądź trudne jest wykonanie tej czynności.

Na uniwersytecie w Cambridge, w laboratorium kierowanym przez Sergeia Skorobogatova i Rossa Andersona, zdecydowano się na wykonanie próby z użyciem światła laserowego i układu firmy Microchip typu PIC16F84. Układ ten to mikrokontroler wyposażony w 68 bajtów statycznej pamięci RAM. W sposób standardowy, (patrz EP 8…10/2003 artykuł pt. „Atak na mikrokontrolery"), zdjęto obudowę struktury układu. Przy pomocy mikroskopu zlokalizowano obszar pamięci RAM w środkowej części struktury układu. Ze względu na ograniczenia budżetowe oraz założenia projektu, który miał wykazać ogólną, niebezpieczną dostępność metody, zdecydowano się na użycie taniej lampy błyskowej typu Vivitar 550FD, używanej przez fotografów. Aczkolwiek natężenie światła uzyskanego z lampy jest znacznie mniejsze niż to możliwe do uzyskania za pomocą lasera, to jednak przy właściwym skupieniu wiązki można osiągnąć pożądany stopień jonizacji półprzewodnika. Lampa błyskowa została zamocowana na obudowie mikroskopu Wentworth Labs MP-901 w miejscu, w którym normalnie mocowana jest kamera filmująca obraz. Wybrano powiększenie 1500×. Mikrokontroler został zaprogramowany w taki sposób, że pamięć mogła być zapisywana i odczytywana z użyciem interfejsu szeregowego. Jak łatwo się domyślić, pamięć była zapisywana pewną znaną wartością, wystawiana na działanie światła lampy błyskowej, a następnie jej zawartość była odczytywana i porównywana ze wzorcem przez oprogramowanie sterujące na komputerze PC. Dodatkowo komputer zmieniał czas trwania błysku tak, aby wybrać optymalną jego energię. Moc wyjściową lampy ustawiono na maksymalną.

Eksperyment (niestety) zakończył się pełnym sukcesem. Używając bardzo prostych „szablonów" wykonanych z folii aluminiowej udało się zmieniać zawartości pojedynczych komórek pamięci statycznej RAM. Stan komórki pamięci był zależny od obszaru wystawionego na działanie światła. W praktyce oznaczało to, że stan pojedynczego bitu mógł być logicznym 0 lub 1 w zależności od tego, który obszar komórki pamięci był oświetlany. Metoda była żmudna, niemniej jednak po rozpoznaniu struktury pamięci można było ustawiać dowolną wartość bajtu w jej obszarze! Potwierdzono w ten sposób, że przy pomocy ogólnie dostępnego wyposażenia możliwa jest zmiana zawartości komórek pamięci. Dodatkowo, mając dostęp do danych odczytywanych z pamięci, można bez większych problemów wykonać jej mapę. Oczywiście kwestią dyskusyjną jest użyteczność i wykorzystanie tej wiedzy, to jednak nie stanowi już przedmiotu niniejszych rozważań.

Światło emitowane przez lampę błyskową nie jest światłem monochromatycznym i dlatego bardzo trudno było kontrolować oświetlany obszar. Niedogodność tę eliminuje zastosowanie światła laserowego mającego znacznie lepsze skupienie. W kolejnym kroku w eksperymencie wykorzystano wskaźnik laserowy, który można nabyć w sklepie za około 30 zł. Osiągnięto identyczne rezultaty, jednak z kilkoma praktycznymi różnicami. Po pierwsze, nie było konieczne użycie mikroskopu i szablonów, ponieważ laser mógł precyzyjnie oświetlać punkt na powierzchni struktury układu, który mógł być odczytywany/oświetlany z częstotliwością około 100 cykli na sekundę. Z drugiej strony, pozycjonowanie lasera musiało być bardzo dokładne ze względu na mniejszy punkt i znacznie mniejszą aperturę oraz oczywiście wymiary elementów w strukturze układu. Wskaźnik był laserem klasy II o mocy mniejszej, niż 1 mW, ale zasilany był podwyższonym napięciem pozwalającym na uzyskanie mocy 10 mW. Długość fali światła wynosiła 650 nm.

Wykonane eksperymenty były wstępem do różnego rodzaju „zabaw" z mikrokontrolerami. Najlepsze z klasy tzw. mikrokontrolerów bezpiecznych nie były podatne na światło laserowe, jako że bardzo często klucz dostępu jest zmienny i częstokroć zależy od przechowywanej informacji (rozmiar bloku danych, suma kontrolna itp.). Z drugiej strony, popularne mikrokontrolery były zupełnie nieodporne na światło lasera: pozwalało ono na dowolne ustawienie stanu bezpieczników blokujących i przerzutników *flip-flop*, a po tym swobodny odczyt zawartości. Niewiele z układów jest bowiem opracowywanych w taki sposób, że uszkodzenie pojedynczego tranzystora nie zmienia stopnia zabezpieczenia układu. W praktyce okazało się, że standardowe układy CMOS praktycznie nie są w żaden sposób zabezpieczone przed opisywanymi wyżej atakami.

Na koniec w laboratorium wykonano testy z użyciem *smartcard* używanych chociażby jako popularne karty płatnicze, karty SIM, czy w systemach kontroli dostępu. Bez wnikania w szczegóły badań dość jest stwierdzić, że w 1997 roku trójka badaczy o nazwiskach Boneh, Demillo i Lipton wykazała, że błąd w czasie obliczeń klucza RSA może spowodować, że karta zwróci uszkodzoną sygnaturę, która może posłużyć do obliczenia sekretnej formuły szyfrującej i w rezultacie umożliwić dostęp do danych. W rezultacie badacze wykazali (niestety w dokumentach niedostępnych dla szerszego grona użytkowników), w jaki sposób można szybko i łatwo, używając opisanych wyżej metod, klonować telefoniczne karty SIM. Na klonowanie kart nie są już dziś potrzebne tygodnie, czy miesiące, ale dysponując odpowiednią wiedzą można to zrobić w przeciągu kilku minut.

Jacek Bogusz, EP
jacek.bogusz@ep.com.pl

Literatura

1. http://www.cl.cam.ac.uk/~sps32/