IOSUD - Universitatea Politehnica Timişoara
Şcoala Doctorală de Studii Inginereşti

CONTRIBUȚII LA DEZVOLTAREA ȘI STANDARDIZAREA WEB-ULUI SEMANTIC

Teză de doctorat – Rezumat
pentru obținerea titlului științific de doctor la
Universitatea Politehnica Timișoara

în domeniul de doctorat: Calculatoare și Tehnologia Informației
autor ing.: Zamfira Constantin Andrei
conducător științific: prof.univ.dr.ing. Horia Ciocarlie
luna **02**, anul **2022**

1. Actualitatea Temei

Pornind chiar de la titlu, în această teză am încercat să aduc contribuții la dezvoltarea noii generații a celui mai important și mai utilizat serviciu de Internet (World Wide Web), care este cunoscută sub numele de Web 3.0, sau *Semantic Web*.

Această versiune a apărut la începutul anilor 2000, fiind inventată de către nimeni altul decât părintele Internetului și al World Wide Web-ului, savantul englez sir Timothy Berners-Lee. Scopul său principal a fost să aducă o structură în dezordinea care exista pe Web, care este dominat de documente nestructurate (text simplu) și semi-structurate (XML, HTML), extinzând aceste documente la date care au o structură semantică. Un alt obiectiv al acestei noi generații a Web-ului este posibilitatea de prelucrare automată a datelor de către mașini, care să înlocuiască munca laborioasă a oamenilor de căutare, găsire, prelucrare, combinare a informațiilor din paginile web; altfel spus, s-a dorit automatizarea proceselor. Acest lucru se poate obține prin adnotarea resurselor web cu ajutorul unor adnotații care păstrează informații despre conținutul lor și care pot fi înțelese de către agenții automați de pe Web, numite *marcaje semantice*. Pentru a se asigura că agenți diferiți au o înțelegere comună a termenilor din aceste descrieri se folosesc ontologii unde se definesc atât termenii cât și relațiile existente între ei. Ontologiile reprezintă "piatra de temelie" a Web-ului Semantic, susținând realizarea proceselor automate, oferind vocabulare cu termenii și relațiile lor. Din cauza faptului că Web-ul Semantic este la începuturile sale, există în prezent o nevoie mare de aceste vocabulare cu date structurate semantic, precum și de agenți inteligenți care trebuie să execute operațiile necesare asupra resurselor de pe Web, cele mai importante fiind cele de raționament.

Pentru îndeplinirea obiectivelor de proiectare a fost creată o arhitectură a proiectului constituită dintr-o stivă de nivele care au fost propuse *ad-hoc* pentru a susține funcționalitatea Web-ului Semantic. Pentru majoritatea acestor nivele ale stivei au fost create limbaje și tehnologii, însă există altele care nu au fost încă standardizate datorită numeroaselor probleme și disocieri care există în acele domenii; acestea sunt nivelele de Logică, Încredere și Dovezi ale arhitecturii SW, pentru care nu au fost (încă) dezvoltate standarde.

Fiind spuse cele de mai sus, am considerat că domeniul Web-ului Semantic este unul potrivit pentru a-mi realiza cercetarea, cu multe arii care necesită soluționări de probleme și la care aș putea contribui și eu.

2. Conținutul tezei

Capitolul 1 – Ce este Web-ul Semantic?

Capitolul 1 al tezei de față este destinat unei introduceri în domeniul în care se realizează cercetarea, în cazul meu Web-ul Semantic. Acest capitol are drept obiectiv să prezinte și să explice cititorului noțiunile care stau la baza domeniului aplicație considerat. În acest scop am început capitolul cu definiția termenilor de 'Internet' și 'World Wide Web', prezentarea arhitecturilor și explicarea funcționalităților lor, care au fost fazele de evoluție prin care au trecut de la apariție și până în prezent. Apoi am prezentat domeniul pe care mi-am realizat cercetarea, cel al Web-ului Semantic. Am pus accentul pe arhitectura acestuia cu nivelele sale fiecare mapat la una (sau mai multe) tehnologii care au fost create și standardizate de către grupul care se ocupă cu realizarea proiectului complet, W3C. Aici a fost prezentată triada de tehnologii fundamentale ale SW, care sunt URI-urile, modelul de date universal RDF și limbajul de reprezentare al ontologiilor, OWL. Fiecare din acestea vor fi preluate și discutate mai pe larg în capitolele următoare (în special în cap.2).

Următorul subpunct principal al capitolului a fost destinat Domeniilor de Aplicație, acolo unde am prezentat și am explicat în ce ramuri ale industriei își găsește proiectul cele mai mari utilizări. Deoarece nu se poate acoperi tot domeniul discutat, am prezentat cinci domenii de aplicație pe care eu le-am considerat a fi cele mai importante, din cele studiate din literatură și am oferit referințe cititorului unde poate găsi și altele care se discută în alte lucrări din domeniu.

Bineînțeles, cum dintr-o lucrare științifică nu poate lipsi un studiu despre stadiul actual al cercetărilor din literatură (eng. *state-of-art*), și aici am realizat unul, care este o secțiune importantă a capitolului (ultima înainte de concluziile finale). Aici am prezentat și am enumerat cele mai importante realizări care au fost făcute pentru a sprijini dezvoltarea domeniului Web-ului Semantic, fiind atât de multe încât am considerat că este mai bine să le punem în categorii; astfel am vorbit despre medii de dezvoltare pentru aplicații SW, sisteme pentru efectuarea de inferențe (raționatoare), depozite de triple RDF, transformatoare între modelele de date (tradiționale-semantice, și vice-versa), vocabulare de ontologii publice din diferite domenii. În finalul secțiunii am făcut o serie de comparații, prezentate sub formă tabelară, și analize in formă textuală în legătură cu capabilitățile care pot fi furnizate de tehnologiile celor două versiuni de Web: cel tradițional și cel Semantic, rezultatele analizelor punând în vedere în mod cert superioritatea tehnologiilor semantice. Pe parcursul acestei secțiuni de review, așa cum se face în mod normal, pe lângă chestiunile discutate am oferit referințe în literatură unde cititorul își poate îmbogăți cunoștințele din domeniul respectiv.

Capitolul 1 se închide cu o secțiune de concluzii care prezintă foarte sumar conținutul capitolului și contribuțiile aduse în acesta.

Capitolul 2 – Logica Drept Suport pe Web-ul Semantic

Acest capitol își propune să discute rolul important pe care îl joacă formalismele logice la realizarea proceselor de pe Web-ul Semantic, în special acela de creare a limbajelor de reprezentare a ontologiilor. Paginile web sunt adnotate cu niște marcaje ce descriu conținutul lor care pot fi înțelese de către agenții de pe Web. Acești termeni, împreună cu relațiile lor sunt definiți în ontologii care sunt accesate de catre agenți heterogeni, pentru a stabili o terminologie comună și partajată, permițând agenților să interpreteze datele în mod corect și neambiguu.

Prima secțiune a acestui capitol discută despre *domeniul ontologiilor*, care, așa cum am spus mai sus, este fundația Web-ului Semantic. O ontologie este definită ca o specificație

formală a conceptualizării unui domeniu de aplicație dorit, încercând să captureze conceptele importante împreună cu relațiile dintre ele. Este o tehnologie pentru reprezentarea cunoștințelor care este utilizată în special în domenii care necesită stocarea unor volume mari de date și efectuarea de operații de inferență asupra lor, cum ar fi deducție, verificarea consistenței, interogare etc. În finalul secțiunii au fost prezentate pe scurt cele mai importante limbaje care au fost create pentru reprezentarea ontologiilor, de la începuturile domeniului și până în prezent.

A doua (și cea mai importată) secțiune a capitolului tratează *domeniul Logicilor Descriptive* (DL). Acestea sunt niște formalisme bazate pe logică pentru reprezentarea cunostințelor dintr-un domeniu de aplicație derivate din Logica de Ordinul 1 (eng. *First Order Logic* - FOL) având drept obiectiv să păstreze procesele de raționament asupra datelor decidabile (spre deosebire de FOL). Fiecare limbaj membru al familiei DL se obține dintr-unul deja existent prin adăugarea de noi constructori de concepte și/sau roluri pentru a putea reprezenta noi tipuri de informații, cel mai primitiv fiind AL (Attributive Language), care conține doar 3 constructori, anume cel de intersecție, restricții de valoare și existențialitate (∩,∀,∃). Aici am prezentat o tabelă cu 12 dintre cele mai importante limbaje DL create în literatură de către marii savanți ai domeniului. În finalul secțiunii 2 am prezentat care sunt cele mai importante domenii de aplicație ale DL-urilor, cu accentul bineînțeles pus pe limbajele de ontologii. Ontologiile de calitate au un rol fundamental pe Web-ul Semantic, iar construcția, integrarea și evoluția lor depind de existența unor semantici bune și tool-uri de raționament puternice. Pentru că DL-urile le dețin pe amândouă le face niște candidați ideali pentru crearea limbajelor de ontologii. Aici am prezentat trei dintre cele mai importante astfel de limbaje existente astăzi: DAML+OIL, OWL și OWL2, sintaxele și semanticile lor, limbajele DL pentru care au fost create, precum și o serie de analize și comparații între capabilitățile oferite.

După ce în secțiunea 2 am vorbit despre DL și bazele de cunoștințe logice, în secțiunea trei am înaintat și am discutat despre următoarea caracteristică importantă a SW, anume procesele de efectuare de raționament și inferență. Există două tipuri de sarcini de inferență în bazele de cunoștințe DL: standard și non-standard. Cele mai comune sunt satisfiabilitatea și incluziunea, și acestea fac parte din categoria celor standard.

După ce am prezentat sarcinile are sens să vorbesc și despre algoritmii creați care efectuează aceste sarcini și de sistemele la scară industrală dezvoltate în scopul efectuării de raționament asupra bazelor de cunoștințe DL. Accentul a fost pus pe algoritmii *tableaux,* care sunt cea mai importantă tehnică pentru realizarea sarcinilor de raționament în DL. Pentru a demonstra modul cum lucrează un algoritm Tableaux am furnizat un exemplu pentru calculul satisfiabilității unui concept din limbajul ALCN, pentru a explica cititorului funcționalitatea acestei tehnici. Celelalte două metode de rationament, anume cea bazată pe incluziune structurală și cea pe automate, nefiind la fel de importante (și mai vechi) decât tehnica *tableaux,* au fost discutate mai pe scurt. Finalul secțiunii cuprinde o listă cu rezultate de raționare din sarcinile standard pentru câteva din limbajele DL, culese din resursele din literatură citite de către mine în realizarea acestui capitol.

Ultima secțiune a acestui capitol important (care este și cel mai amplu al tezei) este, așa după cum era de așteptat, un studiu literar în care am relatat cele mai importante resurse citite de mine din literatura despre domeniul Logicilor Descriptive.

Capitolul se încheie cu secțiunea de concluzii finale, care rezumă conținutul acestuia.

Capitolul 3 – Sistem Raționator pentru Obținerea Saturației Bazelor Logice de Cunoștințe

În acest capitol se propune un sistem de efectuare a raționamentului asupra unei baze logice de cunoștințe reprezentată în sintaxa logicii de ordinul 1 (FOL). Baza de cunoștințe, pe

lângă cele uzuale (fapte și reguli de deducție) mai conține și un tip special, constrângeri, care exprimă condițiile pe care trebuie să le indeplinească datele pentru a fi puse în bază. Acesta este ceea ce diferențiază sistemul nostru de cele existente în prezent.

Capitolul începe cu o secțiune în care se prezintă chestiunile generale ale bazelor logice de cunostințe și a tuturor procedeelor care se vor folosi pe parcurs la construirea sistemului raționator propus, cum sunt cele de *substituție, homomorphism, derivare, saturare, etc.*

Secțiunea 2 face un studiu din literatură despre proiecte existente, iar singurele sisteme cu o funcționalitate cât-de-cât apropiată de nevoile mele sunt framework-urile de reguli business dezvoltate în limbajele de programare OO, dintre care: Drools, EasyRules, RuleBook, Jrules (le-am considerat pe cele din Java deoarece și sistemul meu l-am creat tot in acest limbaj).

Secțunea 3 discută proiectarea iar secțiunea 4 implementarea sistemului propriu-zis, în care am construit modele ale proceselor sub formă de diagrame UML (clase, activitate, obiecte), am explicat cum se reprezintă entitățile și baza de cunoștințe în mediul orientat-obiect (reprezentarea sa logică fiind în FOL) și am detaliat funcționarea sistemului, unde am explicat cum se efectuează calculele asupra bazei de cunoștințe (deducerea de noi fapte prin aplicarea regulilor, verificarea constrangerilor din nivelul ontologic, etc). Tehnicile folosite la implementare au fost *Backtracking* și *Legare Înainte* (eng. *Forward Chaining*). Am oferit cititorului exemple concrete de realizare a procesului de deducție în mediul OO (aplicare a colectiei de obiecte de tip reguli asupra faptelor) pentru a pune în vedere mai bine procesele de efectuare a calculelor din sistem.

Capitolul se încheie cu o secțiune de experimente efectuate cu sistemul și comparații cu alte tehnici similare. Rezultatele pun în evidență superioritatea sistemului propus în ceea ce privește numărul de calcule reduse efectuate datorită utilizării unor tehnici de optimizare a cautării în spațiul soluțiilor posibile, cum este Backtracking, precum și metode de optimizare folosite în procesul de deducție logică, cum este Forward Chaining, toate acestea fiind explicate pe larg în capitolul din teză, precum și să se ofere cititorului referințe în literatură unde poate afla mai multe despre ele.

Capitolul 4 – Sisteme de Detecție și Prevenție a Intruziunilor

Acest capitol reprezintă începutul celei de-a doua părți a tezei mele destinată securității cibernetice și tehnologiilor utilizate la construirea sistemelor de detecție și prevenție a atacurilor (IDPS), dintre acestea numărându-se bineînțeles și cele ale Web-ului Semantic, care reprezintă domeniul de bază al cercetării mele.

Capitolul începe cu o secțiune introductivă în care am prezentat cele mai importante domenii care necesită apărare contra atacurilor cibernetice, am spus ce sunt sistemele IDPS și am enumerat ce tehnologii folosesc în procesul de detecție a atacurilor, care sunt fazele prin care au trecut în evoluția lor, etc.

Secțiunea 2 discută despre componentele tipice din componența unui sistem IDPS, care sunt rolurile fiecăreia în procesul de detecție. Pentru asta am creat modele sub forma de figuri pentru o mai bună prezentare a chestiunilor discutate, înspre a fi mai bine înțelese de către cititor.

În secțiunea 3 a capitolul am discutat despre capabilitățile de securitate ale sistemelor IDPS, cele 4 mari responsabilități ale lor sunt: culegerea datelor de pe rețele, jurnalizare, detecție și (uneori) prevenție.

În secțiunea 4 am prezentat tipul de sisteme de detecție cel mai important și pe care îl voi considera și eu mai departe în teza de față, cel al *IDPS-urilor pentru rețele*. Am reluat ceea ce am discutat în precedentele două secțiuni pentru cazul concret al acestora, și anume componente și capabilități de securitate, toate mapate la cazul detecției atacurilor din rețele de

calculatoare. Am prezentat apoi care sunt capabilitățile de management al IDPS-urilor pentru rețele, care se găsesc sub sarcina administratorilor de securitate. Administratorii au cele mai importante sarcini de efectuat la găsirea unei soluții de securitate, așa cum se afirmă în toate sursele din literatura de specialitate. Aceștia au trei mari sarcini: implementare, operare și mentenanță. Prima fază spune că după ce un produs IDPS de rețea a fost ales, administratorii trebuie să creeze o arhitectură, să testeze componentele în izolare, să securizele componentele și, în final să desfășoare (eng. *deploy*) IDPS-ul. Fiecare din aceste faze a fost discutată pe larg în capitol, împreună cu exemple concrete și modele grafice (figuri) care să expună mai bine vizual procedele. Următoarea etapă se referă la posibilitatea de operare a IDPS-ului de către administratori prin intermediul consolei sale vizuale. Unele sisteme au și controale din linia de comandă – CLI care, spre deosebire de interfețele utilizator (GUI) care în mod normal sunt folosite pentru controlul de la distanță al senzorilor și serverelor, acestea sunt folosite de obicei într-un context local. Consola permite administratorilor să configureze și să actualizeze componentele IDPS (senzorii și serverele de management și baze de date), precum și să monitorizeze stările prin care trec. Marea majoritate a IDPS-urlor permit administratorilor crearea de conturi pentru fiecare tip de utilizator (admin sau user simplu) și să ofere numai privilegiile necesare pentru rolul fiecăruia. Gradul de control oferit poate să difere de la un produs la altul, în funcție de gradul de performanță și implementare al lor. Ultima fază este cea de mentenanță, în care administratorii trebuie să efectueze sarcini precum: verificarea la intervale de timp bine stabilite funcționarea sistemului, aplicarea de actualizări software a componentelor primite de la vânzători, primirea de știri de la furnizori despre probleme de securitate ale IDPS-ului etc. Update-urile de software pot să afecteze doar o parte a componentelor sau chiar întreg sistemul, incluzând senzori, servere de management și console. Administratorii trebuie să valideze integritatea update-urilor înainte de a le aplica deoarece există posibilitatea de a fi modificate sau înlocuite în mod neintenționat.

Capitolul 4 se încheie cu obișnuita secțiune de concluzii în care se rezumă conținutul și se prezintă cititorului contribuțiile aduse de cercetarea mea.

Capitolul 5 – Tehnologii de Inteligență Artificială și Web Semantic folosite la Construcția Sistemelor de Securitate Cibernetică

Acest capitol este destinat ca unul de review în care se fac o serie de discuții și analize cu privire la tehnologiile noi și inovative folosite la construirea sistemelor de securitate cibernetică. Aceste tehnologii sunt împrumutate în mod deosebit din domeniul Inteligenței Artificiale (IA), acolo unde se găsesc cele mai capabile metode și algoritmi pentru realizarea diferitor tipuri de calcule din domenii variate, iar mai nou și din nou apărutul domeniu al Web-ului Semantic care, așa cum a fost discutat până în acest punct, nu este foarte diferit de cel al IA. Capitolul este constituit din 2 părți. Prima tratează folosirea metodelor de IA în detecția intruziunilor iar a doua cele de Web Semantic.

Secțiunea 1 prezintă 5 tehnologii din IA care își au cea mai mare utilitate în domeniul detecției și prevenției atacurilor, și anume:

- Algoritmi Evolutivi, din a cărei clasă fac parte Algoritmii Genetici, Programarea Genetică, Evoluție Gramaticală
- Logica Fuzzy
- Clustering și puncte staționare
- Rețele Neuronale Artificiale
- Data Mining.

Pentru fiecare dintre aceste cinci tehnologii din IA au fost scose în evidență rolul pe care îl pot juca în procesul de detecție a atacurilor. Astfel, Algoritmii Genetici sunt folosiți la proiectarea

automată a modelelor, clasificarea și optimizarea datelor, selecția trăsăturilor, etc. Logica Fuzzy este o tehnică care este utilă în special în analiză și raționament asupra seturilor de date incomplete sau incerte, ceea ce le face niște tool-uri importante în analiza și evaluarea riscurilor de atacuri. Tehnologia de Analiza clusterelor și puncte solitare își găsește utilizare cel mai mult în detecția atacurilor bazate pe anomalii, în care primele reprezintă datele normale iar cele din urmă sunt datele despre atacuri. Ele detectează intruziunile din datele de audit brute (neprelucrate anterior), ceea ce face ca efortul necesar la reglarea IDS-ului să fie mai mic. Tehnica ANN-urilor constituie nucleul științei Machine Learning și, așa cum se poate ușor da seama, își poate găsi aplicații în domenii numeroase și variate datorită numeroaselor capabilități pe care le deține, cum ar fi luarea deciziilor corecte, recunoașterea formelor și tiparelor, etc. În domeniul detecției intruziunilor rețelele neuronale sunt folosite în detecția bazată pe anomalii la crearea și învățarea profilelor de activități benigne din datele de trafic brute și efectuează detecția prin clasificarea noilor evenimente pe baza profilelor create. Ultima tehnică discutată în prima secțiune, cea a Data Mining, este utilă în domeniul detecției atacurilor în special în analiză, mai exact este axată pe reducerea dimensionalităților, clasificare și clustering, având drept rezultat, după părerea unor cercetători renumiți ai domeniului, creșterea vitezei și acurateții din detecție, precum și întărirea securității proprii a sistemului. Această primă secțiune, pe lângă discuțiile ample și analizele făcute în acest domeniu, a mai propus și niște modele ale proceselor de detecție pe baza tehnologiilor AI discutate, modele care au fost create sub formă de figuri grafice, ca scheme bloc ale proceselor, diagrame UML de activitate, pentru a pune în vedere și a face citirorul să înțeleagă mult mai bine cum decurg lucrurile, spre deosebire de descrierea textuală simplă.

Secțiunea 2 a capitolului prezintă tehnologiile nou apărute ale Web-ului Semantic care își găsesc utilizare în domeniul IDS-urilor, dintre acestea cele mai importante sunt cele de "conținut", "ontologie" sau "multi-agenți". Fiecare metodă de securitate care se bazează pe noțiunea de conținut poate utiliza tehnologii din domeniul SW, iar sistemele de detecție a intruziunilor sunt un bun exemplu. Unii savanți renumiți ai Științei Calculatoarelor au deschis o nouă ramură în domeniul securității informației, aceea a folosirii ontologiilor împreună cu avantajele lor. Ei au afirmat că: "ontologiile sunt o nouă paradigmă extrem de promițătoare în aria securității informației prin intermediul cărora avem o unealtă de clasificare a evenimentelor nelimitată". Secțiunea a propus două tabele în care se prezintă câteva din cele mai importante sisteme IDS comerciale și de cercetare culese de mine din literatura de specialitate citită, care folosesc în procesele de detecție tehnici din AI și SW.

Capitolul mai are și o secțiune *state-of-art* în care am prezentat cele mai importante lucrări citite de catre mine în realizarea sa. S-a încheiat cu secțiunea de concluzii aferentă.

Capitolul 6 – Ontologie de Securitate Cibernetică în Rețele de Calculatoare

În acest capitol am propus un model al domeniului securității cibernetice sub forma unei ontologii care poate fi folosită de către sistemele IDS pentru a detecta natura evenimentelor posibil periculoase. Ontologia își propune să descrie domeniul atacurilor de calculatoare capturând cele mai importante concepte și relațiile dintre ele. Pentru construcția și evaluarea sa am folosit metodologii state-of-art ale literaturii, precum limbajele de ontologii ale Web-ului Semantic create de W3C (OWL2), sau SWRL.

Secțiunea 1 este o introducere care prezintă și definește noțiunile importante folosite în capitol, ca tipurile de IDS-uri, fazele prin care au trecut în evoluție, așa cum se specifică în literatura de domeniu, ce sunt ontologiile și care este rolul lor în construcția sistemelor de securitate cibernetică. Secțiunea 2 este un *state-of-art* cu cele mai importante lucrări din literatură citite de mine în realizarea acestui capitol.

Secțiunea 3 descrie construcția modelului ontologic. Pentru aceasta, dintre numeroasele metodologii create în literatura de specialitate, am ales metodologia Ontology Development 101, care este un proces de construire a unei ontologii de domeniu în șapte pași, pe care i-am urmat și eu pentru modelul meu.

Secțiunea 4 prezintă evaluarea modelului propus. Pentru a testa ontologia am luat un firewall de nivel aplicație unde modelul ontologic este ținut în baza de cunoștințe de unde este accesat de către firewall prin intermediul inferențelor pentru a detecta natura evenimentelor nou apărute. Faptul că se bazează pe reguli semantice face modelul să fie mai eficient în ceea ce privește timpii de execuție prin furnizarea unei reduceri substanțiale în stațiul de căutare și rezultând niște rate mai mici de pozitivi falși. Pentru evaluare am folosit metodologia OntoClean, care are în total 15 criterii de evaluare, așa cum este afirmat în literatură, eu însă am ales 8 pe care le-am considerat a fi cele mai importante pentru cazul meu.

În secțiunea 5 am arătat și am explicat cum este încorporată ontologia înăuntrul firewall-ului IDS și cum este accesată de către acesta în procesul de detectare a naturii situațiilor nou apărute. Am lucrat pe un exemplu de atac de tip Cross Site Scripting și am arătat fiecare fază prin care trece această dată prin sistemul IDS, de la intrare și până la jurnalizarea ei ca fiind un atac cunoscut.

Secțiunea 6 prezintă experimentele pe care le-am făcut cu sistemul nostru de detecție. Pentru testate am folosit setul de date despre atacuri Kyoto2006+, care este cel mai bun set pentru acest scop. Sistemul meu bazat pe tehnologii semantice le-am comparat cu cele ale două sisteme din literatură, anume Snort și ModSecurity, rezultatele obținute după 3 metrici, anume *Precizie, Recall și F-metric* au arătat o ușoară superioritate a sistemului meu în ceea ce privește eficiența în viteza de detecție și rata de alarme false.

Capitolul 7 – Sistem de Detecție a Intruziunilor bazat pe Tehnologiile Web-ului Semantic și Inteligenței Artificiale

În acest capitol am propus un sistem IDS distribuit pentru rețele care folosește cele mai recente și inovative tehnologii împrumutate din domeniile Inteligenței Artificiale și Web-ului Semantic cu scopul de a crește eficiența detecției atacurilor reale și de a scădea pe cea a erorilor. Principalul obiectiv al acestui IDS distribuit este să rezolve problemele celor centralizate implicând o arhitectură multi-agent în care fiecărui agent îi revin anumite sarcini de realizat. Baza de cunoștințe este reprezentată sub forma unei ontologii care capturează termenii din domeniul intruziunilor și relațiile lor și este utilizată de către agenți pentru a detecta natura evenimentelor recepționate (date normale sau atacuri). Una din cele mai importante funcții ale sistemului este posibilitatea de a detecta atacuri necunoscute până atunci (Ziua-0), care este implementată de către agentul care efectuează detecție bazată pe anomalii.

Secțiunea 1 este o introducere în care am prezentat noțiunile importante care vor fi folosite pe parcursul capitolului. De asemenea au fost puse în vedere dezavantajele IDS-urilor centralizate, cum ar fi cea a punctului singular de cădere (reprezentat de nodul central), și anume dacă acest nod este atacat atunci întregul IDS cade, sau problema supraîncărcării rețelei datorată transferului informației de la toți senzorii plasați in diverse locuri pe rețea și până la nodul central; aceasta mai poartă numele de problema scalabilității sistemelor.

În secțiunea 3 am prezentat arhitectura pe componente a sistemului meu, și am detaliat fiecare agent ce funcție are. În secțiunea 4 am vorbit despre funcțiile ontologiei pentru atacuri, având rolul unei baze de cunoștințe mai optimizată și care să conțină mai multă inteligență în analiza conținutului. Aceasta conține regulile care permit o reprezentare semantică ce favorizează procesele de inferență și raționament. Regulile sunt extrase din ontologie cu

ajutorul limbajului standard al Web-ului Semantic SWRL, fapt ce extinde ontologia și îi îmbogățește semanticile prin capabilități de raționament deductiv. Apoi am dat un exemplu de regulă SWRL a ontologiei, ce structură și sintaxă are și cum este inferată de către sistem pentru a produce consecințe. Secțiunea 5 discută despre algoritmul de clusterizare folosit în detecția anomaliilor și care este unul bazat pe binecunoscutul K-means din Data Mining.

Acest capitol conclude, așa cum este și normal, cu o secțiune în care am testat sistemul meu. Am evaluat performanța IDS-ului în funcție de două metrici: scalabilitate și detecție. Am simulat atacurile de rețele cu ajutorul tool-ului Metasploit 3.5.1, care a generat următoarele atacuri: Smurf, Backdoor, Spyware Put Hijacker, Nmap TCP Scan, Finger User, RPC Linux Statd Overflow, DNS Zone Transfer, HTTP IIS Unicode. În timpul procesului de evaluare am comparat rezultatele sistemului meu cu cele ale IDS-ului centralizat Snort și ale celui distribuit MONI, acestea arătând superioritatea arhitecturii propuse.

3. Concluzii, contribuții personale, valorificarea cercetării prin publicații

Cercetarea realizată în această teză, așa cum a fost afirmat începând încă cu titlul său și de numeroase ori pe parcurs, și-a avut drept obiective să contribuie și să ajute la dezvoltarea noii generații de Web, apărută la începutul anilor 2000, care se dorește a fi una a datelor structurate universale și a execuției automate a proceselor, care, așa după cum ne putem da seama și din viața de zi cu zi, reprezintă viitorul în industrie. Peste tot vedem mașini și aparate care au luat locul oamenilor în realizarea sarcinilor, cum ar fi cele pentru eliberarea biletelor de călătorie din gări, autogări, sisteme electronice de plată a călătoriei în mijloace de transport cu cardul, fără a mai necesita bilete, sisteme de plată automatizate din supermarket-uri/hipermarket-uri care în ultimii 2-3 ani au luat cu asalt aceste magazine, și exemplele pot continua. Acest lucru se întâmplă și cu World Wide Web-ul, cel mai important și cel mai folosit serviciu din Internet.

Problemele abordate în această teză au urmărit să contribuie la dezvoltarea domeniului cercetat. Ele sunt de două feluri:

- recenzii ale diferitor arii din domeniul Semantic Web în care am prezentat, am definit noțiunile care stau la baza acelei arii, am realizat analize, comparații, am afirmat păreri personale despre acestea și am oferit cititorului, pentru fiecare noțiune/definiție, referințe în literatură unde poate găsi mai multă informație despre chestiunile discutate în capitol;
- construirea de sisteme noi și inovative prin combinarea de tehnologii luate din alte domenii ale Științei Calculatoarelor cu scopul de a rezolva problemele existente dintr-un anumit domeniu.

Prima clasă de contribuții a fost evidențiată în capitolele 1, 2, 4, și 5, acolo unde mi-am concentrat atenția asupra următoarelor domenii: Formalismele logice de reprezentare a cunoștințelor, Sistemele de apărare cibernetice, Tehnologii inovative folosite la construirea sistemelor IDS, și, bineînțeles, prezentarea domeniului principal al cercetării de față, și anume Web-ul Semantic.

Din cealaltă categorie fac parte capitolele 3, 6 și 7, acolo unde am construit sisteme și metode noi de rezolvare a problemelor deschise din anumite domenii. În primul am creat un sistem raționator pentru a realiza deducție asupra unui tip special de baze de cunoștințe compuse din trei tipuri de cunoștințe (fapte, reguli, constrângeri). În al doilea am creat un model al domeniului securității calculatoarelor propus sub forma unei ontologii care își propune să

captureze cât mai multă informație de domeniu împreună cu relațiile dintre concepte, ducând astfel la sporirea capacității de detecție a sistemului IDS și la scăderea ratei de erori ale sale. În ultimul capitol am dorit să continui munca depusă în precedentele două capitole, care sunt toate despre domeniul securității cibernetice. Aici am construit un sistem de detecție a atacurilor în rețele de calculatoare care folosește tehnologii noi și foarte eficiente din domeniile Inteligenței Artificiale și Semantic Web cu scopul de a surveni problemele cu care se confruntă IDS-urile din ziua de azi, în special rata scăzută de detecție a atacurilor reale, greșeli de detecție, neputința detectării atacurilor nemaivăzute până atunci (Ziua-0), precum și întărirea securității proprii a sistemului.

Concluzia mea personală din finalul acestei teze este că am îndeplinit obiectivele pe care mi le-am stabilit inițial pentru cercetarea de față, am adus contribuții la domeniul cercetat (cele enumerate mai sus), lucru dovedit și prin cele 15 publicații internaționale pe care le-am realizat pe durata studiilor de doctorat. Trei dintre lucrări sunt despre Logicile Descriptive, care au fost discutate în capitolul 2 al tezei, una este despre domeniul cercetării mele, Web-ul Semantic, iar încă una despre alt domeniu, înrudit cu Semantic Web și care a cunoscut o evoluție relativ similară cu acesta în ultimele decade, anume Internetul Lucrurilor. Alte două lucrări tratează domeniul Inteligenței Artificiale, în care am analizat cea mai importantă funcție cognitivă umană care ne deosebește și de cel mai evoluat alt animal: comunicarea (verbală sau scrisă).

Dintre lucrările care au propus sisteme și/sau tehnici noi amintesc crearea unei ontologii de domeniul securității cibernetice, un sistem de raționare pentru un tip special de bază de cunoștințe și un sistem IDS distribuit bazat pe tehnologii luate din domeniile Inteligenței Artificiale și Semantic Web. Lista completă a publicațiilor mele științifice poate fi consultată în teză la secțiunea de Referințe, pentru mai multe detalii.

BIBLIOGRAFIE

[1] F.Abdoli, M.Kahani; *Ontology-based Distributed Intrusion Detection System*, Proceedings of 14th International Computer Conference (CSICC), Teheran, Iran (2009)

[2] N.Agarwal, Z.Hussain; *A Closer Look at Intrusion Detection Systems for Web Applications*, Hindawi Journal on Security and Communication Networks, vol.2018

[3] G.Antoniou, P.Groth, F.v.Harmelen, R.Hoekstra; *A Semantic Web Primer*, MIT Press (2012)

[4] M.d'Aquin, E.Motta, M.Sabou, S.Angeletou; *Toward a New Generation of Semantic Web Applications*, IEEE Intelligent Systems, vol.23,no.3, pp.20-28 (2008)

[5] A.Aviad, K.Wecel, W.Abramowicz; *The Semantic Approach to Cybersecurity – Towards Ontology-based Body of Knowledge*, Proceedings of the 14th European Conference on Cyber Warfare and Security (ECCWS,2015)

[6] F.Baader, D.McGuiness, D.Nardi, P.Schneider; *The Description Logic Handbook: Theory, Applications and Implementations*, Cambridge University Press, New York, USA (2003)

[7] T.Berners-Lee; *WWW: Past, Present and Future*, IEEE Computer Magazine, vol.29 no.10 (1996)

[8] T.Berners-Lee, R.Cailliau, J.Groff; *The World Wide Web*, Communications of the ACM (1994)

[9] T.Berners-Lee, M.Fischetti; *Weaving the Web*, Ed. Harper-Collins, San Francisco (1999)

[10] T.Berners-Lee, J.Hendler, O.Lasilla; *The Semantic Web*, Feature Article, Scientific American (2001)

[11] T.Berners-Lee, R.Swick; *Semantic Web Development,* Technical Report AFRL-IF-RS-TR-2006-294, New York (2006)

[12] T.Berners-Lee, J.Hendler, *From the Semantic Web to Social Machines: A research challenge for AI on the World Wide Web*, Journal of Artificial Intelligence, vol.174 no.2 (2010)

[13] I.Brahmi, S.Yahia, H.Aouadi, P.Poncelet; *Towards a Multi-agent based Intrusion Detection System using Data Mining Approaches*, Proceedings of 7th International Workshop on Agents and Data Mining Interaction (ADMI), pp.173-194, Taipei,Taiwan (2011)

[14] N.Dalwadi, B.Nagar, A.Makwana; *Semantic Web and Comparative Analysis of Inference Engines*, International Journal of Computer Science and Information Technologies, vol.3, pp.3843-3847 (2012)

[15] F.Donini, F.Massacci; *Exp-Time Tableaux for ALC,* Artificial Intelligence, vol.124, no.1, pp.87-138 (2000)

[16] D.Fensel, J.Hendler, H.Lieberman, W.Wahlster; *Spinning the Semantic Web: Bringing the Web to its Full Potential*, MIT Press (2003)
[17] F.v. Harmelen, V.Lifschitz, B.Porter; *Handbook of Knowledge Representation*, Foundations of Artificial Intelligence, Elsevier (2008)
[18] I.Horrocks, P.Schneider; *Knowledge Representation and Reasoning on the Semantic Web: OWL*, Handbook of Semantic Web Technologies, pp.365-398, Springer (2011)
[19] J.Pan; *Desciption Logics: Reasoning Support for the Semantic Web*, PhD thesis, Univ. of Manchester (2004)
[20] A.Razzaq, Z.Anwar, F.Ahmad, K.Latif, F.Munir; *Ontology for Attack Detection: An Intelligent Approach to Web Application Security*, Computers&Security, vol.45, pp.124-146, Elsevier (2014)
[21] K.Scarfone, P.Mell; *Guide to Intrusion Detection and Prevention Systems (IDPS)*, Recommendations of the National Institute of Standards and Technology (NIST), Special Publication (2007)
[22] M.Schmidt-Schaus, G.Smolka; *Attributive Concept Descriptions with Complements*, Journal of Artificial Intelligence, pp.1-27 (1991)
[23] N.Shadbolt, T.Berners-Lee, W.Hall; *The Semantic Web Revisited*, IEEE Intelligent Systems, vol.21, pp.96-101 (2006)
[24] J.Song, H.Takakura, Y.Okabe; *Statistical Analysis of Honeypot Data and Building of Kyoto 2006+ Dataset for NIDS Evaluation*, Proceedings of the First Workshop on Building Analysis Datasets and Gathering Experience Returns for Security, pag.29-36, Salzburg, Austria (2011)
[25] Y.Zhu; *Attack Pattern Ontology: A Common Language for Cyber-Security Information Sharing*, TU Delft Publication, Master Thesis (2015)