dr hab. inż. Janusz Stokłosa
prof. nadzw. w Politechnice Poznańskiej

Poznań, 9 marca 2015 r.

Recenzja rozprawy doktorskiej mgr. inż. Piotra Sygi

pt. „**Algorytmy ochrony informacji dla systemów urządzeń o ograniczonych możliwościach**"

wykonana dla Rady Naukowej Instytutu Podstaw Informatyki Polskiej Akademii Nauk

1. Przedłożona mi do recenzji rozprawa doktorska dotyczy wybranych problemów ochrony danych w rozproszonych systemach urządzeń, charakteryzujących się małymi mocami obliczeniowymi, pamięciowymi, komunikacyjnymi, zasobami energii. Pomimo tego, że jest to sytuacja uniemożliwiająca korzystanie z tradycyjnych metod kryptografii, to od takich systemów wymaga się zapewnienie odpowiedniego poziomu bezpieczeństwa przetwarzanym danym.

W rozprawie skoncentrowano się na modelach formalnych dwóch podobnych systemów rozproszonych: systemach identyfikacji radiowej (RFID) i radiowych sieciach sensorów. W systemach RFID ochronie podlega tożsamość użytkownika. W sieciach sensorów skupiono się na ochronie informacji o przebiegu wykonywanego przez sieć algorytmu.

2. Rozprawa składa się z 11 rozdziałów i obszernego wykazu literatury – razem 153 strony. W rozdziale 1 sformułowano tematykę rozprawy, przedstawiono przyjęte konwencje i notacje.

W rozdziale 2 przedstawiono podstawowe definicje i twierdzenia. Mają one charakter pomocniczy.

Rozdział 3 poświęcono systemom RFID. W skład systemu wchodzą znaczniki RFID (tagi) oraz czytniki. Znacznik jest pasywnym układem elektronicznym niewielkich rozmiarów wyposażonym w miniaturowa antenę. Czytnik składa się z nadajnika, odbiornika i dekodera oraz anteny nadawczo-odbiorczej. Znaczniki zdolne są do przesyłania drogą radiową krótkich komunikatów, np. swoich identyfikatorów. Znaczniki pasywne pobierają energię z fali elektromagnetycznej wysyłanej przez czytnik. Znaczniki aktywne korzystają z wewnętrznego źródła zasilania. Dodatkowym elementem systemu jest baza danych przechowująca listę wszystkich identyfikatorów i być może dodatkowe informacje powiązane ze znacznikami.

Znaczniki zwykle pozwalają na komunikację z każdym czytnikiem, nie weryfikując jego uprawnień. Ich moc obliczeniowa jest ograniczona, w rezultacie czego nie można w nich implementować typowych algorytmów kryptograficznych. Wśród podstawowych zagrożeń systemów RFID Autor wymienia: zagrożenia prywatności, klonowanie znaczników i ataki typu odmowa usługi. Zwraca także uwagę, że istnieją możliwości ochrony systemów RFID, dokonując przeglądu literatury w tym zakresie.

Rozdział 4 dotyczy sieci sensorów – sieci składających się z niewielkich urządzeń rozmieszczonych na ograniczonym obszarze w celu realizacji pewnego zadania, np. monitorowania temperatury, ciśnienia, wilgotności. Sieci sensorów mogą pracować zgodnie z ustalonym wcześniej schematem, bądź jako sieci *ad hoc.* Typowy węzeł takiej sieci składa się z mikroprocesora, pamięci, nadajnika/odbiornika radiowego oraz źródła zasilania. Każdy węzeł jest w zasięgu nadawania każdego innego węzła. Moc obliczeniowa węzła jest ograniczona, w rezultacie czego w sieciach takich nie stosuje się typowych, ze względu na uzyskiwany poziom bezpieczeństwa, metod kryptografii w celu ochrony danych. W szczególności rozmiary stosowanych kluczy symetrycznych są niewielkie, co czyni je podatnymi na ataki brutalne (wyczerpujące przeszukiwanie). Węzły komunikują się poprzez jeden wspólny kanał na zasadzie wielodostępu. W pracy przyjmuje się, że węzły są rozmieszczone w sposób losowy, mogą ulegać uszkodzeniom, w rezultacie czego nie dysponują pełną wiedzą o topologii sieci. W tej sytuacji ważne są procedury samoorganizacji sieci; do najistotniejszych Autor zalicza: inicjację, wybór lidera, estymacje rozmiaru sieci.

W sieciach sensorów rozważa się model adwersarza aktywnego, chcącego wpływać na rezultat wykonywanego protokołu. Podatność na tzw. ataki Sybilli (*Sybil attacks*, por. powieść F. R. Schreiber pt. *Sybil*, 1973; także poz. [32] w rozprawie) oznacza, że adwersarz może podsłuchiwać i wysyłać komunikaty w sieci, a także przejmować kontrolę nad niektórymi węzłami sieci, fałszując ich tożsamości. Inną formą ataku są ataki typu odmowa usługi (*Denial of Service*). Autor zwraca uwagę na fakt, że istotny nurt badań nad bezpieczeństwem sieci sensorów dotyczy przekazania sensorom kluczy symetrycznych do szyfrowania wiadomości przeznaczonych dla wybranej grupy węzłów.

W rozdziale 5 zaprezentowano protokół Kameleon. Chroni on prywatność użytkowników systemów RFID. Charakteryzuje się niewielkimi wymaganiami dotyczącymi znaczników. Zapewnia odporność przed szerokim spektrum ataków opartych na fizycznym przejęciu urządzenia wraz z materiałem kryptograficznym.

Protokół jest oparty na pomyśle częstych, losowych zmian identyfikatora znacznika RFID, rozumianego jako ciąg bitów o długości podzielnej przez 4 i parzystej wadze Hamminga. Zmiany te dokonywane są losowo na wielu pozycjach ciągu. Każdorazowo, gdy znacznik wyśle swój aktualny identyfikator, dokonywana jest zmiana identyfikatora w taki sposób, że wybrane losowo n spośród $2n$ bitów ulegają zanegowaniu. Powiązanie pomiędzy starym i nowym identyfikatorem jest silne: odległość Hamminga pomiędzy nimi wynosi n. Aby rozpoznanie znacznika było możliwe, w bazie danych przechowywany jest ostatni identyfikator znacznika i odpowiadający mu fabrycznie przypisany ciąg identyfikacyjny, znany tylko bazie

danych. Po wysłaniu przez znacznik swego aktualnego identyfikatora ID w bazie danych wyszukiwany jest identyfikator ID', taki że odległość Hamminga ID oraz ID' wynosi *n.* W ten sposób można z dużym prawdopodobieństwem zidentyfikować znacznik. W kolejnym kroku w bazie danych identyfikator ID zastępowany jest przez ID'.

Protokół Kameleon został przedstawiony w rozprawie w sposób formalny. Autor przeanalizował jego bezpieczeństwa. Wskazał, że dowolny stan jest osiągalny w dwóch iteracjach. Dokonał analizy rozkładu prawdopodobieństwa uzyskania określonego identyfikatora. Przedstawił sposób rozstrzygania niejednoznaczności identyfikacji.

Rozdział 6 zawiera opis ataku na protokół *Blocker tag*, zaproponowanego w pracy [63] i propozycję ulepszenia protokołu. Protokół *Blocker tag* służy zapewnieniu anonimowości użytkowników w pewnej klasie systemów RFID. Komunikacja pomiędzy znacznikami i czytnikiem odbywa się na zasadzie rozgłaszania. W celu uniknięcia kolizji czytnik musi określić adresata swoich wiadomości, stosując protokół wyodrębniania, np. *Query Tree*. Protokół *Blocker tag* korzysta z dodatkowego urządzenia, jako elementu systemu RFID, zwanego elementem blokującym (*blocker*). Jego zadaniem jest uniemożliwienie adwersarzowi odróżnienia rzeczywistych znaczników RFID od symulowanych przez element blokujący. Istotną wadą protokołu *Blocker tag* jest podatność na atak adwersarza potrafiącego określić moc odbieranego sygnału. Na tej podstawie można odfiltrować sygnał elementu blokującego i określić liczbę znaczników, które odpowiedziały na zapytanie. Autor zdefiniował silnego adwersarza, jako takiego, który zawsze potrafi określić liczbę transmitujących znaczników. Przedstawił atak na system RFID zawierający *n* elementów blokujących.

W dalszej części rozdziału 6 Autor prezentuje ochronę przed silnym adwersarzem za pomocą protokołu *Hedgehog blocker*. Rozpatruje system RFID z elementem dodatkowym zwanym *Hedgehog blocker*. Jest to urządzenie różniące się od znacznika możliwością regulacji swej mocy nadawania. W tym celu Autor proponuje wyposażenie *Hedgehog blockera* w pewną liczbę anten służących do transmisji bitów 0 i taką samą liczbę anten do transmisji bitów 1. W analizie bezpieczeństwa Autor nie zakłada nieznajomości przez adwersarza liczby anten. Korzysta z pojęcia (α,β)-nierozróżnialności dyskretnych zmiennych losowych.

Autor wykonał badania symulacyjne, z których wynika, że adwersarz dokonujący nawet wielu odczytów nie jest w stanie wyznaczyć np. liczby znaczników RFID.

Silny adwersarz, powtarzający zapytania, może wejść w posiadanie dodatkowej wiedzy, neutralizując w ten sposób efekt stosowania *Hedgehog blockera*. Autor opracował rozwiązania tego problemu: algorytm *Sequential Hedgehog* dla znaczników RFID oraz algorytm *Hash Hedgehog*.

Kolejnym elementem, jaki został wprowadzony przez Autora rozprawy do systemu RFID jest urządzenie zwane prewenterem (ang. *preventer*). Istotą działania takiego urządzenia jest powstrzymywanie znaczników przed wysyłaniem odpowiedzi na zapytania czytnika.

Rozbudowując w kolejnym kroku podstawowy system RFID, Autor bierze pod uwagę zastosowanie jeszcze jednego dodatkowego urządzenia: pośrednika (*Proxy Allower*).

Rozdział 7 dotyczy przesyłania informacji w sieciach sensorowych silnie ograniczonych. Autor opracował protokół dla sieci sensorowej typu *single-hop* (każdy węzeł jest w zasięgu każdego innego węzła), składającej się z węzłów o niewielkiej mocy obliczeniowej, wyposażonych w antenę nadawczą oraz z co najmniej jednego urządzenia wyróżnionego, zwanego ujściem, o znacznie większej mocy obliczeniowej (porównywalnej z mocą komputerów osobistych), mogącego wyłącznie odbierać wiadomości. Zakłada się, że węzeł jest zdolny do wyznaczania wartości ustalonej jednokierunkowej i nieodwracalnej funkcji mieszającej. Ujście współdzieli z każdym węzłem sekret. Zadaniem węzłów jest przesyłanie do ujścia informacji, na podstawie których jest wykonywane odpowiednie działanie. Ujście ma dostęp do bazy danych, w której zgromadzone są informacje o identyfikatorach węzłów i współdzielonych z nimi sekretach. Czas jest dyskretny, podzielony na rundy, umożliwiające wykrycie sygnału radiowego o ustalonej częstotliwości nadawania. Metodą odbierania informacji jest wykrywanie fali nośnej.

W modelu zakłada się istnienie adwersarza pasywnego, mającego możliwość podsłuchiwania przesyłanych wiadomości, w szczególności tożsamości transmitujących węzłów. Jego sukcesem będzie także ustalenie czy dwie wiadomości zostały wysłane przez ten sam węzeł.

Autor przedstawił formalny opis protokołu. Założył, że pojedyncze bity są reprezentowane w postaci tzw. *r*-kodów. Taka reprezentacja pozwala na korzystanie z prostej metody szyfrowania symetrycznego. Opisał procedury stosowane w protokole: kodowanie, transmisję, tworzenie preambuły, część identyfikacyjną, część odpowiedzialną za przekazywanie wiadomości właściwej, dekodowanie. W dalszej części wykonał analizę protokołu: poprawność wykonania i analizę złożoności czasowej. Wywody swoje poparł sformułowaniem odpowiednich twierdzeń i przedstawieniem ich dowodów. Omówił także możliwe rozszerzenia protokołu.

Rozdział 8 rozprawy poświęcono meta-algorytmowi, nazwanemu MAO, ukrywającemu przed adwersarzem zewnętrznym przebieg dowolnego protokołu wykonywanego w sieciach sensorów. Algorytm MAO(k) pozwala ukryć wykonanie dowolnego algorytmu z wysokim prawdopodobieństwem. W celu zmylenia adwersarza Autor wprowadził redundancję rund z losowymi stanami kanału komunikacyjnego. Kolejność wstawiania rund nadmiarowych jest pseudolosowo zdeterminowana przez sekret. Adwersarz nie jest więc w stanie rozróżnić rundy nadmiarowej od rundy rzeczywistego algorytmu. Za miarę bezpieczeństwa Autor przyjął (α,β)-ukrycie. W podstawowej wersji (wektor ukrywający wybierany z pewnego zbioru z prawdopodobieństwem jednostajnym) protokół zapewnia ukrycie doskonałe (α = 0) z prawdopodobieństwem 1 – β. Ten protokół został następnie rozszerzony w taki sposób, by możliwe było wykorzystanie szerszej klasy rozkładów prawdopodobieństw słów ukrywających poprzez zmianę parametrów α i β; ta wersja określana jest w rozprawie jako MAO(δ,k). Autor przeprowadził analizę protokołu MAO(δ,k), $\delta \in (0,1)$.

Algorytmowi OSA, aproksymującemu rozmiar sieci sensorów w taki sposób, by adwersarz nie był w stanie, na podstawie przebiegu protokołu, uzyskać informacji o liczbie węzłów,

poświęcono rozdział 9 rozprawy. Aby osiągnąć nakreślony cel można skorzystać z algorytmu MAO, jednak rozwiązanie przedstawione w rozdziale 9 jest efektywniejsze w sensie złożoności czasowej.

W rozdziale 10 zaprezentowano protokół MOC ukrytej inicjacji. Rzecz w tym, by adwersarz obserwujący wykonanie protokołu uzyskał jak najmniej informacji o jego przebiegu, szczególnie o liczbie węzłów sieci. Pod pojęciem inicjacji sieci rozumie się przypisanie każdemu z *n* aktywnych węzłów unikatowego identyfikatora, a także przydzielenie wspólnego sekretu powodującego, że dla adwersarza wyznaczone wartości są nierozróżnialnych od losowych. Wskutek zdarzeń losowych (np. zniszczenia sensorów) może zajść potrzeba powtarzania inicjacji. Algorytm składa się z dwóch faz: wyboru lidera i inicjacji właściwej.

W rozdziale 11 dokonano podsumowania rozprawy i nakreślono perspektywy dalszych badań.

3. Rozprawa ma charakter teoretyczny. Autor trafnie wybrał obszar badań. Oryginalne wyniki zawarł w rozdziałach 5–10. Zostały one zaczerpnięte z pięciu prac opublikowanych lub zgłoszonych do publikacji. Mgr inż. Piotr Syga jest współautorem tych prac. W rozdziale 1.2 określa swój wkład merytoryczny do każdej z nich. Niektóre analizy rozszerza w rozprawie o wnioski wynikające z przeprowadzonych przez niego badań symulacyjnych.

Moim zdaniem, najważniejsze wyniki, przedstawione w rozprawie obejmują:

- w rozdziale 6 fakt 36 z dowodem, że zmienna losowa oznaczająca liczbę *s* użytych anten i zmienna losowa oznaczająca siłę sygnałów znaczników są $(0, \frac{k}{s+1})$-nierozróżnialne, gdzie *k* jest liczbą prawdziwych znaczników RFID będących w zasięgu czytnika;
- w rozdziale 7 twierdzenie 42: Każda wiadomość protokołu jest dekodowalna z prawdopodobieństwem nie mniejszym niż $1 - \varepsilon$, niezależnie od momentu rozpoczęcia transmisji przez inne węzły;
- w rozdziale 8 twierdzenie 52 mówiące o tym, że w przypadku dowolnego *N*-rundowego algorytmu protokół MAO(δ,*k*) zapewnia (α,β)-ukrycie dla $\alpha = \delta^{-2N} - 1$ i $\beta = 1 - \left(1 - \delta^{\frac{L-1}{2}} \frac{2}{L+1}\right)^N$, dla pewnego $L \in \mathbb{N}_+$;
- w rozdziale 9 twierdzenie 55: Protokół OSA zapewnia $\left(0, \frac{1}{m \log_2 N}\right)$-ukrycie przebiegu aproksymacji rozmiaru sieci, gdzie *m* oznacza liczbę potoków danych, a *N* – górne ograniczenie liczby węzłów w sieci. Ponadto analiza złożoności czasowej doprowadziła do wniosku, że wykonanie algorytmu trwa dokładnie $(m \log_2 N)^2 + m \log_2 N$ rund;
- w rozdziale 10 twierdzenie 57: Dla ustalonego ograniczenia górnego liczby aktywnych węzłów *N*, algorytm MOC zapewnia $\left(0, \frac{2}{N}\right)$-ukrycie liczby aktywnych węzłów.

Praca napisana jest językiem zrozumiałym, zwięzłym. Jestem zdania, że tytuł rozprawy można by skorygować: zamiast mówić o „systemach urządzeń o ograniczonych możliwościach”, z informatycznego punktu widzenia lepsze byłoby sformułowanie: „Algorytmy ochrony informacji dla systemów urządzeń o ograniczonych zasobach”.

Usterek natury redakcyjnej spostrzegłem niewiele, wśród nich:

1. Na str. 24_4 (i dalszych) zamiast „funkcje haszujące" poprawniej byłoby „funkcje mieszające", a jeszcze lepiej, gdy o takich jest mowa, „kryptograficzne funkcje mieszające" lub po prostu „funkcje skrótu".
2. Na str. 34^{10} jest „wybór leadera", na str. 124_6 – „wybór (...) lidera".
3. Strona 34_2 – zamiast „protokół Diffie-Hellmana" powinno być „protokół Diffiego-Hellmana".
4. Na str. 44, rys. 5.8, brak opisu osi wykresu.
5. Na str. 49, rys. 5.10, nieszczęśliwie opisano oś rzędnych jako „Nr of rounds". W legendzie wykresu powinny być określenia pisane po polsku.
6. Na str. 61_1 mowa jest o „jednokierunkowej (w praktyce) funkcji haszującej". Podobnego określenia użyto też na str. 81_{21}: „nieodwracalnej (w praktyce) funkcji haszującej". Prosiłbym o wyjaśnienie jaka jest różnica pomiędzy jednokierunkową funkcją haszującą" a „jednokierunkową w praktyce funkcją haszującą" oraz „nieodwracalną funkcją haszującą a „nieodwracalną w praktyce funkcja haszującą".
7. Na str. 70_{11-10} użyto niezręcznego określenia „obliczać funkcję".
8. Na str. 104 w tytule rozdziału 8.3.1 jest MAO(k, δ), natomiast w tekście MAO(δ,k).
9. Na str. 114^4 zamiast „*trade-offu*" napisałbym „kompromisu".

Usterki te nie mają jednak wpływu na moją pozytywną ocenę zawartości merytorycznej rozprawy.

5. Reasumując stwierdzam, że:
   - tematyka rozprawy jest aktualna i ważna,
   - Autor rozwiązał zdefiniowane przez siebie problemy naukowe i użył do tego celu odpowiednich metod,
   - rozprawa świadczy o dużej wiedzy Autora i znajomości literatury z zakresu sieci sensorów.

Uważam, że przedstawiona mi do recenzji rozprawa spełnia wymagania stawiane dysertacjom doktorskim w Ustawie o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki z dnia 14 marca 2003 r. (Dz. U. nr 65, poz. 595). Wnoszę o dopuszczenie mgr. inż. Piotra Sygi do publicznej obrony rozprawy.