Umjetna inteligencija u sigurnosnim sustavima

Balta, Nikola

Sveučilište Jurja Dobrile u Puli

Fakultet informatike

NIKOLA BALTA

UMJETNA INTELIGENCIJA U SIGURNOSNIM SUSTAVIMA

Završni rad

Pula, kolovoz, 2020.

Sveučilište Jurja Dobrile u Puli

Fakultet informatike

NIKOLA BALTA

UMJETNA INTELIGENCIJA U SIGURNOSNIM SUSTAVIMA

Završni rad

JMBAG: 0303054098, redoviti student

Studijski smjer: informatika

Predmet: Poslovni informacijski sustavi

Znanstveno područje: Društvene znanosti

Znanstveno polje: Informacijske i komunikacijske znanosti

Znanstvena grana: Informacijski sustavi i informatologija

Mentor: doc. dr. sc. Darko Etinger

Komentor: doc. dr. sc. Nikola Tanković

Pula, kolovoz, 2020.

IZJAVA O AKADEMSKOJ ČESTITOSTI

Ja, dolje potpisani Nikola Balta, kandidat za prvostupnika informatike ovime izjavljujem da je ovaj Završni rad rezultat isključivo mojega vlastitog rada, da se temelji na mojim istraživanjima te da se oslanja na objavljenu literaturu kao što to pokazuju korištene bilješke i bibliografija. Izjavljujem da niti jedan dio Završnog rada nije napisan na nedozvoljen način, odnosno da je prepisan iz kojega necitiranog rada, te da ikoji dio rada krši bilo čija autorska prava. Izjavljujem, također, da nijedan dio rada nije iskorišten za koji drugi rad pri bilo kojoj drugoj visokoškolskoj, znanstvenoj ili radnoj ustanovi.

Student

_____

U Puli, _____, _____ godine

IZJAVA

o korištenju autorskog djela

Ja, Nikola Balta, dajem odobrenje Sveučilištu Jurja Dobrile u Puli, kao nositelju prava iskorištavanja, da moj završni rad pod nazivom **„Umjetna inteligencija u sigurnosnim sustavima“** koristi na način da gore navedeno autorsko djelo, kao cjeloviti tekst trajno objavi u javnoj internetskoj bazi Sveučilišne knjižnice Sveučilišta Jurja Dobrile u Puli te kopira u javnu internetsku bazu završnih radova Nacionalne i sveučilišne knjižnice (stavljanje na raspolaganje javnosti), sve u skladu s Zakonom o autorskom pravu i drugim srodnim pravima i dobrom akademskom praksom, a radi promicanja otvorenoga, slobodnoga pristupa znanstvenim informacijama.

Za korištenje autorskog djela na gore navedeni način ne potražujem naknadu.

U Puli, _____ (datum)

Potpis

_____

SADRŽAJ

1. UVOD

Kao što se može zaključiti iz naslova rada, ideja iza ove teme je bila pokazati utjecaj umjetne inteligencije u sustavima zaštite računala. Međutim, tokom svog rada i istraživanja o toj temi sam odlučio u radu ne samo prikazati ulogu umjetne inteligencije u zaštiti sustava, nego omogućiti čitatelju da kroz rad stekne uvid ne samo u taj dio komponiranja umjetne inteligencije u sigurnosne sustave, nego pružiti i uvid u sve grane koje su povezane sa time. Da kroz čitanje rada dobije nekoliko informacija o umjetnoj inteligenciji, zaštiti sustava, tipovima napada i zatim prikazati kako se i u kojim oblicima u sve to uklapa umjetna inteligencija.

U uvodnom dijelu sam ispričao djelomičnu povijest umjetne inteligencije, prikazao njezine korijene i samu ideju o stvaranju strojeva koji imaju sposobnost ljudskog razmišljanja, da ta ideja seže u daleku prošlost čovječanstva. Grana umjetne inteligencije koja je predmet ovog rada jeste strojno učenje engl. *machine learning*. Strojno učenje predstavlja način učenja kojemu je cilj imitirati čovjekovo učenje. Nakon predstavljanja strojnog učenja, idući dio je bio sama sigurnost sustava, što je to i od čega sve imamo potrebu štititi sustave, uređaje i ljude koji se iz bilo kojeg razloga, privatnog ili poslovnog koriste internetom.

Središnji i glavni dio rada se bavi istraživanjem strojnog učenja u zaštiti sustava. Za što se sve strojno učenje koristi u obrani sustava, što on doprinosi samim sustavima, i na koji način čovjek zajedno sa računalom može surađivati da se postignu do sada ne bih se usudio reći nemoguće stvari, ali samome čovjeku sigurno teže dostižne, bez pomoći umjetne inteligencije. Strojno učenje se najviše primjenjuje za automatsko otkrivanje prijetnji putem modela učenja. Računalu se daju upute kako prepoznati određene napade i ono na temelju tih napada dalje samostalno otkriva iste ili slične, sa ili bez potrebe za ljudskom asistencijom. U većini slučajeva potreba za ljudskom asistencijom nije nužna, ali bilo da se radi o istraživanjima ili o primjerima iz prakse, najbolji rezultati se postižu kada automatizirani dijelovi sustava koji su samostalni surađuju zajedno sa čovjekom, odnosno ljudima, tu se zapravo probijaju granice koje ni stroj, niti čovjek ne mogu sami prijeći.

U posljednjem dijelu sam se osvrnuo na tehniku *Honeypota*, jedan od najstarijih trikova IT-a. *Honeypot* je zapravo zamka za napadače čija je svrha prikupljanje informacija. Kroz to poglavlje sam htio prikazati da uz pomoć umjetne inteligencije nečemu „starome“ možemo dodati nešto novo i prikazati ga u potpuno drugome svjetlu.

Na samome kraju sam spomenuo primjenu umjetne inteligencije i za „drugi tim“, za napadače. Umjetna inteligencija jeste inteligencija, ali je i dalje umjetna i napravljena od strane čovjeka, za koju svrhu se ona koristi ovisi o samome čovjeku. Umjetna inteligencija predstavlja tehnologiju koju su razvili ljudi, kao i svaki naš izum, što se može jasno vidjeti iz povijesti, može se koristiti u dobre, ali i u loše svrhe.

2. Povijest umjetne inteligencije

Područje umjetne inteligencije (*AI*) nastoji razumjeti inteligentne entitete. Razlika između drugih znanosti kao što su psihologija i filozofija, koje se također bave istraživanjem inteligencije, umjetna inteligencija se više bavi izgradnjom, to jest izradom inteligentnih entiteta i te iste nastoji razumjeti. Čak i u ranim fazama svoga razvoja umjetna inteligencija je proizvela mnoge značajne i impresivne proizvode. Iako se budućnost ne može predvidjeti, jasno je da računala koja posjeduju inteligenciju, bilo na ljudskoj razini, ili na nekoj višoj, sa sigurnošću se može reći da će takva računala imati veliki utjecaj na ljudsku svakodnevnicu i budućnost čovječanstva. Umjetna inteligencija kao znanstvena disciplina je jako mlada, i spada u novije discipline. Naziv ove znanosti je utemeljen 1956. g.. U svojem nazivu sadrži riječ „inteligencija“ ali inteligencija je predmet istraživanja više od 2000 godina, filozofi su od davnina pokušavali odgonetnuti kako funkcioniraju učenje, sjećanje, razum i kako se svi ti procesi rade. Danas područje AI-a obuhvaća široko polje pod-područja od percepcije i logičkog razmišljanja, do specifičnih zadataka kao što su igranje šaha, dokazivanje matematičkih poučaka, dijagnosticiranje bolesti i mnoge druge. Što je zapravo umjetna inteligencija? Kako ju možemo definirati? Definicije umjetne inteligencije se mogu podijeliti u dvije sfere, prva je način razmišljanja i prosuđivanje a druga bi bila ponašanje.

Definicije umjetne inteligencije se mogu svrstati u četiri kategorije :

1. Sustavi koji razmišljaju kao ljudi
2. Sustavi koji razmišljaju racionalno
3. Sustavi koji se ponašaju kao ljudi
4. Sustavi koji se ponašaju racionalno

Umjetna inteligencija vuče korijene iz mnogih drugih znanosti kao što su filozofija, psihologija, matematika i računalno inženjerstvo. Temelji iz filozofije su bili pitanja razuma i razmišljanja, iz psihologije način razmišljanja i percipiranja svijeta i ponašanja ljudi, matematika je umjetnoj inteligenciji pridonijela u tri područja: računanje, logika i vjerojatnost. Prikazivanje računanja putem algoritama prvi je primijenio arapski matematičar al-Khowarazmi u 9. stoljeću. Vjerojatno najveću zaslugu ima računalno inženjerstvo, jer da bi umjetna inteligencija funkcionirala potrebna joj je osim „ljudske inteligencije“ i uređaj na kojem će raditi. Uređaj koji na

najbolji način pokazuje „inteligenciju“ je bilo i ostalo računalo. Umjetna inteligencija se uspjela razviti tako dobro i zbog softverske strane informatike koja joj je pružila operativne sustave, programske jezike i alate za pisanje modernih programa. Umjetna inteligencija je isto tako uzvratila razvojem mnogih ideja za informatiku, neke od njih bi bile razvoj prevoditelja, vezane liste kao strukture podataka, automatsko upravljanje memorijom i neki ključni temelji objektno orijentiranog programiranja i razvoj korisničkih sučelja.

Može se reći da umjetna inteligencija ima jako puno definicija koje ju opisuju, subjektivno definicija koja opisuje ovu znanost je „Field of computer science that studies how machines can be made to act intelligently.“ (Jackson, 1986.). Što bi u prijevodu bilo „Područje informatike koje proučava kako napraviti stroj koji je inteligentan“.

3. Strojno učenje

Strojno učenje je grana umjetne inteligencije koja se bavi gradnjom algoritama koji se oslanjaju na skup primjera neke pojave ili fenomena. Ti primjeri mogu biti prirodni, umjetni tj. da su napravljeni ljudskom rukom ili mogu biti generirani putem nekog drugog algoritma. Na slici jedan su prikazane grane umjetne inteligencije.

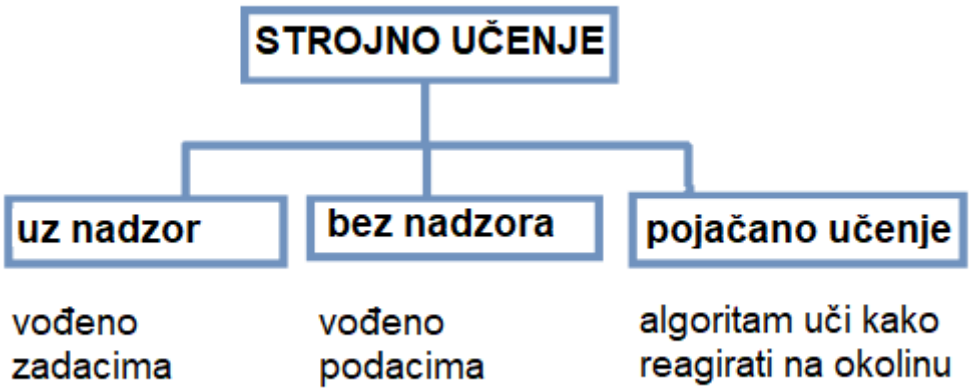
Slika 1 – Grane umjetne inteligencije

Strojno učenje se isto tako može definirati kao proces rješavanja praktičnih problema na način da se prvo skupljaju i grupiraju podaci, a zatim uz pomoć algoritama se napravi statistički model koji se bazira na tim podacima. Taj model se koristi za rješavanje nekog problema.

Umjetna inteligencija i strojno učenje se sve više i više isprepliću i povezuju danas unutar raznih industrija i aplikacija zbog naglog razvoja računala, kapaciteta za pohranu podataka i samih zbirki podataka. Strojno učenje uči računalo kako samostalno donositi odluke, razlikuje se od tradicionalnog programiranja po tome što se u tradicionalnom programiranju računalu daju eksplicitne instrukcije koje daju neke rezultate. Strojno učenje koristi statističke modele, matematičke optimizacije i rudarenje podataka (*data mining*), tu računala pokušavaju donositi odluke na temelju svog ponašanja.

Postoji nekoliko tipova strojnog učenja:

1. Učenje uz nadzor (*Supervised learning*)
2. Učenje bez nadzora (*Unsupervised learning*)
3. Pojačano učenje (*Reinforcement learning*)



Slika 2 – Tri kategorije strojnog učenja

3.1 Učenje uz nadzor (*Supervised learning*)

Kod učenja uz nadzor stroj uči uz uzorak podataka koji je definiran na takav način da govori stroju što ti podaci predstavljaju. Algoritam koji uči uz nadzor prima jednu ulaznu varijablu I (*input*) i izlaznu varijablu O (*output*) i algoritmi se koriste za kreiranje i učenje funkcija mapiranja (f) uz pomoć ulaza i izlaza, tj. *inputa* i *outputa*. Cilj algoritma je da postigne optimalnu funkciju mapiranja da za svaki novi ulaz (I) predvidi novi izlaz (O). Znači, algoritam prima kolekciju inputa tj. ulaza sa pripadajućim outputima tj. izlazima i algoritam uči tako da u danim ulazima, odnosno izlazima traži greške i modificira dani model po potrebi. Algoritmi učenja uz nadzor koriste uzorke za predviđanje vrijednosti kada primaju neoznačene podatke, odnosno podatke koji još nemaju svoju vrijednost. To se postiže uz pomoć klasifikacije, regresije i predviđanja. Klasifikacijski problem je kada varijabla izlaza spada pod neku vrstu kategorije npr. „crno“ ili „bijelo“ ili „pozitivan“ ili „negativan“. Problem regresije je kada varijabla izlaza ima neku stvarnu, mjerljivu vrijednost „kuna“ ili „visina“. Na temelju toga stroj bi trebao moći analizirati nove podatke i dati točne rezultate, odnosno odgovore. Učenje uz nadzor ima primjenu u medicinskoj dijagnostici i prepoznavanju govora.

3.2 Učenje bez nadzora (*Unsupervised learning*)

Kod učenja bez nadzora stroj uči uz pomoć podataka koji nisu označeni. Podaci kod ove vrste učenja imaju samo ulazne (I) tj. *input* podatke bez izlaznih (O) tj. *output* podataka. Cilj ovakvog učenja je modelirati građu podataka na takav način da se iz toga dobiju dodatne informacije o tim podacima. Algoritmi bi sami trebali otkriti strukturu podataka, sami doći do značenja tih podataka i iz njih donijeti samostalne zaključke i smisao samih podataka. Razlika od algoritama sa nadzorom i bez je ta da algoritmi bez nadzora nemaju povijest podataka kojoj mogu pristupiti da bi pretpostavili izlazne podatke. Računalo bez nadzora ne zna što podaci predstavljaju niti koje izlaze treba dati na temelju ulaza, ono treba samostalno shvatiti uzorke i strukturu ulaznih podataka koji nemaju nikakvu oznaku i doći do očekivanog izlaza. Problemi učenja bez nadzora se mogu svrstati u dvije kategorije, problemi grupiranja i problemi asocijacije. Problem grupiranja je takav da se pokušava naći nasljedno grupiranje podataka, kao što je grupiranje kupaca po kupovnom ponašanju. Kod problema asocijacije se pokušavaju naći pravila koja opisuju veliki dio podataka, npr.

kupci koji kupe proizvod „X“ kupe i proizvod „Y“. Primjer algoritma za učenje bez nadzora je klasifikacija filmova na „Netflixu“.

3.3 Pojačano učenje (*Reinforcement learning*)

Kod pojačanog učenja računalo surađuje sa svojom okolinom da ostvari neki cilj. Ima dosta sličnosti sa učenjem bez nadzora jer računalo uči koristeći neoznačene podatke, ali za razliku od učenja bez nadzora u pojačanome učenju računalo dobiva povratne informacije o izlaznim podacima. Računalo je u stanju percipirati stanje okoliša u kojemu se nalazi i ovisno o stanju može izvoditi određene zadatke. Različita stanja okoline daju različite rezultate. Postoje dva tipa pojačanog učenja a to su, pozitivno i negativno pojačano učenje.

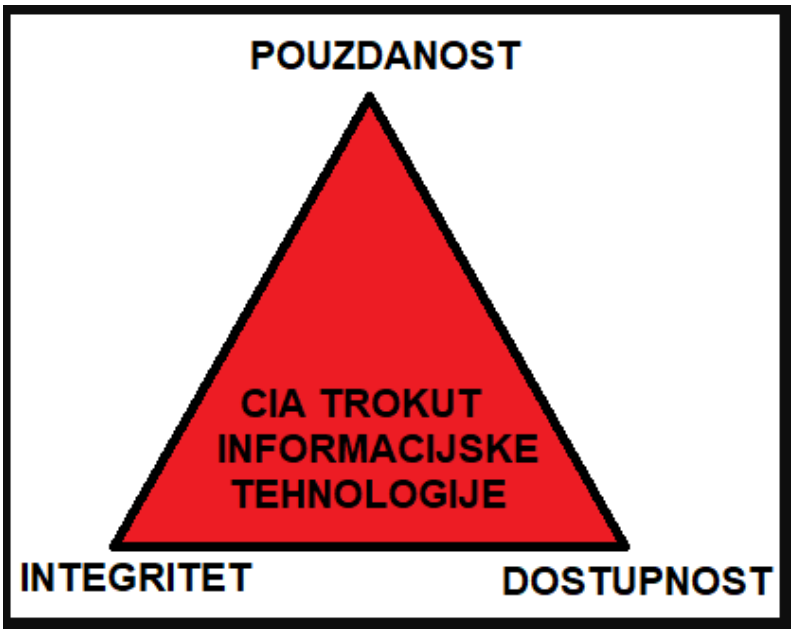
1. Pozitivno pojačano učenje–je definirano tako da kada se desi neki događaj na temelju nekog ponašanja računala, snaga i frekvencija tog ponašanja se povećava, ima pozitivan efekt na ponašanje.
Prednosti ovakvog učenja su : maksimizacija performansi i održavanje promjena na duže vrijeme.
Mane ovakvog učenja su : može doći do preopterećivanja stanja što dovodi do krivih rezultata.

2. Negativno pojačano učenje–je definirano tako da osnažuje ponašanje kada se izbjegne negativni uvjet. Prednosti ovakvog učenja: povećavanje ponašanja i pružanje manjka minimalnog standarda performansi. Mane ovakvog učenja su: daje dovoljno informacija za zadovoljavanje minimuma ponašanja.

Pojačano učenje se koristi u robotici za industrijsku automatizaciju.

4. Zaštita digitalnih sustava

Zaštita sustava je praksa koja se bavi osiguravanjem računalnih mreža, sustava i drugih digitalnih infrastruktura od zlonamjernih napada. Istraživači koji se bave ovim područjem nastoje očuvati tzv. *CIA* trokut informacijske tehnologije.



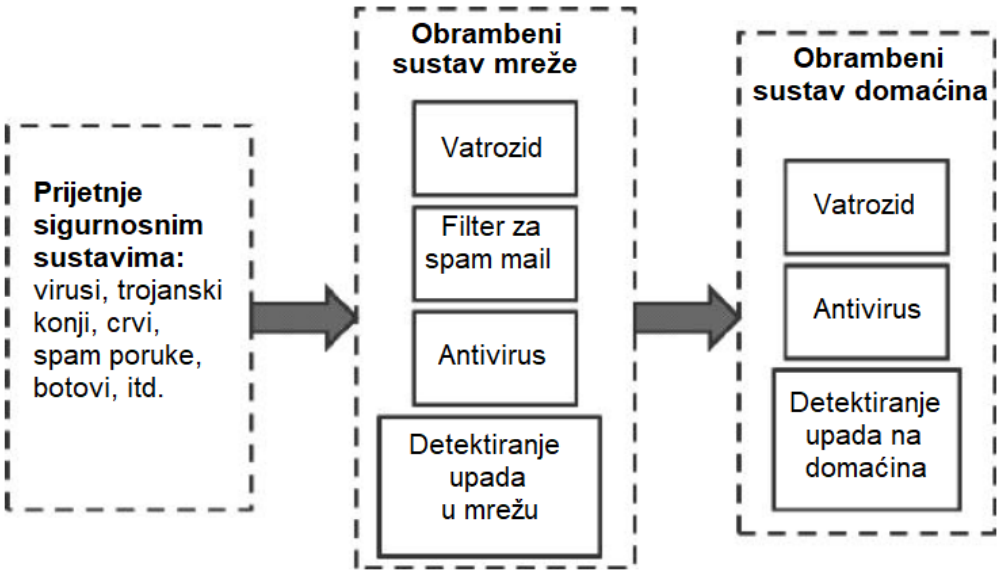
Slika 3 – CIA trokut informacijske tehnologije

Confidentiality (pouzdanost) – sprječavanje neovlaštenog pristupa podacima

Integrity (integritet) – sprječavanje neovlaštenog mijenjanja podataka

Availability (dostupnost) – sprječavanje neovlaštene nedostupnosti podataka

Sustavi zaštite kroz različite načine obrane štite računala i računalne mreže od napadača koji žele ukrasti osjetljive informacije njihovih korisnika (financijske, medicinske, privatne).



Slika 4 – Konvencionalni sustav digitalne zaštite

Iz slike 4 možemo vidjeti na koji se način digitalni sustavi zaštite brane od opasnosti u koje spadaju virusi, trojanski konji, crvi, spam i sl. Digitalni sustavi se brane od napada na dvije razine, jedna razina su obrane mreže, a druga obrane domaćina. Sustavi koji brane mrežu kontroliraju tok unutar mreže, vatrozid mreže, filtriraju spam, upravljaju antivirusnim sustavom i praćenjem neovlaštenog upada u mrežu. Sustavi koji brane domaćina upravljaju kontrolom dotoka podataka na radnoj stanici pomoću vatrozida, antivirusa i praćenjem neovlaštenog upada u mrežu domaćina. Konvencionalni pristup digitalnoj obrani su dizajnirani unutar vatrozida, alata za autentifikaciju i mrežnim serverima koji prate i blokiraju viruse i druge zlonamjerne napade. Npr. *Microsoft* operativni sustav ima ugrađeni „Kerberos“ kriptografski sustav. Taj sustav štiti korisnikove podatke. Iz istog razloga se na sustave instaliraju antivirusni programi. Ovi pristupi štite digitalne sustave.

Ali ovakav pristup ne funkcionira kod pokušaja zaštite aplikacija zbog grešaka prilikom dizajniranja i implementacije softvera i mrežnih infrastruktura. Aplikacijski sustavi se štite zakrpama (*patch*), ali ni to nije rješenje jer napadači konstantno pronalaze nove načine za iskoristiti greške u sustavima. Zato nije dovoljno graditi sustave koji štite od već otkrivenih i poznatih napada već se moraju primjenjivati razne metodologije viših razina za otkrivanje novonastalih digitalnih napada koje bi ojačale digitalne sustave obrane.

Sustavi obrane se mogu podijeliti na komponente kao što je vidljivo iz slike 4. Alati koji prate podatke npr. Libpcap za Linux sustave ili Winpcap za Windows prate tragove informacija preko mreže. Događaji se mogu podijeliti na događaje domaćina ili na događaje mreže, ovisno od kuda dolaze. Ako dolaze iz log datoteka onda pripadaju domaćinu, ako su iz mrežnog prometa onda pripadaju mreži. Događaj domaćina se sastoji od naredbi koje izvršava korisnik i od niza sistemskih poziva koje izvršava sustav koje pokreće neka aplikacija, npr. „Gmail“. Mrežni događaji uključuju podatke mrežnog prometa npr. niz paketa internet protokola (IP) ili transmisijskog protokola kontrole (TCP). Dio predprocesorskih podataka filtrira napade čije je potpise već naučio. Dio sustava koji izvlači svojstva se koristi u analizi događaja, ti događaji uključuju nizove sustavskih poziva, početno vrijeme i trajanje mrežnog toka, broj bajtova i paketa i sl. U dijelu za analiziranje mnoge metode detektiranja upada u mrežu se implementiraju radi

istraživanja ponašanja digitalne infrastrukture koja se možda je, možda nije prethodno pojavljivala, npr. praćenje zlonamjernog prometa. Odluke odgovora se izvršavaju kada se napad identificira.

Rješavanju problema zaštite digitalnih sustava se može pristupiti na dva načina, proaktivna rješenja i reaktivna rješenja. Proaktivni pristup predviđa i eliminira ranjivosti digitalnog sustava i uvijek je u stanju pripravnosti da brani sustav od napada. Da bi proaktivna rješenja radila potrebna im je autentifikacija korisnika npr. lozinka ili biometrijska identifikacija. Drugi način su reaktivna rješenja kao što su sustavi koji detektiraju upad u mrežu (*intrusion detection systems - IDSs*). *IDSs* detektiraju upade iz log datoteka i mrežnog toka podataka, procjenjuju štetu na sustavu i omogućavaju praćenje napadača i tako sprječavaju takve napade u budućnosti.

Slika 5 – Prilagodljivi obrambeni sustav

4.1 Tipovi digitalnih napada

Postoje razni oblici digitalnih napada. Neke vrste *ransomware* napada tzv. napadi za otkupninu u kojima napadači otmu neki proizvod prije nego bude pušten u prodaju ili neke alate koji su potrebni tvrtkama za poslovanje i zauzvrat od njih traže novac. Drugi oblik napada predstavlja upadanje u sustav radi krađe bitnih podataka koje oštećeni korisnici otkriju mjesecima kasnije, a u većini slučajeva nikada. Neki od osnovnih tipova digitalnih napada su :

1. *Malware*
2. Pecanje (*phishing*)
3. Socijalni inženjering (*social engineering*)
4. Napadi čovjeka u sredini (*man in the middle*)
5. Napadi nultog dana (*zero day attack*)

4.1.1 *Malware*

Riječ *malware* je kombinacija riječi engl. „*malicious*“ i engl. „*software*“, što bi značilo zlonamjerni softver. Predstavlja svaki oblik softvera čija je svrha poremetiti operacije računala, prikupiti osjetljive informacije o korisniku ili pristupiti privatnim dijelovima sustava. Može biti u obliku koda, skripte, ili zlonamjerne aplikacije. U globalnom smislu zlonamjerni napadi se fokusiraju na sustave različitih vladinih organizacija ili poslovnih korporacija radi krađe osjetljivih podataka ili onemogućavanja obavljanja njihovih zadataka, ali isto tako *malware* se koristi kod krađe osobnih podataka pojedinaca kao što su bankovni računi i brojevi kreditnih kartica. *Malware* se distribuira putem društvenih mreža, piratiziranog softvera, e-mailova i web stranica. Štete uzrokovane *malwareom* mogu biti manje i popravljive ili puno veće i opasnije. Oblici štete su gubitak podataka, zaraženi podaci se mogu automatski obrisati, krađa identiteta preko keylogger programa. Najveća šteta su vjerojatno financijski gubici bilo kod napada na tvrtke ili osobe, ukoliko napadač dođe do financijskih podataka korisnika (bankovni računi, kreditne kartice) može nanijeti veliku financijsku štetu.

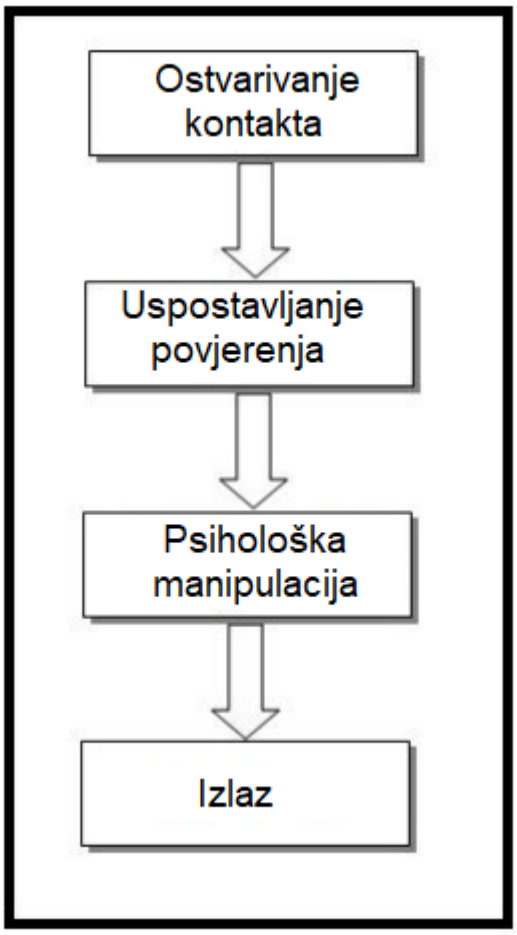
4.2 Pecanje (phishing)

Pecanje (*phishing*) je praksa slanja zlonamjernih e-mailova koji naizgled djeluju da su poslani od strane legitimnih izvora. Koriste iste nazive, logotipove, i sintaksu službenih stranica da navedu žrtve da kliknu na zlonamjerne linkove. Korisnici misle da im primljena poruka treba, pruža neke bitne informacije ili zahtjeva njihovu reakciju. Napad pecanjem postoji od 90-ih godina, najčešće se koristio unutar tvrtki prilikom komunikacije iste sa klijentima. Nakon što korisnik klikne na poveznicu napadač dobije pristup osjetljivim podacima kao što su kreditne kartice, ili podaci za logiranje.

Slika 6 – Napad pecanja

4.3 Social Engineering

Socijalni inženjering (*social engineering*) se može definirati kao najjednostavniji način za krađu osjetljivih podataka od korisnika gdje napadači iskorištavaju znatiželju ljudi i njihovo povjerenje i tako dobiju pristup do osjetljivih informacija i dijelova sustava. U kontekstu zaštite digitalnih sustava primarno se socijalni inženjering koristi za dobivanje povjerljivih informacija od tvrtke ili da ih se prisili na kršenje sigurnosnih protokola te tako zaraze prethodno zaštićene sustave. Napadač koji provede sate pokušavajući probiti lozinke zaposlenika neke tvrtke, može jednostavno nazvati nekog od zaposlenika, predstaviti se kao help desk i pitati korisničke podatke.



Slika 7 – Životni ciklus socijalnog inženjeringa

Socijalni inženjering se dijeli u četiri ciklusa koji su prethodno prikazani na slici 7:

1. Ostvarivanje kontakta–Prva faza se sastoji od prikupljanja informacija o ljudima i načinu poslovanja tvrtke, odnosno mete. Ulaganjem u ovu fazu se mogu otkriti ranjivosti sustava. Ova faza se odvija prije faze napada.
2. Uspostavljanje povjerenja–U ovoj fazi napadač inicira komunikaciju sa metom i preuzima interakciju te se tako povezuje sa metom da zadobije njegovo/njezino povjerenje u svrhu izvlačenja informacija u kasnijim fazama napada.
3. Psihološka manipulacija–Ovdje napadač iskorištava stečeno povjerenje iz prethodne faze da bi dobio povjerljive informacije i na lakši način upao u sustav. Napad se izvršava u ovoj fazi.

4. Izlaz–Nakon što napadač dobije potrebne informacije mora izaći iz sustava bez da ostavi traga ili prebaci sumnju na sebe. Nakon ove faze, ukoliko je uspješno izvedena napadaču se teško ulazi u trag.

5. Strojno učenje u sigurnosnim sustavima

Strojno učenje predstavlja djelotvoran alat koji se može primijeniti u zaštiti sustava. To je vidljivo iz raznih algoritama protiv pecanja i sustava za praćenje upada u mrežu. Unutar industrije digitalne zaštite sustava strojno učenje ima veliki potencijal. Nove metode drastično povećavaju preciznost praćenja napada i ranjivosti i isto tako tehnološkim napretkom računala, povećavanjem njihovih analitičkih sposobnosti mogućnosti analiziranja sve većih količina podataka su postale svakodnevnica. Primjena umjetne inteligencije u sustavima obrane omogućava istima da se samostalno bore protiv napada, bez ljudske instrukcije i interakcije.

5.1 Spam i *phishing* detektiranje

Detektiranje pecanja i spama koristi različite tehnike koji smanjuju utrošeno vrijeme na praćenje takvih napada i potencijalne opasnosti uzrokovane neželjenim e-mailovima. U današnje vrijeme pecanje predstavlja prvi korak u napadu na mrežu, bilo da su mete tvrtke i korporacije ili pojedinci. Detektiranje istih je otežano zbog neprestanih strateških promjena koje koriste napadači, gotovo da niti jedan napad nije identičan i zbog toga napadači zaobilaze tradicionalne filtre za anti-spam. Tu dolazi strojno učenje. Današnji sustavi koji koriste strojno učenje u filtriranju spama i pecanja imaju daleko bolje rezultate od ručnog ili manualnog filtriranja. Prva istraživanja su se fokusirala na problem filtriranja spama gdje je se za rješenje problema koristio *Naive Bayes* algoritam. *Naive Bayes* algoritam se koristio zbog svoje robusnosti, klasificirao je podatke na opasne i bezopasne i osim toga bio je jeftino rješenje.

Cilj pecanja je krađa osobnih podataka, znanstvenici su grupirali anti-phishing metode na tri dijela :
detektivne metode, preventivne metode i ispravljačke metode, prikazane na slici 8.

Detektivna rješenja	Preventivna rješenja	Ispravljačka rješenja
1. Nadzor životnog ciklusa računa 2. Nadzor pečata 3. Onemogućavanje dupliciranja 4. Filtriranje sadržaja 5. Anti *malware* 6. Anti *spam*	1. Autentikacija 2. Izrada zakrpe i promjena vođenja 3. Autentikacija e-maila 4. Sigurnost web aplikacija	1. Rušenje stranice za pecanje 2. Digitalna forenzika i istraživanje

Slika 8 – Anti-phishing metode

5.1.1 E-mail spam filtriranje

Tehnike strojnog učenja i statistički pristup modeliranju za cilj ima izgradnju modela i klasifikatora za filtriranje spam mailova od korisnika. Za izgradnju modela su potrebni prekvalificirani e-mailovi za razlikovanje spam mailova od normalnih. Proces izgradnje modela se naziva trening. Modeli i algoritmi strojnog učenja su pokazali najuspješnije rezultate od svih prethodnih tehnika filtriranja. Primjer korištenja strojnog učenja je i Google, oni su primijenili filtriranje sa strojnim učenjem u Gmail sustavu koji ne samo da blokira nadolazeći spam već i DoS napade, viruse i druge napade.

5.2 *Malware* detektiranje

Detektiranje zlonamjernog softvera predstavlja veliki problem jer noviji i moderniji zlonamjerni softver ima mogućnost automatskog generiranja različitih verzija zlonamjernih programa sa istim efektima. Kod promjena verzija mijenja se i exe datoteka. Zbog ovakvih promjenjivih svojstava zlonamjernog softvera tradicionalni pristup identifikacije istih postaje beskoristan. Zlonamjerni softver se može podijeliti u više kategorija ovisno o svrsi : virusi, crvi, trojanski konji, *adware*, *spyware*, *ransomware* i alati za upravljanje na daljinu. Tehnike strojnog učenja se mogu koristiti za analiziranje zlonamjernog softvera i njihovu podjelu i kategorizaciju.

5.3 Sustavi detektiranja upada (*IDSs*)

Sustav detektiranja upada je obrambena mjera koja prati aktivnosti računalnih mreža i izvještava mrežnog administratora o potencijalnim opasnostima. Napadači na razne načine pokušavaju upasti u mrežu i doći do privatnih podataka, stoga je njihova sigurnost i zaštita glavni dio bilo kakve organizacije. Sigurnost *IDSa* se temelji na kombinaciji autentikacije i autorizacije kontrole pristupa. *IDSovi* zajedno sa antivirusnim sustavima i vatrozidom čine sustav zaštite od napada. *IDS* hvata snimke (*snapshot*) cijelog mrežnog sustava i onda koristi informacije koje dobiva iz uzoraka da odredi kada i kako bi se mogao dogoditi napad. Komponente pripreme za upad u sustav su: poznavanje potencijalnih napada i prevencija potencijalnih napada. Za sprječavanje budućih napada od ključne važnosti može biti poznavanje prošlih napada koji su se dogodili nad sustavom. *IDSovi* su se tradicionalno bazirali na uzorcima poznatih napada, ali moderni pristupi otkrivanja nepravilnosti i opasnosti se baziraju na strojnom učenju. Dva problema *IDS*a u kojima je izražena primjena strojnog učenja su: detektiranje botnetova i algoritmi generiranja domena (*DGAs*). Botnet predstavlja mrežu zaraženih računala preko koje napadači organiziraju nove napade na sustav. Detektiranje botnetova identificira komunikaciju između zaraženih računala unutar mreže sa vanjskim serverima. DGA su velika prijetnja korporacijama, oni generiraju nova imena domena i uz pomoć njih zaražena računala komuniciraju sa vanjskim serverima tako što periodički generiraju nove nazive domaćina (*hostnames*).

5.3.1 Vrste IDS sustava

Sustavi detektiranja upada (*IDSs*) se dijele na tri vrste :

1. *Network Intrusion Detection System* (sustavi detektiranja upada u mrežu)
2. *Network Node Intrusion Detection System* (sustavi detektiranja upada u čvorove mreže)
3. *Host Intrusion Detection System* (sustavi detektiranja upada domaćina)

Najjednostavnije rečeno prva dva tipa sustava prate mrežni promet, dok treći tip sustava prati aktivnosti i datoteke na uređajima domaćina.

NIDS–se postavlja na strateške točke mreže sa namjerom da pokriva mjesta na kojima je mrežni promet najskloniji napadima. U pravilu se postavlja na sve razine

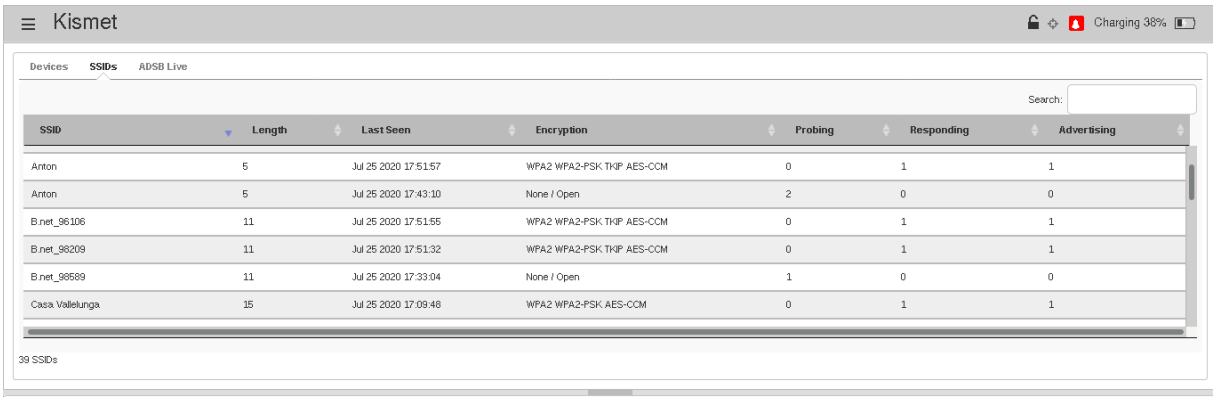
mreže i pasivno uspoređuje promet sa poznatim napadima. *NIDS* nije nepogrješiv, budući da analiziraju jako veliku količinu podataka zna se dogoditi da mu neki napadi „promaknu“ ili da ih ne uspije otkriti, u tim slučajevima je potrebna uloga mrežnog administratora.

NNIDS–je identičan *NIDSu* samo se on primjenjuje na jednom domaćinu, a ne na cijeloj mreži.

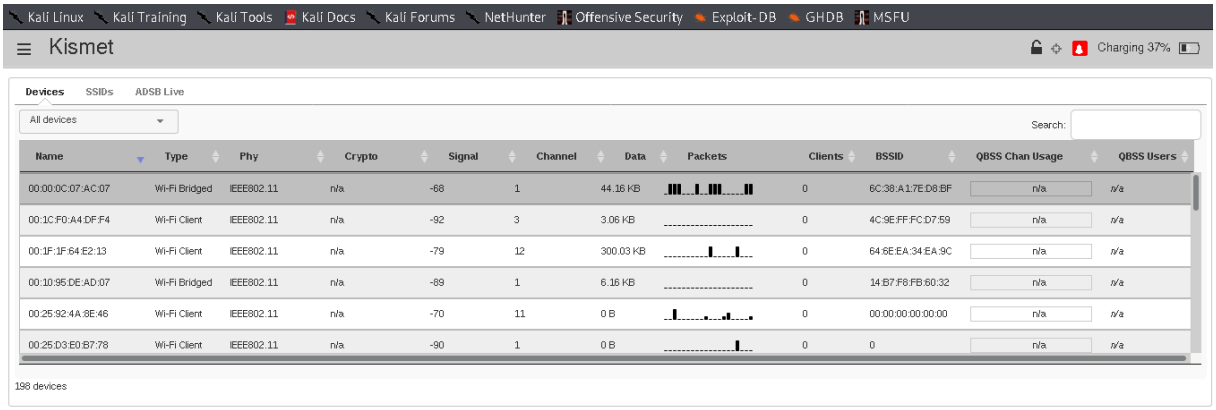
HIDS–se pokreće na svim uređajima koji imaju pristup mreži i internetu unutar organizacije. Ovaj sustav ima prednost nad *NIDSima* jer je više fokusiran na unutarnji mrežni promet. Uspoređuje snimke datoteka i ako primijeti nepravilnosti odmah ih javlja administratoru.

5.3.2 Kismet

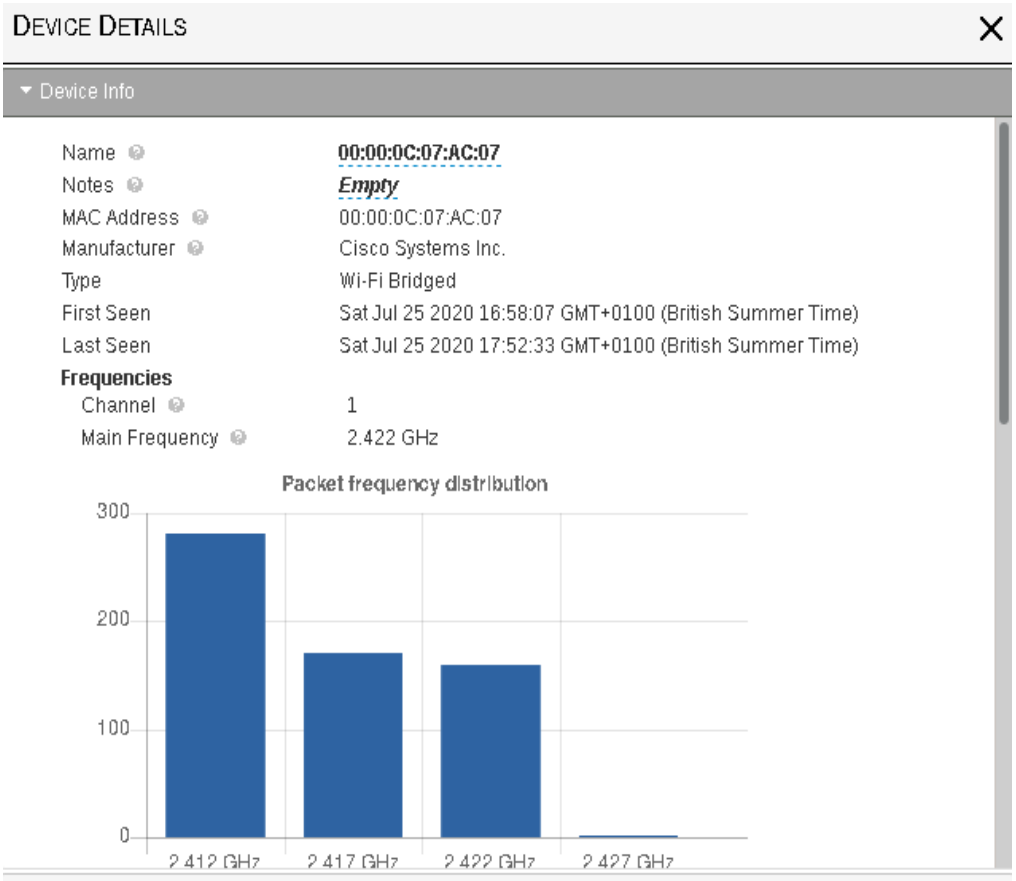
Kismet je *IDS* koji se fokusira na bežične protokole i bluetooth. Za omogućavanje rada Kismet programa potrebno je imati mrežnu karticu koja podržava opciju nadziranja (kartica izgubi pristup internetu, ali može pratiti pristupne točke i promet). Kismet osim što otkriva bežične pristupne točke isto tako otkriva neautorizirane pristupne točke čiji je broj veći nego što se čini. Kismet prvo skenira okolne mreže i zatim omogućava korisniku da prati prijenos paketa na zasebnom uređaju. Nedostatak Kismeta je taj što je ograničen na wi-fi mreže i što dugo vremena pretražuje iste. Prednosti Kismeta su mogućnost proširivanja sučelja sa dodacima i aktivna zajednica sa kojom možete stupiti u direktni kontakt za bilo kakvu pomoć. Kroz nekoliko idućih slika ću prikazati dio sučelja Kismet-a.



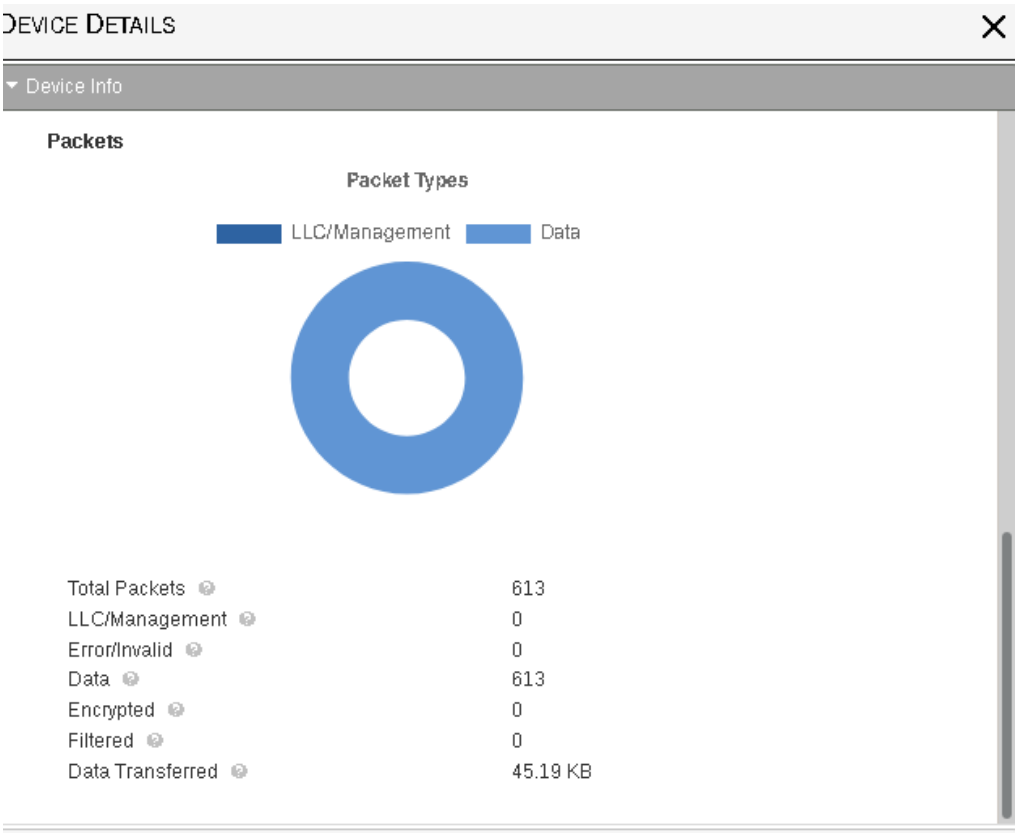
Slika 9 – Popis otkrivenih SSID-eva (naziva wi-fi mreža)



Slika 10 – Popis svih uređaja koje je otkrio Kismet



Slika 11 – Detalji odabranog uređaja I. dio



Slika 12 – Detalji odabranog uređaja II. dio

Iz slika 11 i 12 se vide informacije o uređaju kao što su naziv uređaja, proizvođač, način spajanja i distribucija paketa, osim toga i podjela podataka na upravljačke i podatkovne. Kismet prati ukupni broj paketa, pakete sa greškom, kriptirane pakete i na kraju računa veličinu prenošenih podataka.

Slika 13 – Wi–fi detalji uređaja

Na slici 13 su prikazani ukupni paketi koji dolaze preko wi–fi mreže i Kismet nudi grafički prikaz sveukupnih paketa a isto tako i podatkovnih. (Broj paketa je veći nego na prethodnoj slici jer Kismet prati prijenos u realnom vremenu 613; 623).

Slika 14 – Grafički prikazi paketa

Posljednja opcija je grafički prikaz paketa kroz vrijeme rada (minute, sati, dani) i posebno grafovi za podatkovne pakete.

6. Ostale primjene strojnog učenja u zaštiti sustava

Kako prijetnje za sigurnost sustava postaju sve opasnije istraživanja se fokusiraju na korištenje alata strojnog učenja za identifikaciju, reagiranje i zaustavljanje digitalnih napada. Strojno učenje se može primijeniti na analiziranje napada, isto tako može automatizirati rutinske zadatke.

Najznačajniji slučajevi uporabe strojnog učenja u zaštiti sustava su :

1. Detektiranje spama
2. Analiza *malwarea*
3. Detekcija upada u mrežu
4. Analiziranje mobilnih krajnjih točaka
5. Zatvaranje *zero-day* ranjivosti uz pomoć strojnog učenja

Neke upotrebe su pojašnjene u prethodnom dijelu, osim preostalih bitno je navesti i detektiranje i klasifikaciju prijetnji, ocjenu rizika mreže i automatizaciju i optimizaciju rutinskih zadataka.

6.1 Detektiranje i klasifikacija prijetnji

Algoritmi strojnog učenja se mogu implementirati u sustave na takav način da se na temelju modela koji uči kroz analiziranje velikih količina podataka identificiraju uzorci zlonamjernih aktivnosti, kao rezultat toga novootkrivene slične aktivnosti se automatski rješavaju. Modeli treniraju na skupu podataka od prije definiranih napada i kasnije se na temelju njih grade modeli koji prate nove napade. Strojno učenje se može koristiti i pri klasifikaciji zlonamjernog softvera, sustav ih automatski dijeli i grupira na temelju njihovog ponašanja. Uz takav način rada analitičarima se skraćuje i olakšava posao i oni onda puno brže mogu identificirati nove prijetnje i na njih pravilno reagirati. Na primjer koristeći informacije iz *WannaCry ransomware* napada, model koji je napravljen na temelju tih podataka može naučiti kako identificirati slične napade i tako olakšati njihovo prijevremeno otkrivanje. Isto tako tehnike strojnog učenja se koriste u klasifikaciji IP prometa koji može pomoći u automatiziranju procesa kod *IDSa* (sustava za detektiranje upada u mrežu) za identificiranje uzoraka ponašanja kod DDOS (*distributed denial of service*) napada.

6.3 Ocjena rizika mreže

Ocjena rizika mreže je korištenje kvantitativnih mjera za dodjeljivanje ocjena rizika različitim dijelovima mreže i na temelju toga se raspoređuju prioriteti za podjelu resursa zaštite sustava unutar pojedinih organizacija. Ovaj proces se može automatizirati preko strojnog učenja. Model može na temelju analiziranja prijašnjih digitalnih napada zaključiti koji su najranjiviji dijelovi mreže i kojim vrstama napada su mreže najizloženije i na temelju njih dati ocjenu rizika. Prednost kod korištenja strojnog učenja je ta što se rezultati temelje isključivo na podacima, dok su se prijašnji temeljili na znanju o mrežama i iskustvu analitičara. Na temelju ocjena organizacije se mogu pripremiti na određene vrste napada prije nego se oni dogode. Provedena su istraživanja u kojima su se koristili algoritmi kao što su *K–Nearest*

Neighbour, Support Vector Machines i *Random Forest* za analizu i grupiranje mrežnih podataka.

6.4 Automatiziranje rutinskih zadataka i optimiziranje izvještaja

Strojno učenje se može iskoristiti za automatiziranje zadataka koje sigurnosni analitičari svakodnevno moraju raditi. Analiziranjem starih izvještaja analitičara o identificiranju i odgovaranju na određene napade, može se konstruirati model koji bi na temelju tih podataka samostalno naučio odgovarati na te iste napade bez sudjelovanja čovjeka. Još uvijek je nemoguće u potpunosti zamijeniti čovjeka, tj. analitičara, ali sa uključivanjem strojnog učenja u procese digitalne zaštite sustava čovjek i računalo postižu rezultate na brži, bolji i kvalitetniji način.

Zbog sve većeg razvoja umjetne inteligencije eksponencijalno raste i proces automatizacije zadataka. Sve više zadataka koje je do sada obavljao čovjek, sada se prepuštaju računalima, i svakim danom ih se automatizira sve više. U nekim slučajevima potpuno automatiziranje zadataka ne daje najbolje rješenje, već kombiniranjem umjetne i ljudske inteligencije se postižu bolji rezultati. Danas se može primijetiti rast tvrtki koji se ne fokusiraju samo na automatiziranje procesa koristeći umjetnu inteligenciju nego rade proizvode koji poboljšavaju i nadopunjuju rad analitičara. Primjer takve tvrtke je Palantir, njihovi proizvodi olakšavaju analiziranje ogromnih količina podataka za analitičare.

7. Automatizirana *malware* obrana

Klasični sustavi ne uspijevaju uvijek ispravno obraditi ogroman broj *malware* napada, međutim, sustavi uz pomoć umjetne inteligencije se mogu istrenirati da identificiraju i najmanje *ransomware* i *malware* napade prije nego oni dospiju do sustava. Da identificiraju prijetnje, analitičari su morali provjeravati nizove slova u binarnom kodu u potrazi za potpisom napadača. Postoje algoritmi ponašanja koji ne analiziraju kod direktno nego koriste modele vjerojatnosti za pronalazak zlonamjernog koda. Algoritmi ponašanja su jako skupi i nisu uvijek djelotvorni. Za razliku od njih puno bolji je heuristički algoritam. On uz pomoć baze podataka o karakteristikama zlonamjernog i „dobrog“ koda zaključuje da li je neki kod opasan ili ne. Glavna karakteristika ovog algoritma, kao i svakog drugog algoritma strojnog učenja je ta da se može prilagoditi i evoluirati.

7.1 *AI^2 : Big data* obrana

U svom znanstvenom radu istraživači sa MIT-evog instituta i laboratorija za informatiku i umjetnu inteligenciju u suradnji sa startup tvrtkom koja se bavi strojnim učenjem *PatternEx* su predstavili platformu *AI^2* koja predviđa digitalne napade puno bolje od postojećih sustava tako što konstantno dobiva nove upute od ljudskih stručnjaka. Njihov sustav spaja ljudsku intuiciju sa modernim strojnim učenjem i rezultat toga je aktivni sustav koji može učiti. Sustav ima četiri ključna svojstva : bihevioralnu analitičku platformu za obradu podataka, metode za detektiranje, mehanizam za prikupljanje povratnih informacija od ljudi i modul učenja uz nadzor. U svome radu su sustav testirali na 3.6 milijarde linija log datoteka, što dokazuje da sustav ima sposobnost učenja. Veeramachaneni i Arnaldo su identificirali tri najveća izazova za industriju IT sigurnosti:

1. Nedostatak označenih podataka–mnogim tvrtkama nedostaju označeni primjeri prošlih napada što im onemogućava korištenje modela za učenje uz nadzor.

2. Neprestani razvoj napada–čak i kada je učenje omogućeno neće uvijek biti uspješno jer napadači neprestano mijenjaju načine napada.

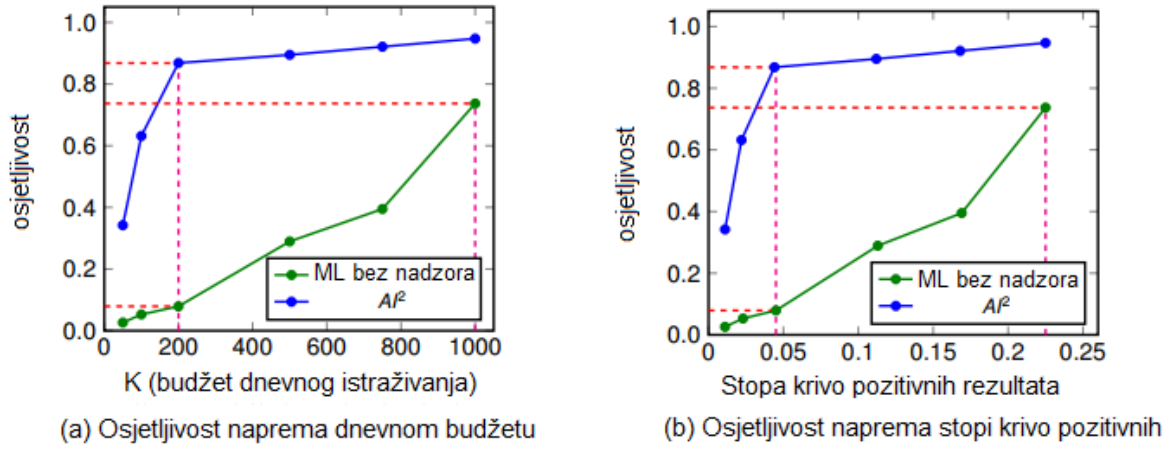
3. Ograničeno vrijeme za istraživanje i budžet–oslanjanje samo na analitičare za istraživanje napada je veliki financijski i vremenski trošak.

Oni su predstavili rješenje koje kombinira iskustvo analitičara sa najnovijim tehnikama strojnog učenja, sustav su nazvali *AI^2*. *AI^2* uči i automatski kreira modele koji kada obrađuju nove podatke daju pretpostavke koje su jednako kvalitetne kao i ljudske.
Sustav je razvijen sa AMS pristupom (*Active Model Synthesis*) koji se sastoji od šest koraka:

1.Model računa ponašanje podataka u velikom skupu podataka (*data set*)
2.Analitičaru predstavi jako mali set podataka koji se generira preko sustava strojnog učenja za detektiranje napada bez nadzora čovjeka
3.Prikuplja povratne informacije o podacima

4.Samostalno uči nove modele uz nadzor koristeći te povratne informacije
5.Kombiniranjem modela uz nadzor i modela bez nadzora predviđa napade
6.Ponavljanje koraka 1. – 5.

Na slici 15 su predstavljeni rezultati korištenja sustava u razdoblju od tri mjeseca.



Slika 15 – Rezultati korištenja *AI²* sustava

AI² sustav efektivno iskorištava vrijeme-sustav detektira 86.8% napada na samo 200 događaja, u usporedbi sa do sada korištenim sustavima to je deset puta više. Ako se broj događaja poveća na npr. 1000, sustav bez nadzora postiže 73.7% točnosti detektiranja napada uz postotak „krivo–pozitivnih“ od otprilike 22%, dok *AI²* sustav ima 86% točnosti i postotak „krivo–pozitivnih“ od 4.4%, što je pet puta manje od dosadašnjih sustava. U svom radu su predstavili sustav koji prikuplja i analizira podatke samostalno na temelju strojnog učenja, sustav dobiva povratne informacije na temelju kojih gradi nove modele koji detektiraju napade, bez sudjelovanja čovjeka.

AI² se sastoji od četiri komponente:

1. Sustav za obradu podataka–platforma koja obrađuje podatke i kvantificira ih
2. Sustav za vanjsko otkrivanje–sustav koji iz podataka gradi model preko učenja bez nadzora koristeći neku od tri metode: gustoća, dekompozicija matrica ili repliciranje neuronskih mreža
3. Mehanizam povratnih informacija–ova komponenta daje analitičaru određeni broj događaja, nakon toga ih analitičar označava da li su opasni ili ne, povratne informacije se šalju u modul

4. Modul učenja uz nadzor–od povratnih informacija sustav pravi novi model koji predviđa da li su novi događaji opasni ili ne, kako dobiva više povratnih informacija tako model postaje sve bolji i precizniji

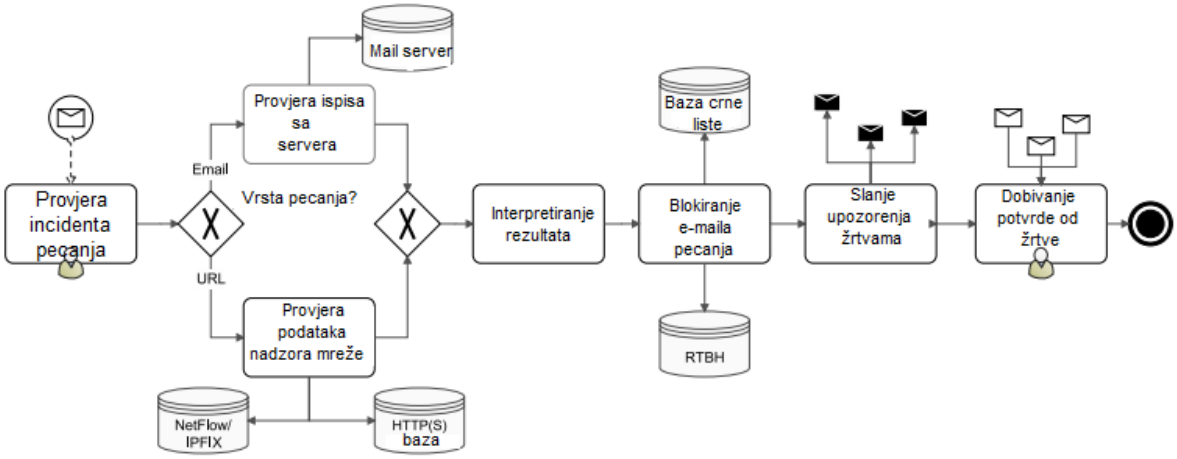
Prema riječima Nitesha Chawle profesora informatike sa sveučilišta Notre Dame ovaj znanstveni rad kombinira snage analitičara i njihove intuicije sa strojnim učenjem i rezultat toga je povećavanje broja točno otkrivenih napada i smanjivanje broja „krivo – pozitivnih“ napada.

8. PhiGARo : *Framework* za automatsko detektiranje pecanja

Bez obzira na to na koji način se štiti infrastruktura IT-a ljudski faktor će uvijek biti slaba točka. Pecanje kao takav je najpoznatiji oblik socijalnog inženjering napada. Mete napada pecanja nisu samo financijske institucije, nego i zaposlenici tvrtki i korisnici društvenih mreža. Cilj napadača ne mora uvijek biti financijska dobit, može isto tako biti i prikupljanje osobnih podataka ili slanje spam poruka. Obrađivanje napada pecanja iziskuje ponekad više vremena nego što to mrežni administratori imaju, a isto tako ovise o izvještajima korisnika koji su žrtve takvih napada. Članovi instituta za informatiku iz Brna, Husak i Čegan su došli na ideju za *framework* koji bi automatizirao detektiranje i izvještavanje o *phishing* napadima, te tako skratio i olakšao vrijeme potrebno za rješavanje istih.
Glavni dio njihove ideje je PhiGARo, alat za automatsku obradu napada pecanja. Alat otkriva žrtve pecanja i sprječava nastanak daljnje štete. Napad pecanja se obrađuje automatski ali svejedno ovisi o podnošenju izvještaja o napadu, Husak i Čegan su došli na ideju da se dio podnošenja izvještaja automatizira što bi omogućilo alatu da samostalno otkriva i obrađuje napade, te pri tome sprječava daljnju opasnost samog napada. Oni su za otkrivanje napada pecanja koristili *honeypot*, kao nezavisni dio mreže sustava. Mailovi pecanja koji dospiju na *honeypot* server bi se automatski slali PhiGARo alatu koji bi bio neovisan o izvještajima administratora.

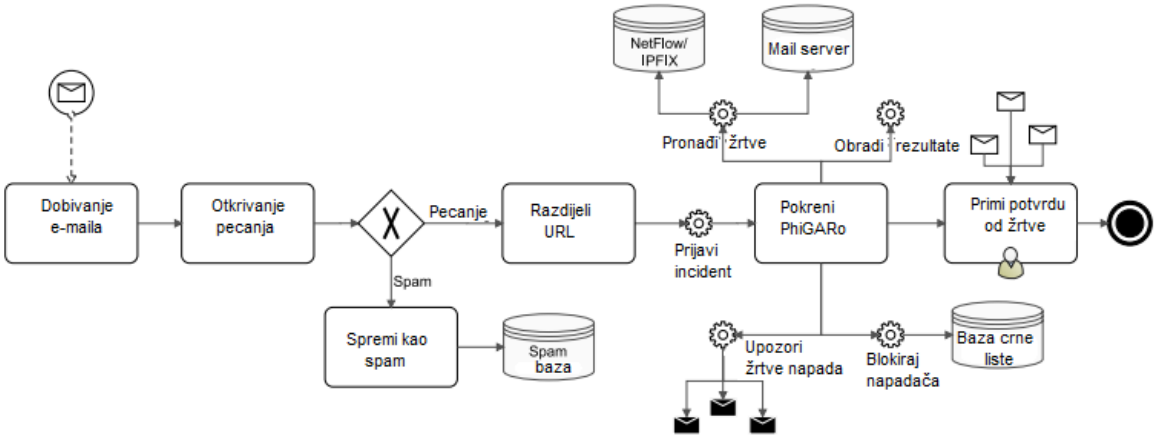
Husak i Čegan predlažu *framework* čije su rješenje podijelili u dva dijela : prvi dio bi bio otkrivanje napada pecanja, a drugi bi bio njegova obrada. PhiGARo je postao

ključni dio njihovog *frameworka* jer je automatizirao proces odgovaranja na napad pecanja.



Slika 16 – Dijagram PhiGARo sustava

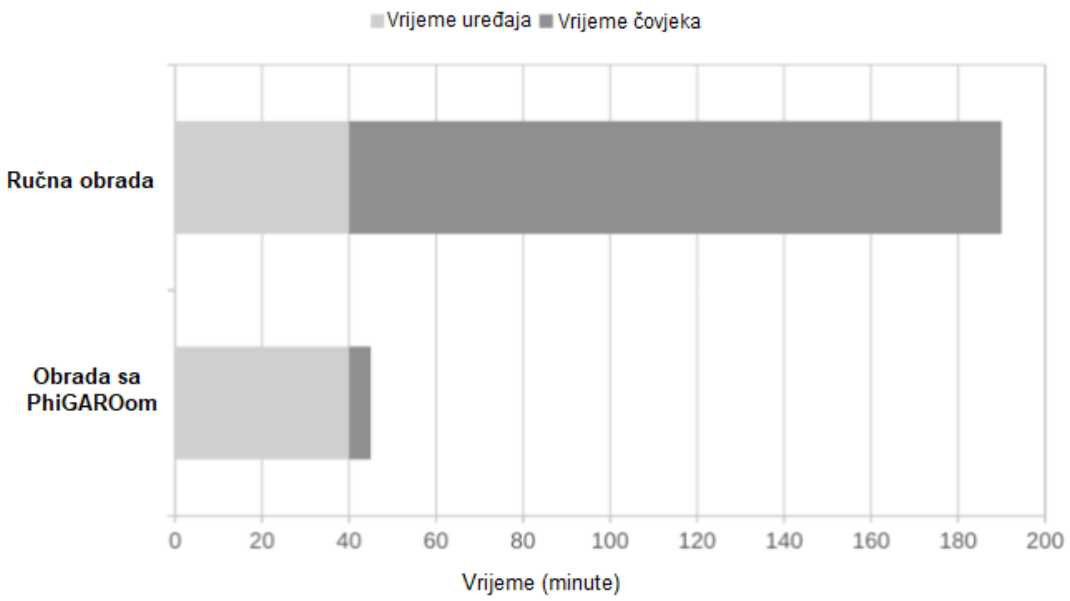
Prvo što PhiGARo napravi je to da procjeni na koji način napadač pronalazi žrtvu. Ako napad sadrži neki *URL* (*Uniform Resource Locator*) pronalazi ga preko nadziranja mreže, a ako je putem e-maila onda koristi sistemske zapise e-mail servera. Napadačev *URL* se prvo raščlanjuje na skraćene *URLove* i na preusmjerene, te se tako dolazi do originalnog *URLa*. *NetFlow* se koristi za pronalazak žrtvi napadača koje su pristupile zlonamjernom *URLu*. PhiGARo pronalazi IP adresu *phishing* stranice i pronalazi mrežni promet između te adrese i mreže, i zatim označi te adrese kao potencijalne žrtve napada pecanja. Ako napad nije preko *URLa*, već preko e-maila, PhiGARo traga za odgovorenim mailovima tako što traži adrese pošiljatelja i primatelja, nakon toga se žrtva identificira preko e-mail adrese. PhiGARo ima mehanizam koji blokira promet između zaštićenog sustava i IPa *phishing* stranice, dok moduli za izvještavanje prosljeđuju izvješća dalje. Zadnji korak obrade je obavještavanje žrtava, PhiGARo šalje e-mail upozorenje svim identificiranim žrtvama. Izvještaj se sastoji od objašnjavanja napada, primjera poruke pecanja, dokaza i *screenshota phishing* stranice. Cijeli ovaj proces je prikazan preko dijagrama na slici 16.



Slika 17 – Dijagram automatske detekcije i izvještavanja

Na slici 17 je prikazan dijagram automatiziranog otkrivanja pokušaja pecanja i obrade napada uz pomoć PhiGARo-a. Nakon što primi poruku, ukoliko je identificirana kao pokušaj pecanja, prvo se pretražuje *URL* i on se razdjeljuje. Standardizirani izvještaj pecanja se generira i šalje PhiGARo-u koji se automatski pokreće, bez naredbi korisnika.

Husak i Čegan su testirali sustav na 133 napada pecanja i njihovi rezultati su prikazani na slici 18.
Iz slike se može zaključiti da je utrošeno vrijeme za obradu napada automatizacijom drastično manje od ručne obrade.



Slika 18 – Prosječno vrijeme ručne obrade napada i obrade sa PhiGARo-om

9. Razvoj *honeypotova* sa umjetnom inteligencijom

Honeypotovi u IT-u predstavljaju resurse koji se namjerno ubacuju u mrežu da bi namamili napadače i tokom njihovih napada prikupljali informacije o načinu napada, njihovim alatima i tehnikama. Obično kod *honeypotova* ne postoji nikakva interakcija, jer kod korištenja *honeypota* svaka interakcija se smatra zlonamjernom. Vrijednost *honeypotova* je u tome što one prikupljaju informacije. Oni unutar mreže predstavljaju domaćine kojima je cilj namamiti napadače. Postoji više načina za postavljanje *honeypota*: kao fizički uređaj, virtualni uređaj ili kao virtualni domaćin. Fizički *honeypot* je stvarna platforma sa vlastitom IP adresom. Virtualni *honeypot* uređaj se može napraviti sa nekim softverom za virtualizaciju kao što su VMW ili Virtualbox, prednost virtualnih *honeypotova* je ta što se na jednoj platformi može pokrenuti više *honeypotova*. Ako se *honeypot* radi kao virtualni domaćin, to se ostvaruje preko alata *Honeyd*, koji omogućava pokretanje više virtualnih domaćina na jednom uređaju.

Honeypotovi se dijele prema razini interakcije–interakcija predstavlja razinu komunikacije koja je dozvoljena između napadača i samog *honeypota*. Interakcija isto tako označava koliko se *honeypot* može izložiti napadaču, odnosno koliko ga sam napadač može iskoristiti. Što je *honeypot* izloženiji napadima to može prikupiti više informacija. Podjela se vrši na: *honeypotove* niske interakcije i na *honeypotove* visoke interakcije.

9.1 *Honeypot* niske interakcije

*Honeypo*t niske interakcije *(Low–interaction honeypot)* kako mu i ime govori ima najmanju sposobnost interakcije između napadača i *honeypota*, isto tako predstavlja najjednostavniju vrstu *honeypota*. Ovaj tip *honeypota* se postavlja kao virtualni domaćin sa nekim dodatnim servisima, kao što su FTP (File Transfer Protocol) servisi, koji se postavljaju preko *honeypot frameworka Honeyd*. *Honeyd* je open source *framework* koji jednostavno postavlja *honeypotove* niske interakcije na virtualne mašine. Najveća prednost ovakvih *honeypotova* je rizik, razina rizika je najmanja jer nude najmanje usluge. Napadači se mogu samo spojiti na *honeypot* i skenirati virtualne ulaze *(portove),* još jedna prednost ove vrste *honeypotova* jeste lakoća održavanja i razvijanja. Nedostatak ovih *honeypotova* jest mala količina informacija koje mogu prikupiti od napadača.

9.2 *Honeypot* visoke interakcije

Honeypot visoke interakcije (*High–interaction honeypot*) predstavlja složeno rješenje *honeypot* sustava jer nudi paket pravih servisa napadačima koji napadače ne ograničavaju. Ovaj tip *honeypota* omogućava analitičarima prikupljanje velikih količina podataka o napadačevim aktivnostima. Prikupljene informacije se mogu dalje iskoristiti za proučavanje napadačevog ponašanja, alata, motiva i ponekad i identiteta. Njihova najveća prednost nad *honeypotovima* niske interakcije jeste količina prikupljenih informacija. Nedostatci *honeypotova* visoke interakcije su troškovi održavanja, vrijeme potrebno za postavljanje ovakvih sistema i za njihovo korištenje i održavanje potrebni su stručnjaci.

9.3 Ekspertni sustavi i rasuđivanje na temelju slučaja

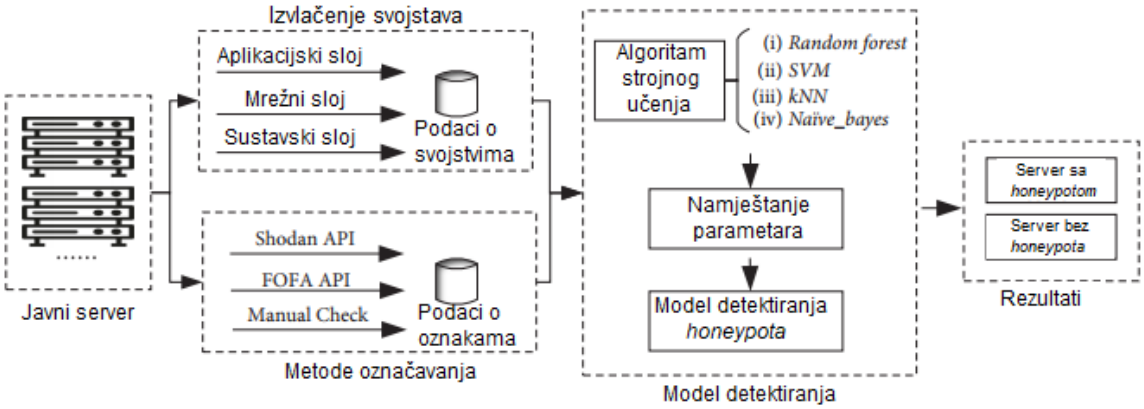
U svome znanstvenom istraživanju Zakaria i Kiah predstavljaju dva pristupa korištenja umjetne inteligencije u kombinaciji sa *honepotovima*. Ekspertni sustavi (*Expert system*) i rasuđivanje na temelju slučaja (*Case–based reasoning*). Ekspertni sustav je program koji simulira rješavanje problema poput ljudskog stručnjaka u nekoj određenoj domeni. ESi su jedna od najupotrebljivijih tehnika umjetne inteligencije na svijetu. ES se sastoji od radne memorije (*Working Memory*), baze znanja (*Knowlage Base*) i alata za zaključivanje (*Inference Engine*). Radna memorija predstavlja prostor u koji se spremaju činjenice koje unosi korisnik, njih obrađuje alat za zaključivanje koji će pretražiti bazu znanja da donese rješenje. Danas se ekspertni sustavi mogu lako napraviti, postoje alati u C-u i Javi koji omogućavaju jednostavno izrađivanje ekspertnih sustava. Najveći izazov kod izgradnje ovih sustava je izrada baze znanja, jer kvaliteta baze ovisi i znanju i sposobnostima *developera* koji ju radi.

Rasuđivanje na temelju slučaja (*Case–based reasoning*) predstavlja način rješavanja novih problema iskorištavanjem starih rješenja. U ovakvim sustavima znanje se temelji na prošlim iskustvima i implementira se kroz primjere slučajeva. Slučaj (*case*) je model znanja za određeno iskustvo povezano sa nekom domenom. Slučaj ima dva dijela: opis problema i rješenje problema. Slučajevi se spremaju u bazu za lakše dohvaćanje. *CBR* sustavi rješavaju nove probleme tako što prilagođavaju stara rješenja dovoljno dobro da se mogu primijeniti i na rješavanje novih problema. *CBR* ciklus se sastoji od četiri koraka:

Retrieve (dohvaćanje)–trenutni problem se spaja sa najsličnijim rješenjem iz prošlosti i dohvaća se iz baze, najčešće preko K-*nearest neighbour* algoritma

Reuse ili *Revise* (iskorištavanje ili izmjena)–mijenjaju se značajke starih problema da se prilagode za rješavanje novog, ovaj korak se radi pomoću zamjene ili prilagodbe parametara

Retain (spremanje)–u ovom koraku sustav uči i sprema iskustva, novo rješenje se sprema u bazu za daljnju upotrebu, što više slučajeva obradi, sustav postaje sve učinkovitiji



Slika 19 – *Framework* za automatsko detektiranje

Postoje razni alati za izgradnju *CBR* sustava kao što su myCBR i Jcolibri. Jedna varijanta Jcolibria je Jcolibri Studio koji je integriran unutar *Eclipsea* i omogućava brzu izgradnju *CBR* sustava.

10. Strojno učenje u digitalnom kriminalu

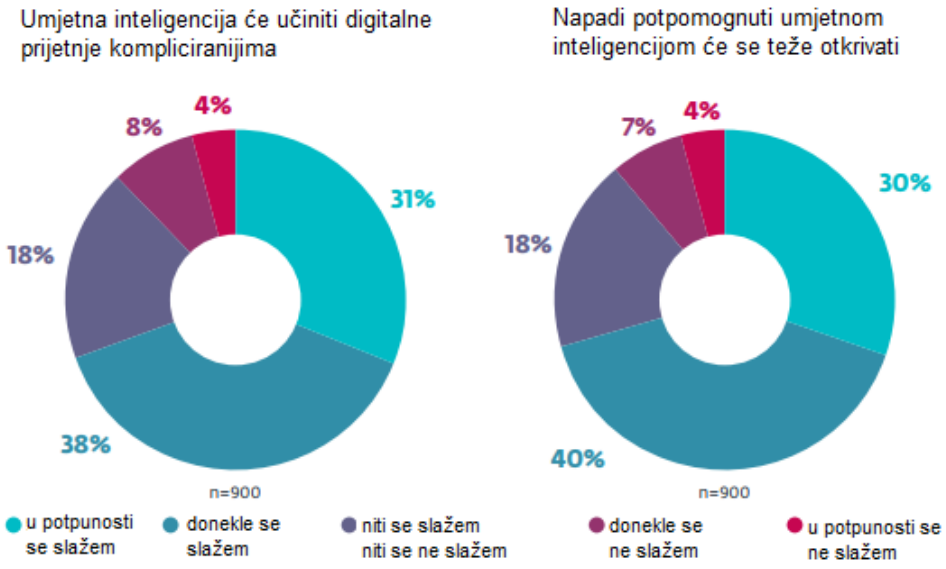
Strojno učenje može biti saveznik u obrani i zaštiti sustava, ali isto tako nove tehnike strojnog učenja koriste i napadači. Kako se strojno učenje razvija za digitalnu obranu, isto tako se može razvijati i za digitalne napade, koji mogu ciljano napadati modele strojnog učenja. Obje strane, i sigurnosni analitičari kao i digitalni napadači pokušavaju primijeniti nove tehnike umjetne inteligencije u svome radu. Neke od poznatijih tehnika strojnog učenja koje se koriste za napade su: neautorizirani pristup i izbjegavajući *malware*.

Područje koje je bilo jako pogođeno strojnim učenjem je strojni vid (*machine vision*), gdje se računalo trenira da prepoznaje objekte na slikama, ista tehnologija se koristi kod vozila koja upravljaju sama sobom. Računalo se može istrenirati da zaobiđe i captcha autorizaciju, jednu od najčešćih na internetu.

Captcha autorizacija traži od korisnika da označi određenu vrstu objekta na slikama prije autorizacije, taj proces se može ubrzati putem neuronskih mreža. Brojna istraživanja su dokazala da strojno učenje može olakšati upad u sustave koji koriste ovakvu vrstu zaštite. Neki od primjera su PassGAN, sustav koji generira lozinke koristeći GAN (*Generative Adversarial Networks*). Sustavi strojnog učenja koji proučavaju e-mail i komunikaciju korisnika uče kako kopirati stil pisanja korisnika i prilikom izvršavanja napada, otežavaju razlikovanje e-mailova napadača od pravih e-mailova.

Postoje slučajevi gdje se metode strojnog učenja koriste za generiranje zlonamjernog koda, takav kod sustavi za obranu nisu u mogućnosti otkriti, čak i oni koji isto rade pomoću strojnog učenja. Primjer je DeepLocker, zlonamjerni softver baziran na umjetnoj inteligenciji koji je razvio tim IBM-ovih istraživača koji ima sposobnost pronalaska svoje mete preko prepoznavanja lica, glasa i pronalaska geolokacije.

Na slici 14 su predstavljeni rezultati ankete provedene na oko 1000 ispitanika iz IT svijeta. Rezultati pokazuju da velika većina ispitanika misli da će umjetna inteligencija jako unaprijediti zlonamjerne prijetnje i napade.



Slika 20 – Rezultati ankete na 900 ispitanika iz IT područja

11. ZAKLJUČAK

Činjenica je da se tehnologija danas razvija nevjerojatnom brzinom. Pretpostavke najvećih stručnjaka u prošlosti nisu mogle predvidjeti ovako brz tehnološki razvoj u posljednjih 40 godina. Tehnologija koja je doživjela najveći razvoj jeste informacijska tehnologija. Posebna grana koja se razvila unutar informacijske tehnologije je grana umjetne inteligencije, koja trenutno dobiva posebnu pažnju bilo sa istraživačke strane ili sa praktične. Umjetna inteligencija, njezinom primjenom u bilo kojoj znanosti (medicina, inženjerstvo, zaštita sustava) došlo je do velikih promjena, neki dijelovi su se automatizirali, unaprijedili i poboljšali, a u nekim dijelovima računala su zamijenila ljude. Umjetna inteligencija je ljudima olakšala odrađivanje repetitivnih zadataka, koji su iziskivali veliki utrošak vremena, automatiziranjem takvih zadataka ljudima se oslobodilo vrijeme da se bave drugim i važnijim stvarima.

U današnjem vremenu informacija je postala najvrjednija valuta, vrijednost informacije nema cijenu. Informacija kao takva je postala meta napadača koji ju žele otuđiti, pokvariti, iskoristiti za svoje osobne potrebe i najčešće nanijeti štetu bilo tvrtkama ili individualnim osobama. Tu na snagu stupa zaštita sustava, primjenom umjetne inteligencije u zaštiti sustava veliki broj zadataka se uspio poboljšati i olakšati. Korištenjem modela strojnog učenja za otkrivanje napada dostignuta je razina otkrivanja napada prije nego se oni dogode, jer računala uz pomoć umjetne inteligencije imaju veliku sposobnost učenja koja čini proces otkrivanja *malwarea* i napada bržim nego što je bio ikad prije. Naravno da sposobnosti umjetne inteligencije nisu neograničene, one i dalje ovise o sposobnostima ljudi koji ju razvijaju, kako, koliko brzo i koliko efikasno.

Velika većina ljudi pri spomenu riječi umjetna inteligencija ima pomisao na preuzimanje ljudskog načina života od strane računala, mašina, robota i drugih opasnosti tehnologije, koje samo čekaju pravi trenutak za to. Tokom istraživanja gradiva, članaka, znanstvenih radova i drugih materijala za pisanje ovog rada sam stekao dojam da se čovjek ne treba bojati umjetne inteligencije, već da je ona njegov saveznik bilo da čovjek ima dobre ili neke druge namjere. Znači da moć umjetne inteligencije i dalje ovisi o čovjeku i o njegovom načinu iskorištavanja iste.

Mogućnosti čovjeka su ograničene, kao i mogućnosti njegovih izuma, jedan od njih je i umjetna inteligencija, čovjek sam nije sposoban probiti neke granice, ali u kombinaciji i kroz suradnju sa umjetnom inteligencijom, zajedničkim radom se mogu dostići mjesta na koja niti čovjek niti računalo samostalno ne bi nikada dosegli.

LITERATURA

1. Russel S., Norvig P.; Artificial intelligence : a modern approach, Prentice-Hall Inc., 1995.

2. Burkov A., The Hundred-Page Machine Learning Book, Andriy Burkov, 2019.

3. https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/ (pristupljeno : 15. 07. 2020.)

4. https://towardsdatascience.com/machine-learning-for-cybersecurity-101-7822b802790b (pristupljeno : 15. 07. 2020.)

5. https://duncanwinfrey.com/cia-security-triangle (pristupljeno : 15. 07. 2020.)

6. https://builtin.com/cybersecurity (pristupljeno : 16. 07. 2020.)

7. D. Sumeet, D. Xian; Data Mining and Machine Learning inc Cybersecurity, Taylor and Francis Group, LLC, 2011.

8. J. Steinberg; Cybersecurity for Dummies, John Wiley & Sons, 2020.

9. https://www.csoonline.com/article/2117843/what-is-phishing-how-this-cyber-attack-works-and-how-to-prevent-it.html (pristupljeno : 16. 07. 2020.)

10. Breda F., Barbosa H., Morais T.; Social Engineering and Cyber Security, Conference paper, 2017.

11. Salahdine F., Kaabouch N.; School of Electrical Engineering and Computer Science,

University of North Dakota,Grand Forks; Social Engineering Attacks: A survey, 2019.

12. https://www.exploit-db.com/docs/english/18135-social-engineering---the-human-factor.pdf (pristupljeno : 16. 07. 2020.)

13. Proko E., Hyso A., Gjylapi D.; Machine Learning Algorithms in Cyber Security, RTA-CSIT, 2018.

14. Ford V., Siraj A.; Applications of machine learning in cyber security,

In Proceedings of the 27th International Conference on Computer Applications in Industry and Engineering, 2014.

15. https://www.dnsstuff.com/intrusion-detection-system#what-is-an-intrusion-detection-system (pristupljeno : 16. 07. 2020.)

16. https://www.dnsstuff.com/network-intrusion-detection-software (pristupljeno 26. 07. 2020.)

17. Rege M.; Mbah R.B.K.; Machine Learning for Cyber Defense and Attack, The Seventh International Conference on Data Analytics, DATA ANALYTICS 2018, 2018.

18. https://www.csoonline.com/article/3227906/what-is-wannacry-ransomware-how-does-it-infect-and-who-was-responsible.html (pristupljeno : 16. 07. 2020.)

19. https://www.mdsny.com/5-top-machine-learning-use-cases-for-security/ (pristupljeno : 17. 07. 2020.)

20. Veeramachaneni K., Arnaldo I.; AI2 : Training a big data machine to defend, IEEE International Conference on Big Data Security, 2016.

21. http://news.mit.edu/2016/ai-system-predicts-85-percent-cyber-attacks-using-input-human-experts-0418 (pristupljeno : 17. 07. 2020.)

22. https://towardsdatascience.com/cyber-security-ai-defined-explained-and-explored-79fd25c10bfa (pristupljeno : 17. 07. 2020.)

23. M. Husak.; J. Čegan.; PhiGARo: Automatic Phishing Detection and Incident Response Framework, 9th International Conference on Availability, Reliability and Security, ARES, 2014.

24. Maiorana E., Huang C., Han J., Zhang X., Liu J.; Automatic Identification of Honeypot Server Using Machine Learning Techniques, 2019.

25. https://www.csoonline.com/article/3384702/what-is-a-honeypot-a-trap-for-catching-hackers-in-the-act.html (pristupljeno : 18. 07. 2020.)

26. Kubovič O.; Košinar P.; Janošik J.; Can Artificial Intelligence Power Future Malware, Research Desk, 2020.

27. https://www.weforum.org/agenda/2019/06/ai-is-powering-a-new-generation-of-cyberattack-its-also-our-best-defence/ (pristupljeno : 18. 07. 2020.)

POPIS SLIKA