CARNet
HRVATSKA AKADEMSKA I ISTRAŽIVAČKA MREŽA
CROATIAN ACADEMIC AND RESEARCH NETWORK

Napad na MD5 algoritam

CCERT-PUBDOC-2009-04-260

+CERT.hr u suradnji s LSS

Sigurnosni problemi u računalnim programima i operativnim sustavima područje je na kojem CARNet CERT kontinuirano radi.

Rezultat toga rada je i ovaj dokument, koji je nastao suradnjom CARNet CERT–a i LS&S–a, a za koji se nadamo se da će Vam koristiti u poboljšanju sigurnosti Vašeg sustava.

CARNet CERT, www.cert.hr

Nacionalno središte za **sigurnost računalnih mreža** i sustava.

LS&S, www.LSS.hr

Laboratorij za sustave i signale pri Zavodu za elektroničke sustave i obradbu informacija Fakulteta elektrotehnike i računarstva Sveučilišta u Zagrebu.

Sadržaj

1. Uvod

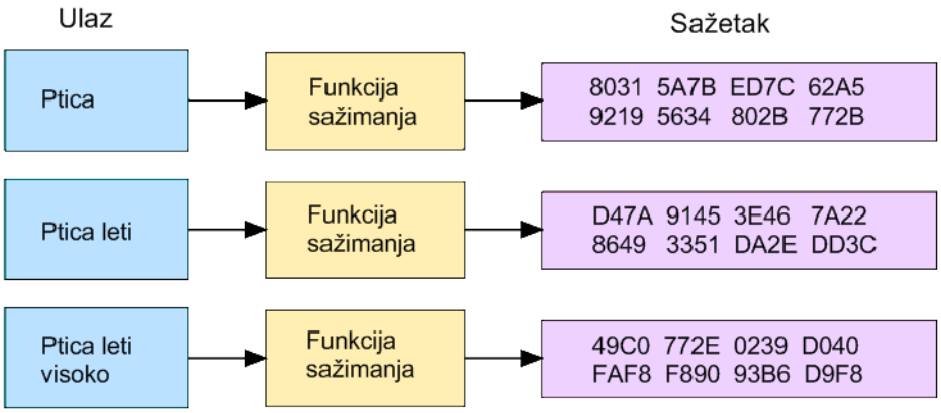
Kriptografski sažetci (eng. hash, cryptographic digest) u širokoj su upotrebi u današnje vrijeme. Za izračunavanje sažetka poruke koriste se posebne funkcije (eng. hash functions). Te funkcije stvaraju svojevrsni digitalni otisak određene veličine. Spomenuti je otisak zapravo niz znakova koji se obično zapisuju u heksadekadskoj notaciji. Dobre funkcije za izračunavanje sažetaka za različite ulazne podatke daju različite sažetke, odnosno dobiveni je sažetak jedinstven za svaku pojedinu poruku. To svojstvo kriptografskih sažetaka koristi se u kriptografiji za očuvanje integriteta (besprijekornosti) i neporecivosti poruke te dokazivanje identiteta. Funkcije za računanje sažetka poruke imaju mnogo primjena u području sigurnosti, a neke od njih su upotreba u digitalnim potpisima i PKI (eng. Public Key Infrastructure) sustavima, zatim u sustav tzv. e-novca (novac koji se razmjenjuje isključivo elektroničkim putem, odnosno putem Interneta) te u mnogim kriptografskim protokolima.

Ukoliko se otkrije da algoritam za stvaranje sažetaka ne računa jedinstvene sažetke, odnosno ako za dvije različite ulazne poruke postoji isti sažetak tada se kaže da je došlo do sukoba i algoritam više nije siguran. Danas je vrlo raširena upotreba MD5 algoritma za računanje sažetaka. Otkriveno je još 1993. godine, a potvrđeno 2004. godine da MD5 algoritam sadrži ranjivosti te da više nije dovoljno siguran za upotrebu u sustavima koji trebaju pružiti zaštitu korisnicima.

U ovom dokumentu definirane su funkcije za računanje sažetaka općenito, njihova primjena u PKI sustavima i u digitalnom potpisu. Osim toga, opisan je MD5 algoritam i njegove ranjivosti, tehnike za otkrivanje sukoba u MD5 algoritmu, kao i primjeri zlouporabe ranjivosti MD5 algoritma.

2. Funkcije za računanje sažetka poruke

Osnovu postupka za utvrđivanje besprijekornosti čini matematički postupak za izradu kriptografskog sažetka poruke. Kriptografski sažetak izrađuje se funkcijom za računanje sažetka. Te funkcije imaju ključnu ulogu u kriptografiji i zaštićenom načinu komuniciranja. Funkcije za izračunavanje sažetka kao ulaz primaju poruku i proizvode izlaz koji se naziva sažetak poruke (eng. hash, message digest), kao što je moguće vidjeti na slijedećoj slici.



Slika 1. Računanje sažetka poruke

Preciznije, funkcija za izračunavanje sažetka je deterministički algoritam koji preslikava niz proizvoljne duljine u niz utvrđene duljine, npr. *n* bitova. To je jednosmjerna funkcija koja iz poruke proizvoljne duljine računa sažetak stalne duljine. Funkcija je jednosmjerna jer je izračunavanje sažetka vrlo lako, dok je iz sažetka praktički nemoguće izračunati izvorni tekst. Jednosmjerne funkcije omogućuju očuvanje besprijekornosti (integriteta) i neporecivosti poruke. Sažetak poruke se matematički izračunava za svaku pojedinu poruku i različit je od poruke do poruke. Izračunava se matematičkim metodama čiji se opis nalazi u funkcijama sažimanja. Spomenute funkcije komprimiraju niz bitova poruke u niz određene veličine i vrijednosti. Kompresija ili računanje sažetka mora se obaviti tako da je nemoguće od neke druge poruke istom metodom dobiti istu vrijednost sažetka. Dakle, osnovna ideja funkcija za izračunavanje sažetka poruke je dobivanje sažetog zapisa, jedinstvenog za pojedini niz znakova, iz kojeg je nemoguće nekim postupkom dobiti izvornu poruku. Funkcije mogu neku poruku komprimirati na mnogo načina i ograničiti duljinu njenog sažetka na recimo 128 bitova. Vrijednost sažetka (tih 128 bitova) mora biti jedinstvena samo za tu poruku.

Osnovna obilježja funkcija za izračunavanje sažetka poruke su:

1. Računaju sažetke fiksne duljine iz ulaznog niza podataka proizvoljne duljine.
2. Ireverzibilne su, odnosno iz sažetka se ne može izračunati izvorni niz podataka.
3. Postoji mogućnost sukoba (eng. collision) – zbog fiksne duljine sažetka, dva različita ulazna niza podataka mogu rezultirati istim vrijednostima sažetka .
4. Kako bi se izbjegli predvidivi sukobi, podaci koji se malo razlikuju rezultiraju potpuno različitim vrijednostima sažetka.

Iako se upotrebom sažetka veličine 128 bitova može dobiti velik broj različitih poruka, moguće je da se na određenom uzorku poruka dobiju dva ista sažetka, tj. da dođe do sukoba (kolizije). Kako bi se to izbjeglo, u praksi su se počeli koristiti duži sažetci poruka (160 bitova i više). Standardne duljine sažetka su od 64 bita do 256 bita (64, 128, 160, 256) i ovise o algoritmu sažimanja. Vjerojatnost pojave sukoba predstavlja bitnu mjeru kvalitete pojedine funkcije za računanje sažetka poruke. Stoga je kvaliteta funkcije usko povezana s duljinom rezultirajućeg sažetka.

Najjednostavniji oblik funkcije za izračunavanje sažetka je uzastopna uporaba XOR funkcije na nizu bitova koji se dobivaju dijeljenjem izvorne poruke na dijelove jednake duljine. Najpoznatiji algoritmi za izračunavanje sažetka su MD5 , SHA-1, Tiger i dr. U slijedećoj tablici dan je popis popularnih algoritama i neke njihove značajke.

Algoritam	Veličina sažetka (u bitovima)	Otpornost na sudare (složenost)	Otpornost na inverziju (složenost)
HAVAL	256/224/192/160/128	Da	
MD2	128	Skoro	
MD4	128	Da (2^8)	S greškama (2^102)
MD5	128	Da (2^5)	Ne
PANAMA	256	Da	
RadioGatún	Proizvoljne veličine	Ne	
RIPEMD	128	Da	
RIPEMD-128/256	128/256	Ne	
RIPEMD-160/320	160/320	Ne	
SHA-0	160	Da (2^39)	
SHA-1	160	S greškama (2^63)	Ne
SHA-256/224	256/224	Ne	Ne
SHA-512/384	512/384	Ne	Ne
Tiger(2)-192/160/128	192/160/128	Ne	
WHIRLPOOL	512	Ne	

Slika 2.Tablica s popisom popularnih algoritama

Funkcije za izračunavanje sažetka pogodne su i koriste se za provjeru integriteta podataka. Osim toga, spomenute funkcije imaju široku primjenu na području računalne sigurnosti. Primjenjuju se :

- u digitalnim potpisima,
- u PKI sustavima,
- u kodovima za autentikaciju poruke (eng. Message Authentication Codes – MAC),
- u protokolima SSL/TLS (eng. Secure Socket Layer/Transport Layer Security)
- PGP (eng. Pretty Good Privacy) sustavima za kriptiranje i autenticiranje podataka,
- u protokolu S/MIME (eng. Secure / Multiple Internet Mail Extension) za osiguranje integriteta poruka,
- u protokolu IPSec (eng. IP Security) koji predstavlja skup protokola za sigurnu razmjenu IP paketa
- za indeksiranje podataka u tablicama sažetaka (eng. hash tables),
- za otkrivanje jednakih podataka,
- za jedinstvenu identifikaciju datoteka,
- kao zbroj za provjeru (eng. checksum),
- za otkrivanje korupcije podataka itd.

Kodovi za autentikaciju poruke omogućuju autentikaciju poruke simetričnim tehnikama. Simetrični kriptosustavi koriste isti ključ za kriptiranje i dekriptiranje poruka. MAC algoritam stvara kratki niz podataka koji se koristi za provjeru autentičnosti poruke i sastoji se od funkcije sažimanja i tajnog ključa. Kodovi za autentikaciju poruka računaju se i provjeravaju upotrebom istog ključa, tako da ga samo oni kojima je poruka namijenjena mogu provjeriti.

3. Primjena funkcija za računanje sažetka

3.1. Digitalni potpis

Digitalnim potpisom (eng. digital signature - DS) se utvrđuje autentičnost elektroničkih dokumenata, kao što su elektroničko pismo, web stranica ili slika. Dokument je autentičan ako je poznat njegov autor i ako je moguće dokazati da nije neovlašteno izmijenjen. Vjerodostojnost (eng. authentication) potpisanih dokumenata provjerava se upotrebom kriptografskih metoda.

Postupak digitalnog potpisivanja dokumenta sastoji se od:

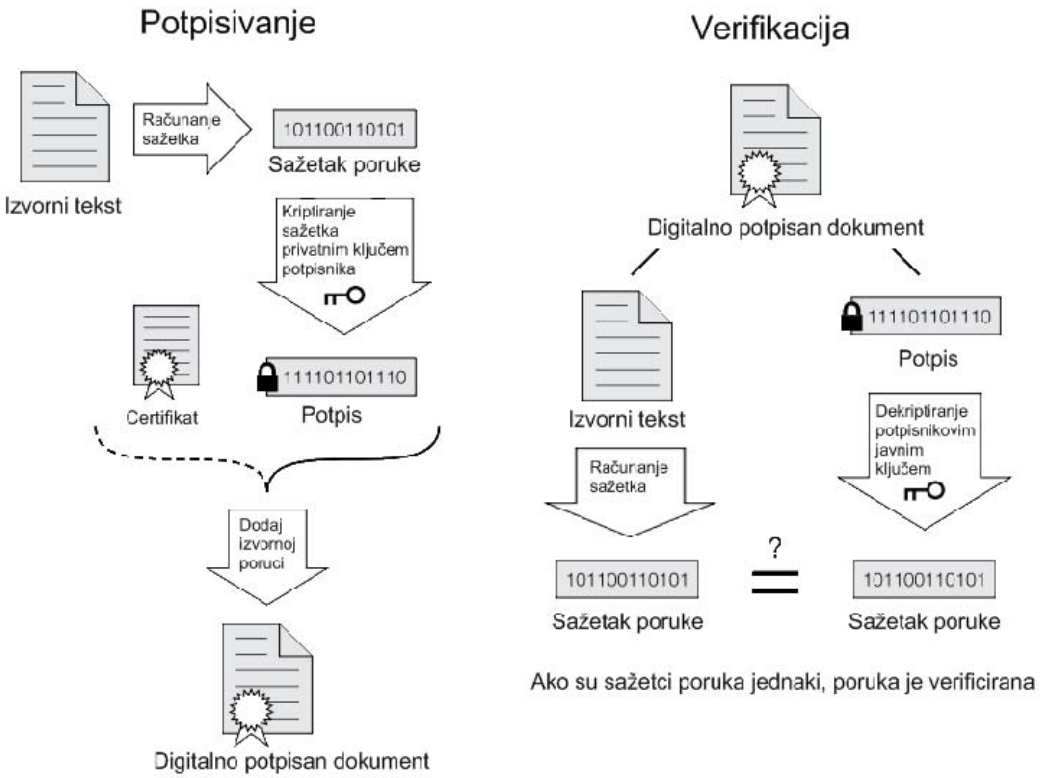
- izračunavanja sažetka poruke (izvornog teksta) – na primjer MD5 algoritmom i
- kriptiranja sažetka poruke.

Digitalni potpis osigurava:

- autentičnost - identitet pošiljatelja utvrđuje se dekriptiranjem sažetka poruke,
- integritet ili besprijekornost poruke - utvrđuje se je li poruka izmijenjena na putu do primatelja
- neporecivost - pošiljatelj ne može poreći sudjelovanje u transakciji jer jedino on ima pristup do svog privatnog ključa kojim je potpisao poruku

U stvaranju digitalnog potpisa, odnosno u postupku kriptiranja sažetka poruke koristi se asimetrični kriptografski postupak. Potpisnik stvara par ključeva, privatni i javni. Privatnim ključem kriptira sažetak poruke te tako stvoreni digitalni potpis šalje ili objavljuje zajedno s potpisanom porukom. Osnova sigurnosti digitalnog potpisa je u tajnosti privatnog ključa dok je javni ključ svima dostupan. Najčešći algoritam koji se koristi za kriptiranje sažetka poruke je RSA.

Sljedeći primjer opisuje digitalno potpisivanje dokumenta.



Slika 3. Digitalno potpisivanje dokumenta i verifikacija

Neka Ana i Matko razmjenjuju poruke. Ana želi osigurati autentičnost, besprijekornost i neporecivost poruke. Kako bi to postigla ona digitalno potpisuje poruku. Pri tome koristi svoj privatni ključ za kriptiranje sažetka poruke izračunatog na primjer MD5 algoritmom (privatni ključ inače služi za dekriptiranje!). Ana šalje poruku:

$$M = (P, RSA(H(P), K_{DA}),$$

gdje je P izvorna poruka (razgovijetni tekst), RSA algoritam kriptiranja, *H(P)* je sažetak poruke (izračunat korištenjem npr. MD5 algoritma) i K_{DA} je Anin privatni ključ.

Dakle, Ana je kriptirala sažetak poruke svojim privatnim ključem i u poruku koju šalje Matku uključila izvornu poruku i kriptirani sažetak te poruke.

Matko Aninim javnim ključem (koji inače služi za kriptiranje!) obavi dekriptiranje:

$$H(P) = RSA^{-1}(RSA(H(P), K_{DA}), K_{EA}),$$

gdje je H(P) sažetak poruke, RSA^{-1} postupak dekriptiranja, K_{DA} Anin privatni ključ, K_{EA} Anin javni ključ i $RSA(H(P), K_{DA})$ sažetak poruke kriptiran RSA algoritmom upotrebom Aninog privatnog ključa.

Matko je dekriptiranjem sažetka Anine poruke saznao dvije činjenice:

- da je poruku uistinu poslala Ana, jer samo ona i nitko drugi ne poznaje njezin privatni ključ
- da je pristigla poruka P besprijekorna

Kriptirani sažetak čini svojevrsni potpis, tj. digitalni potpis koji je Ana poslala uz poruku. Detaljni oblik ovog digitalnog potpisa ovisi o sadržaju poruke. On je za svaku poruku drugačiji.

3.2. Digitalni certifikat

Korisnici sve više koriste Internet za razmjenu osjetljivih podataka, kao što su web stranice banaka. Često osjetljivost podataka upućuje na to da se ne smiju razotkriti neovlaštenim osobama. Razmjena takvih podataka zahtjeva da korisnik ima osiguranje (uvjerenje) da je web stranica koju posjećuje valjana stranica organizacije s kojom korisnik posluje. Digitalni certifikati pružaju to uvjerenje.

Digitalni certifikat je elektronički dokument koji utvrđuje identitet i autenticira korisnika kada obavlja određene transakcije na Internetu. Certifikati koriste digitalne potpise za povezivanje javnih ključeva s podacima o identitetu vlasnika, kao što su ime osobe ili organizacije, adresa i sl., i time sprečavaju neovlaštenu izmjenu podataka. Certifikat sadrži identitet i javni ključ te ih povezuje u digitalni potpis. Javni se ključevi koriste u asimetričnoj kriptografiji. Asimetrični kriptosustavi zasnivaju se na određenim svojstvima brojeva koja se istražuju u teoriji brojeva. Takvi kriptosustavi imaju različite ključeve kriptiranja i dekriptiranja. Kod kriptiranja se obični (razgovijetni) tekst kodira kao niz prirodnih brojeva koji se odabranom funkcijom kriptiranja i javnim ključem kriptiranja preračunavaju u niz brojeva kriptiranog teksta. Funkcija kriptiranja mora biti takva da iz niza znakova kriptiranog teksta napadač ne može odrediti izvorni niz znakova. Međutim, poznavanje ključa dekriptiranja (privatnog ključa) omogućuje lako izračunavanje izvornog teksta.

Digitalne certifikate izdaje organizacija koja se naziva certifikacijska ustanova ili certifikator (eng. Certificate Authority – CA). Certifikator jamči da je prilikom stvaranja digitalnog potpisa provjerio identitet vlasnika javnog ključa i da je provjerio da taj vlasnik posjeduje odgovarajući privatni ključ. U digitalnom potpisu koristi se funkcija sažimanja, kao što je MD5 algoritam. Tkogod ima javni ključ certifikatora, on može provjeriti potpis certifikatora na certifikatu. Tako CA jamči da javni ključ u certifikatu pripada osobi čiji je identitet zapisan u istom certifikatu.

Svjetski prihvaćen standard za digitalne certifikate je X.509. Spomenuti certifikat sadrži slijedeće komponente:

- serijski broj
- vrijeme valjanosti
- naziv izdavača – identitet CA koji je izdao certifikat
- naziv subjekta – identitet stranke (osoba, organizacija, web stranica, drugi certifikator ili certifikator koji izdaje taj certifikat) koja se certificira
- javni ključ subjekta – javni ključ stranke koja se certificira
- osnovna ograničenja – polje koje sadrži bit koji ukazuje je li ovo certifikat CA ili korisnika
- digitalni potpis – stvara ga CA koji izdaje certifikat upotrebom svojeg privatnog ključa

Digitalnim potpisom u certifikatu CA jamči da je provjerio (poštujući svoje politike) da je identitet subjekta u certifikatu, upravo identitet vlasnika javnog ključa koji se također nalazi u istom certifikatu te da taj vlasnik posjeduje pripadni privatni ključ. Potpis u certifikatu može provjeriti bilo tko upotrebom certifikata CA koji je izdao javni ključ

3.3. PKI sustav

Infrastruktura javnog ključa (eng. Public Key Infrastructure - PKI) detaljno je opisana u dokumentu Nedostaci PKI infrastrukture (CCERT-PUBDOC-2009-02-255). PKI sustav je složen sustav koji se temelji na asimetričnoj kriptografiji.

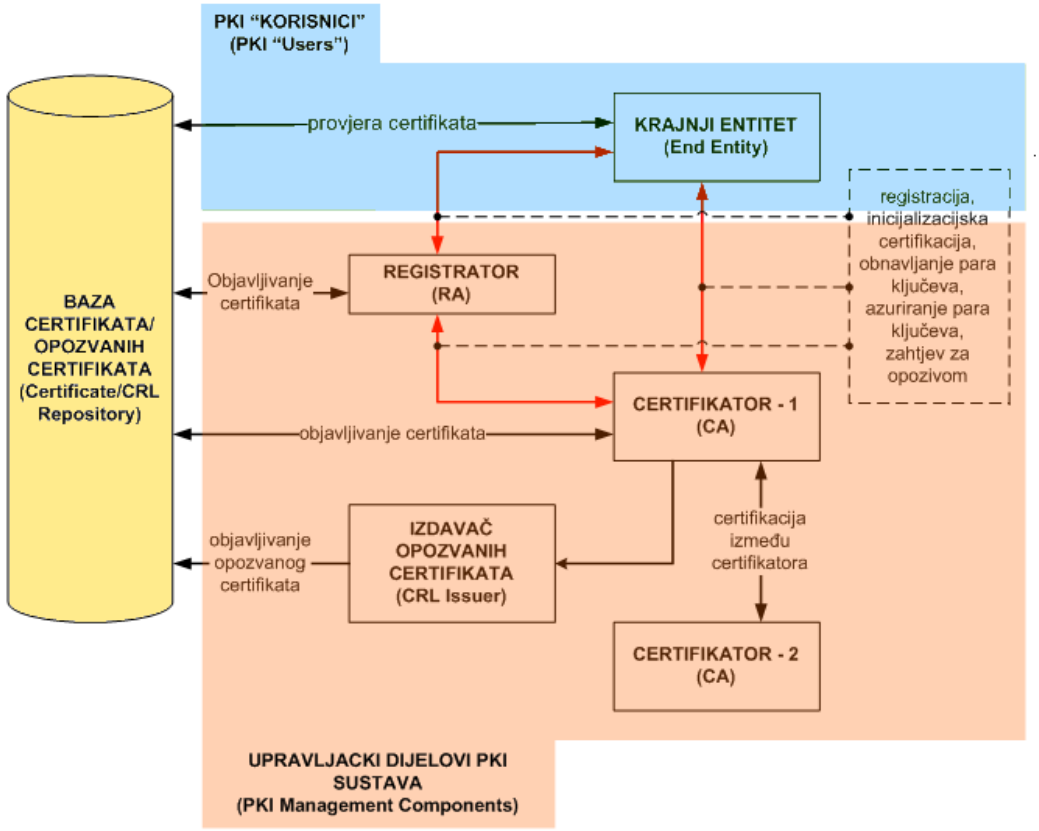
Infrastruktura privatnog ključa je skup sklopovlja, programskih paketa, ljudi, politika i procedura koje su potrebne za stvaranje, upravljanje, spremanje, distribuciju i opozivanje digitalnih certifikata. PKI sustav koristi i digitalne potpise pri stvaranju certifikata koje uključuje upotrebu sažetka poruke. Mnogo današnjih sustava još uvijek za računanje sažetka poruke koristi MD5 algoritam.

PKI je sporazum koji veže javne ključeve s njihovim korisničkim identitetima preko certifikacijske ustanove (certifikatora). Korisnički identitet mora biti jedinstven za svakog certifikatora. Povezivanje se ostvaruje putem registracije koje može obavljati program ili osoba, ovisno o razini povezivanja koje se obavlja. Dio PKI sustava koji osigurava povezivanje je registrator. Certifikator osigurava za svakog korisnika da je u certifikatima javnih ključeva nemoguće krivotvoriti korisnički identitet, javni ključ, njihovo povezivanje, uvjete utvrđivanja valjanosti i druge atribute.

Namjena PKI sustava je sigurna komunikacija preko nesigurnih kanala, a objedinjuje certifikate, certifikacijsku ustanovu (certifikator), spremnik certifikata i opozvanih certifikata, korisnike certifikata i sve njihove međusobne interakcije (interakcije između pojedinih elemenata sustava). PKI sustav omogućuje autentikaciju i pruža brojne usluge, kao što su osiguravanje povjerljivosti podataka, njihovog integriteta te upravljanje ključevima (eng. key managament), odnosno certifikatima.

ITU-T (eng. Telecommunication Standardization Sector) standard za PKI koji se koristi od 1988. godine je X.509. Spomenuti standard određuje, između ostalog, standardne formate za certifikate javnih ključeva, certifikate opozvanih popisa, certifikate atributa i algoritam za utvrđivanje valjanosti procesa izdavanja certifikata.

Osnovni dijelovi PKI sustava su krajnji entitet ili korisnici PKI sustava, certifikacijska ustanova ili certifikator (eng. Certification Authority - CA), registracijski centar ili registrator (eng. Registration Authority - RA), spremnik ili baza valjanih i opozvanih certifikata (eng. Certificate/CRL Repository) i izdavač opozvanih certifikata (eng. CRL Issuer). Slijedeća slika opisuje PKI sustav temeljen na X.509 standardu.



Slika 4. PKI sustav temeljen na standardu X.509

Slika opisuje proces objave certifikata, provjere certifikata, objavljivanje opozvanog certifikata te certifikaciju između dva certifikatora. Krajnji entitet ili krajnji korisnici ne moraju biti fizičke osobe, već mogu biti uređaji, kao što su poslužitelji i usmjerivači (eng. router), programi, odnosno sve što može biti identificirano certifikatom. Certifikacijska ustanova (CA) potpisuje i izdaje certifikate. Osnovna zadaća certifikatora je izdavanje certifikata, njihovo objavljivanje i po potrebi opoziv. Certifikator svojim potpisom jamči ispravnost podataka u certifikatu. On izravno ili preko registracijskog centra (RA) registrira krajnje entitete ili korisnike i utvrđuje njihov identitet na odgovarajući način. Certifikator obavlja ponekad i funkciju sigurnog pohranjivanja ključeva. Izvor je povjerenja u PKI sustavu, a povjerenje je osnova na kojoj se zasniva PKI sustav. Registracijski centar ili registrator je opcionalna komponenta PKI sustava. Može biti i dio certifikatora. Uloga registratora je vezana uz registriranje krajnjih entiteta, uz provjeru posjeduje li korisnik privatni ključ koji odgovara javnom ključu na certifikatu. On može biti posrednik između korisnika certifikacijske ustanove prilikom obavještavanja o neovlaštenoj izmjeni privatnog ključa. Registrator je također korisnik PKI sustava i prema tome ima svoj javni ključ i certifikat.

Baza certifikata je sustav ili skup distribuiranih sustava koji pohranjuju certifikate i popis opozvanih certifikata dostupnih svim unutarnjim, ali i vanjskim korisnicima PKI sustava koji koriste certifikate za identifikaciju.

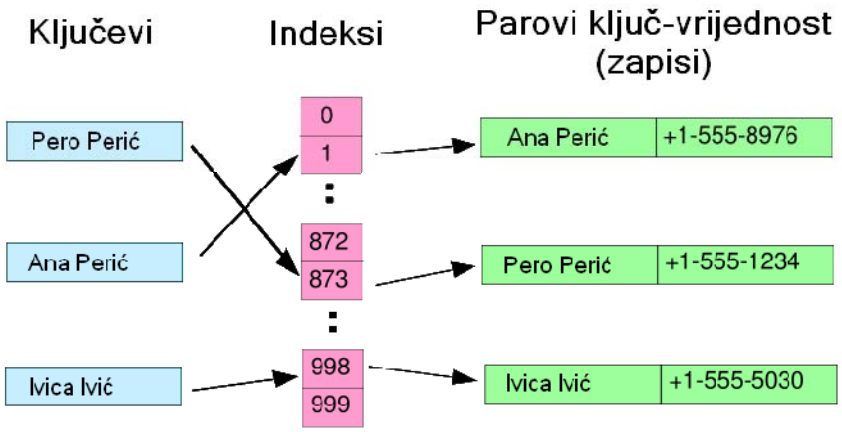
Izdavač opozvanih certifikata izdaje popis opozvanih certifikata. Valjanost certifikata je određena vremenskim razmakom, no certifikati mogu postati nevažeći i prije isteka vremenskog perioda. Na primjer certifikat se može opozvati ukoliko je privatni ključ neovlašteno promijenjen. Svaki opozvani certifikat identificiran je svojim serijskim brojem u popisu opozvanih certifikata koja je javno dostupna svima.

Uz sve prednosti PKI-a, kao i velik broj standarda vezanih uz njega, on nije ostvaren u stvarnosti onako kako se očekivalo. Postoji malen broj ostvarenih sustava temeljenih na infrastrukturi javnog ključa, a razloge tome je u složenosti realizacije i velikim troškovima izgradnje ovog sustava.

4. MD5 algoritam

MD5 (Message Digest algorithm 5), definiran dokumentom RFC 1321, je popularan algoritam za stvaranje sažetka poruke. Razvio ga je Ronald L. Rivest 1991. godine, i temelji se na starijem - MD4 algoritmu. MD4 algoritam je razvijen s namjerom da bude brz na 32 bitnim procesorima, ali zbog toga je bio rizičan u smislu kriptoanalitičkih napada. S druge strane, tvorci MD5 algoritma su se odrekli određenog koeficijenta brzine izvođenja kako bi ostvarili veću sigurnost.

MD5 algoritam je u širokoj upotrebi i pruža određenu dozu sigurnosti da će dokument koji se šalje stići na odredište besprijekoran. Na primjer, na poslužiteljima se često uz datoteke čuvaju i njihovi MD5 sažetci koji služe kao zbroj za provjeru (kako bi se potvrdilo da je datoteka koju korisnik preuzima upravo ona koja se i navodi). Korisnik može prenijeti datoteku na svoje računalo, izračunati njezin sažetak i usporediti ga sa zbrojem za provjeru na poslužitelju. Operacijski sustavi temeljeni na sustavu Unix koriste MD5 sažetke kao zbrojeve za provjeru. Poznato je da mnoge aplikacije koriste MD5 algoritam, a one uključuju komunikaciju SSL/TLS (eng. Secure Sockets Layer/ Transport Layer Security) i IPSec (eng. Internet Protocol Security) protokolima te ostale kriptografske protokole. Također, algoritam se koristi u mnogim implementacijama mehanizama za postavljanje vremenskih oznaka (eng. timestamps), u tablicama sažetaka (eng. hash tables), aplikacijama za provjeru integriteta, distribuiranim datotečnim sustavima, stvaranju slučajnih brojeva, PKI sustavima te za digitalno potpisivanje dokumenata. Na slijedećoj slici dan je primjer tablice sažetaka.



Slika 5. Tablica sažetaka za zapis telefonskih brojeva

Tablica sažetaka koristi se za brzo pretraživanje podataka. Svaki se ključ (npr. ime osobe) pridružuje različitom indeksu. Za pretvaranje ključeva u indekse koristi se funkcija sažimanja.

MD5 algoritam uzima ulaznu poruku proizvoljne duljine i stvara sažetak duljine 128 bitova. Ulazna se poruka ili izvorni tekst dijeli na blokove duljine 512 bitova. Zadnji blok teksta, koji ne mora biti potpun, nadopunjuje se na 512 bitova tako da se:

- iza zadnjeg bita teksta dodaje jedna jedinica,
- nakon te jedinice upisuje se toliko nula da u bloku preostanu 64 bita i
- u te se preostale bitove upisuje bitovna duljina izvorne poruke (bez nadopunjujućih bitova)

Svaki se blok dijeli na 16 podblokova duljine 32 bita koji se mogu nazvati:

$$M_0, M_1, M_2, ..., M_{15}$$

Svaki podblok M_j sudjeluje u izračunavanju sažetka četiri puta te se obavlja u 64 koraka podijeljena u četiri kruga. U svakom se koraku *i* ($1 \le i \le 64$) u izračunavanje uključuje i 32-bitna konstanta K_i, koja se dobiva kao

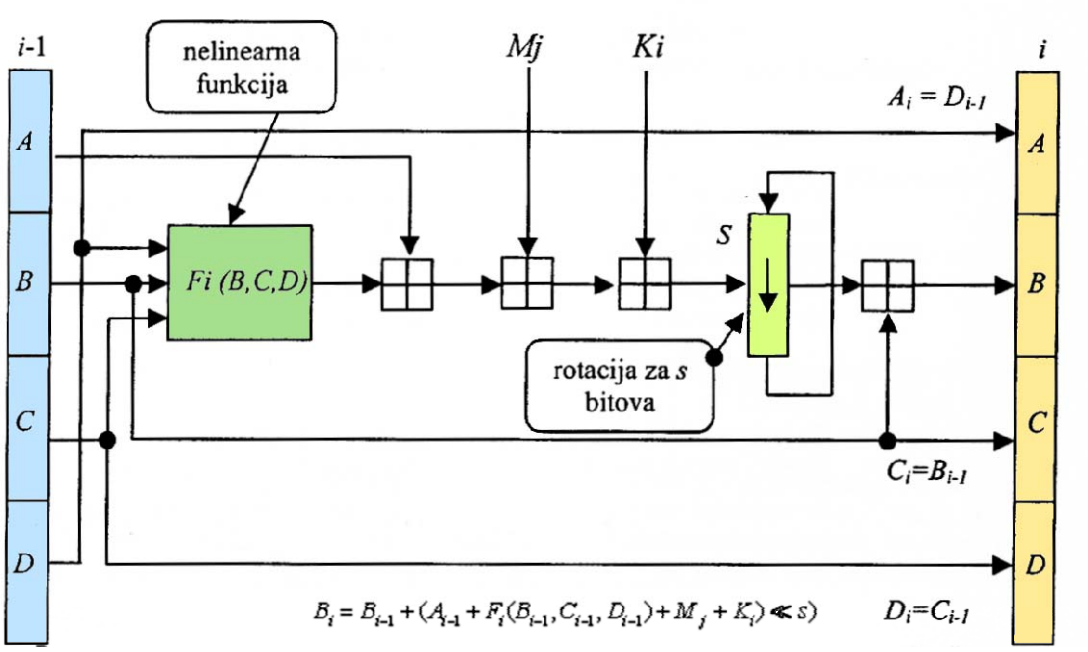
$$K_i = 2^{32} \cdot abs(\sin(i))$$

Sažetak S od 128 bitova sastoji se od nadovezane četiri 32-bitne varijable *A*, *B*, *C* i *D*. One se mogu početno inicijalizirati s konstantama:

$$A_0 = 01234567_{16} \qquad B_0 = 89ABCDEF_{16} \qquad C_0 = FEDCBA98_{16} \qquad D_0 = 76543210_{16}$$

Konstante A_0, B_0, C_0, D_0 čine inicijalizacijski vektor $IV = (A_0, B_0, C_0, D_0)$. Sve vrijednosti zapisane su u heksadekadskom brojevnom sustavu.

U *i*-tom se koraku izračunava:



A, B, C, D – 32 bitne riječi koje opisuju stanje algoritma

F – nelinearna funkcija

M_j – 512-bitni blok ulazne poruke

K_i – 32-bitna konstanta definirana za svaku operaciju

S – rotacija u lijevo za s mjesta, s definiran za svaku operaciju

operacija + modulo 2^{32}

Slika 6. MD5 postupak sažimanja poruke u i-tom koraku

U prvom krugu ($1 \le i \le 16$) se koristi nelinearna funkcija

$$F_i(X, Y, Z) = (X \wedge Y) \vee ((\neg X) \wedge Z)$$

Podblokovi se dovode redom:

$$M_0, M_1, M_2, M_3, ..., M_{14}, M_{15}$$

a rotacije u lijevo odvijaju se za s = 5, s = 19, s = 14 i s = 20 bitova u četiri uzastopna koraka, što se ponavlja četiri puta.

U drugom se krugu ($17 \le i \le 32$) koristi funkcija:

$$F_i(X, Y, Z) = (X \wedge Z) \vee (Y \wedge (\neg Z))$$

Podblokovi se dovode redom:

$$M_1, M_6, M_{11}, M_0, M_5, M_{10}, M_{15}, M_4,$$
$$M_9, M_{14}, M_3, M_8, M_{13}, M_2, M_7, M_{12}$$

a rotacije u lijevo obavljaju se za 5, 19, 14 i 20 bitova po četiri uzastopna koraka četiri puta.

U trećem se krugu ($33 \le i \le 48$) koristi funkcija:

$F_i(X,Y,Z) = X \oplus Y \oplus Z$, gdje je $\oplus$ XOR funkcija

Podblokovi se dovode redom:

$$M_5, M_8, M_{11}, M_{14}, M_1, M_4, M_7, M_{10},$$
$$M_{13}, M_0, M_3, M_6, M_9, M_{12}, M_{15}, M_2$$

a rotacije u lijevo obavljaju se za 4, 11, 16 i 23 bitova u četiri uzastopna koraka četiri puta.

U četvrtom se krugu ($49 \leq i \leq 64$) koristi funkcija:

$$F_i(X,Y,Z) = Y \oplus (X \vee (\neg Z))$$

Podblokovi se dovode redom:

$$M_0, M_7, M_{14}, M_5, M_{12}, M_3, M_{10}, M_1,$$
$$M_8, M_{15}, M_6, M_{13}, M_4, M_{11}, M_2, M_9$$

a rotacije u lijevo obavljaju se za 6, 10, 15, i 21 bitova po četiri uzastopna koraka četiri puta.

Konačnim vrijednostima varijabli A, B, C i D koje se nadovezuju u 128 bitovni sažetak, pribrajaju se konstante A_0, B_0, C_0, D_0, što daje:

$$A = A_{64} + A_0, \quad B = B_{64} + B_0, \quad C = C_{64} + C_0, \quad D = D_{64} + D_0$$

Na primjer, za poruku:

```
Zec trči preko livade
```

dobije se algoritmom MD5 sljedeći sažetak:

```
78d69ce92609dcb3590dce43382d53bb
```

Samo nagađanjem mogu se dobiti dvije poruke koje imaju isti sažetak poruke ili proizvesti poruku sa zadanim sažetkom poruke. Vjerojatnost da se dobiju 2 ista sažetka poruke ovisi o 2^{64} ($1.84467441 \times 10^{19}$, više nego što ima zvijezda u našoj galaksiji) operacija, a da se dobije željeni sažetak poruke ovisi o 2^{128} ($3.40282367 \times 10^{38}$)operacija.

1996. godine pronađena je ranjivost algoritma, odnosno sukob u algoritmu koji je, iako nije predstavljao kritičnu opasnost po algoritam, naveo stručnjake da preporučuju upotrebu drugih algoritama, kao što je SHA-1. Unatoč tome, MD5 je ostao u širokoj upotrebi do 2004. godine kada su otkrivene ozbiljne sigurnosne mane algoritma koje njegovu buduću upotrebu čine upitnom. U slijedećim poglavljima će biti više riječi o ranjivostima MD5 algoritma i upotrebi tih ranjivosti za različite vrste napada na zaštićene sustave.

5. Ranjivost MD5 algoritma

Sigurnost kriptografskih funkcija za računanje sažetka, kao što je MD5, ne temelji se na teoriji brojeva, kao što je to slučaj kod asimetričnih kriptografskih metoda. Kako bi funkcije sažimanja bile sigurne potrebno je zadovoljiti tri svojstva, odnosno za funkciju *h* s domenom *D* i kodomenom *R* treba vrijediti (za zadanu funkciju $h : D \to R$ skup *D* se naziva domenom, a skup *R* kodomenom funkcije.):

- **Prva otpornost na inverziju**: za zadani $y \in R$, ne smije biti moguće izračunati $x \in D$, takav da vrijedi $h(x) = y$, odnosno ne smije postojati $h^{-1}(y) = x$.
- **Druga otpornost na inverziju**: Za zadani $x \in D$, ne smije biti moguće izračunati određeni $x_0 \in D$, tako da je $h(x) = h(x_0)$.
- **Otpornost na sukob**: Ne smije biti moguće pronaći određeni x, $x_0 \in D$, takav da je $h(x) = h(x_0)$

1993. godine B. den Boer i A. Bosselaers otkrili su dvije poruke čiji su MD5 sažetci bili jednaki, odnosno otkrili su sukob u MD5 algoritmu. Ustanovili su da postoje pseudo-sukobi MD5 algoritma. Pseudo-sukob je sukob kod kojeg upotreba dva različita inicijalizacijska vektora daje jednaki sažetak. Dva se vektora za stvaranje dva sažetka razlikuju samo u najznačajnijem bitu svake od četiri varijable (A_0, B_0, C_0, D_0).

Zatim je 1996. godine H. Dobbertin izveo uspješan napad na MD5 algoritam i otkrio sukob upotrebom jednog proizvoljnog inicijalizacijskog vektora za dvije ulazne poruke. Iskoristio je dvije različite 512-bitne poruke s proizvoljnom vrijednosti inicijalizacijskog vektora:

$$A_0 = 12AC2375_{16} \qquad B_0 = 3B341042_{16} \qquad C_0 = 5F62B97C_{16} \qquad D_0 = 4BA763ED_{16}$$

To je bilo dovoljno da kriptografi preporuče upotrebu drugih algoritama za stvaranje sažetaka, kao što su SHA-1 i RIPEMD-160.

2004. godine znanstvenici sa sveučilišta Shandong u Kini, na čelu s Xiaoyun Wang, objavili su sukobe za MD5 algoritam, kao i za još neke druge funkcije za izračunavanje sažetka. Njihov analitički napad trajao je sat vremena na grozdu računala (eng. cluster), a za otkrivanje sukoba iskoristili su paradoks rođendana (u teoriji vjerojatnosti paradoks rođendana ili problem rođendana tvrdi da će u skupu slučajno odabranih ljudi biti nekoliko parova koji će imati rođendan na isti dan). 2005. godine ista je skupina znanstvenika demonstrirala konstrukciju dva PKI certifikata po standardu X.509 sa različitim javnim ključevima, ali istim MD5 sažetkom. 2006. godine Vlastimil Klima poboljšao je metode koje su koristili znanstvenici sa sveučilišta Shandong i objavio je algoritam koji je nazvao tuneliranje. Tom metodom moguće je pronaći sukob proizvoljnih poruka za manje od minute. 2007. godine skupina znanstvenika objava je metodu napada s nazivom o"Okrivanje sudara odabranog prefiksa" (dijelovi dokumenta koji se nalaze ispred dijela u kojem dolazi do sudara). 2008. godine grupa istraživača objavila je da su uspjeli uspješno iskoristiti MD5 sukobe za lažiranje certifikatora u PKI sustavu. Certifikati certifikacijske ustanove izgledali su valjano kada su bili podvrgnuti provjeri MD5 sažetkom, iako zapravo to nisu bili valjani certifikati. U nastavku je dan kronološki slijed otkrivanja nedostataka MD5 protokola.

Godina	Znanstvenici	Metode
1993.	B.den Boer i A. Bosselaers	pseudo-sukobi MD5 algoritma
1996.	H. Dobbertin	otkrivanjem sudara s poluslobodnim startom
2004.	X.Wang i H.Yu	diferencijalna kriptoanaliza i paradoks rođendana
2005.	X.Wang i H.Yu	konstrukcija dva PKI certifikata s istim MD5 sažetkom
2006.	Vlastimil Klima	diferencijalna kriptoanaliza i tuneliranje
2007.	skupina znanstvenika	otkrivanje sudara odabranog prefiksa
2008.	skupina znanstvenika	konstrukcija lažnog certifikatora upotrebom sukoba MD5 algoritma

Slika 7. Popis napada na MD5 algoritam

5.1. Tehnike pronalaska sukoba

Neformalno se može reći da je funkcija za računanje sažetka otporna na sukob ako je teško pronaći dva izvorna teksta za koje se dobije isti izlaz, odnosno isti sažetak. Za funkciju se kaže da je skoro otporna na sukob ako je teško pronaći dva izvorna teksta takva da se njihovi sažetci razlikuju u samo nekoliko bitova.

5.1.1. Upotreba inicijalizacijskog vektora

Prema inicijalizacijskim vektorima koji se koriste u traženju sudara, napadi na funkcije za računanje sažetaka mogu se podijeliti na slijedeće:

- **Napad otkrivanjem sudara** – sudari se pronalaze upotrebom konstantnih inicijalizacijskih vektora za dva različita izvorna teksta (sudar 1. tipa).
- **Napad otkrivanjem sudara s poluslobodnim startom** (eng. semi-free-start): sudari se pronalaze upotrebom istog slučajno ili proizvoljno odabranog inicijalizacijskog vektora za dva različita izvorna teksta (sudari 2. tipa).
- **Napad otkrivanjem pseudo-sudara**: sudari se pronalaze upotrebom slučajno ili proizvoljno odabrana dva različita inicijalizacijska vektora za dva različita izvorna teksta (sudari 3. tipa).

Napadi otkrivanjem sudara za MD5 algoritam otkrivaju sudare u više podblokova podataka (izvorni je tekst podijeljen na podblokove). Moguće je pronaći sudare u MD5 algoritmu bez otkrivanja sudara 1. tipa. Izvođenje napada uključuje otkrivanje skorih sudara (dva podbloka podataka su takvi da se njihovi sažetci razlikuju u samo nekoliko bitova) nakon obrade prvih podblokova (x_1, x_1') izvornog teksta. Takvi se sudari koriste za dobivanje sudara 3. tipa za druga dva različita podbloka podataka (x_2, x_2').

5.1.2. Upotreba paradoksa rođendana

Veličina sažetka MD5 algoritma je 128 bitova, što je dovoljno malo da se iskoristi paradoks rođendana. U teoriji vjerojatnosti paradoks rođendana ili problem rođendana tvrdi da će u skupu slučajno odabranih ljudi biti nekoliko parova koji će imati rođendan na isti dan. U skupini od barem 23 slučajno odabranih osoba, vjerojatnost je 50% da postoji barem jedan par čiji je rođendan isti dan. Za 57 ili više ljudi vjerojatnost postojanja takvog para je 99%. Ako pak u skupini postoji 367 osoba, vjerojatnost je 100% (jer je maksimalan broj rođendana 366).

Neka postoji popis skupine od 23 slučajno odabranih osoba. Ako se uspoređuje rođendan prve osobe na popisu s ostalima, 22 su mogućnosti za postojanje istog rođendana. Ukoliko se uspoređuje svaka osoba na popisu sa svima ostalima postoji 253 različitih mogućnosti za postojanje istog rođendana, odnosno u skupini od 23 osoba može postojati $\frac{23 \cdot 22}{2} = 253$ parova s istim rođendanom.

Slika 8.Paradoks rođendana za 23 slučajno odabrane osobe

Približna vjerojatnost da dvije osobe izabrane slučajno iz cijele populacije imaju isti rođendan je $\frac{1}{365}$ (ako se ne računa prijestupna godina) uz pretpostavku da su svi rođendani jednako vjerojatni. Ako se problem rođendana poopći, može se primijeniti na otkrivanje sudara kod funkcija sažimanja. Očekivani broj *n*-bitnih sažetaka koji se mogu stvarati prije nego što dođe do sudara je $2^{\frac{n}{2}}$. Ovo je razlog zašto je mali broj sudara kod računanja sažetaka neizbježan.

Ako je zadana funkcija *h*, cilj napada je pronaći dvije ulazne varijable x_1 i x_2 takve da vrijedi $h(x_1) = h(x_2)$. Takav par naziva se sukob. Metoda kojom se nalazi sukob računa funkciju *h* za različite ulazne vrijednosti koje mogu biti odabrane slučajno ili pseudo-slučajno. Postupak se odvija sve dok se isti rezultat ne nađe više puta.

Kod izvođenja ovakvog napada postoje dva povezana kriptografska izazova. Jednostavniji je otkrivanje dvije poruke koje imaju isti sažetak. Rivest i Dusse su zaključili da je težina pronalaska takve dvije poruke za MD5 algoritam reda 2^{64} operacija. Drugi je izazov pronaći izvornu poruku ako je zadan sažetak. Težina obavljanja spomenutog zadatka je reda 2^{128} operacija.

5.1.3. Diferencijalna kriptoanaliza

Jedna od najvažnijih metoda analize napada na funkcije za računanje sažetka poruke je diferencijalni kriptoanalitički napad. Općenito, spomenuti se napad uglavnom koristi za razbijanje kriptiranih blokova teksta. Diferencijalna analiza je u osnovi napad odabranim otvorenim tekstom i oslanja se na analizu razlika između dva otvorena teksta koji su kriptirani istim ključem (u slučaju primjene na sažetke poruke koristi se ista funkcija sažimanja). Diferencijalna kriptoanaliza može koristiti XOR logičku funkciju za otkrivanje razlika među tekstovima. Prvi koji je upotrijebio diferencijalnu kriptoanalizu kao metodu otkrivanja kriptiranog teksta bio je Murphy, a kasnije su metodu unaprijedili E.Biham i A.Shamir, koji su analizirali sigurnost DES algoritma kriptiranja. Oni su opisali diferencijalnu kriptoanalizu kao metodu koja analizira utjecaj određenih razlika u parovima izvornih tekstova na razlike u rezultirajućim parovima kriptiranih tekstova. Razlika dva parametra *X* i *X'* definira se kao $\Delta X = X' - X$. Za dvije poruke *M* i *M'*, $M = (M_0, M_1, ..., M_{k-1})$, $M' = (M_0', M_1', ..., M'_{k-1})$, potpuni diferencijal (razlika) za funkciju za računanje sažetka poruke definira se na sljedeći način:

$$\Delta H_0 \xrightarrow{(M_0, M_0')} \Delta H_1 \xrightarrow{(M_1, M_1')} \Delta H_2 \xrightarrow{(M_2, M_2')} \cdots\cdots \Delta H_{k-1} \xrightarrow{(M_{k-1}, M_{k-1}')} \Delta H,$$

gdje je ΔH_0 početna vrijednost razlike koja je jednaka 0. ΔH je izlazna razlika za dvije poruke. $\Delta H_i = \Delta IV_i$ je izlazna razlika za *i*-tu iteraciju i predstavlja početnu razliku za slijedeću iteraciju. Očito je da ako je $\Delta H = 0$ da postoji sukob za *M* i *M'*. Razlika koja daje sukob naziva se diferencijal sukoba.

Uz XOR funkciju često se u diferencijalnoj kriptoanalizi koristi i operacija modulo. Upotrebom kombinacije operacija modulo i XOR za svaki prolaz dobiva se više informacija nego da se koristi svaka funkcija zasebno.

6. Primjeri napada na MD5 algoritam

6.1. Otkrivanje sudara diferencijalnom kriptoanalizom

X.Wang i H.Yu sa sveučilišta Shandong u Kini koristili su diferencijalnu kriptoanalizu za otkrivanje sukoba u MD5 algoritmu. Proučavali su je li moguće pronaći par poruka, od kojih se svaka sastoji od dva 512-bitna bloka, koji proizvode sudare nakon drugog bloka. Preciznije, definirali su parove (M_0, M_1) i (M_0', M_1') takve da vrijedi:

$$\begin{aligned}(a,b,c,d) &= \mathrm{MD5}(a_0,b_0,c_0,d_0,M_0),\\ (a',b',c',d') &= \mathrm{MD5}(a_0,b_0,c_0,d_0,M_0'),\\ \mathrm{MD5}(a,b,c,d,M_1) &= \mathrm{MD5}(a',b',c',d',M_1'),\end{aligned}$$

gdje su a_0, b_0, c_0, d_0 inicijalne vrijednosti konstanti za MD5 algoritam. Pokazali su da je takve sukobe moguće naći efikasno, gdje za pronalazak prvih blokova (M_0, M_0') treba 2^{39} (otprilike 550 milijardi) MD5 operacija, a za pronalazak drugih blokova (M_1, M_1') treba 2^{32} MD5 operacija. Primjena napada na računalu IBM P960 traje otprilike sat vremena za pronalazak M_0 i M_0'. Za najbrži slučaj potrebno je petnaest minuta. Nakon toga, potrebno je samo između 15 sekundi i 5 minuta za pronalazak M_1 i M_1'.

Sljedeći postupak opisuje kako pronaći sudar podataka veličine dva 512-bitna bloka slijedećeg oblika:

$$H_0 \xrightarrow{(M_0, M_0'), 2^{-37}} \Delta H_1 \xrightarrow{(M_1, M_1'), 2^{-30}} \Delta H = 0.$$

1. Ponavljaj slijedeće korake dok se ne pronađe prvi blok:
   (a) Odaberi slučajnu poruku M_0
   (b) Promjeni M_0 posebnim tehnikama modifikacije poruka.
   (c) Tada, M_0 i $M_0' = M_0 + \Delta M_0$ daju prvu iteraciju razlike $\Delta M_0 \rightarrow (\Delta H_1, \Delta M_1)$ s vjerojatnošću 2^{-37}.
   (d) Provjera primjenom MD5 algoritma na M_0 i M_0'.
2. Ponavljaj slijedeće korake dok se ne otkrije sudar:
   (a) Odaberi slučajnu poruku M_1.
   (b) Promjeni M_1 posebnim tehnikama modifikacije poruka.
   (c) Tada, M_1 i $M_1' = M_1 + \Delta M_1$ daju drugu iteraciju razlike $(\Delta H_1, \Delta M_1) \rightarrow \Delta H = 0$ s vjerojatnošću 2^{-30}.
   (d) Provjera je li pronađeni par daje sudar.

Slijedeća tablica pokazuje dva para sudara za MD5 algoritam. H je vrijednost sažetka gdje je najznačajniji oktet posljednji (eng. little-endian) i bez proširenja izvorne poruke, a H* je vrijednost sažetka gdje je najznačajniji oktet prvi (eng. big-endian) s proširenjem izvorne poruke.

M_0	2dd31d1 c4eee6c5 69a3d69 5cf9af98 87b5ca2f ab7e4612 3e580440 897ffbb8 634ad55 2b3f409 8388e483 5a417125 e8255108 9fc9cdf7 f2bd1dd9 5b3c3780
M_1	d11d0b96 9c7b41dc f497d8e4 d555655a c79a7335 cfdebf0 66f12930 8fb109d1 797f2775 eb5cd530 baade822 5c15cc79 ddcb74ed 6dd3c55f d80a9bb1 e3a7cc35
M_0'	2dd31d1 c4eee6c5 69a3d69 5cf9af98 7b5ca2f ab7e4612 3e580440 897ffbb8 634ad55 2b3f409 8388e483 5a41f125 e8255108 9fc9cdf7 72bd1dd9 5b3c3780
M_1'	d11d0b96 9c7b41dc f497d8e4 d555655a 479a7335 cfdebf0 66f12930 8fb109d1 797f2775 eb5cd530 baade822 5c154c79 ddcb74ed 6dd3c55f 580a9bb1 e3a7cc35
H	9603161f a30f9dbf 9f65ffbc f41fc7ef
H'	a4c0d35c 95a63a80 5915367d cfe6b751
M_0	2dd31d1 c4eee6c5 69a3d69 5cf9af98 87b5ca2f ab7e4612 3e580440 897ffbb8 634ad55 2b3f409 8388e483 5a417125 e8255108 9fc9cdf7 f2bd1dd9 5b3c3780
M_1	313e82d8 5b8f3456 d4ac6dae c619c936 b4e253dd fd03da87 6633902 a0cd48d2 42339fe9 e87e570f 70b654ce 1e0da880 bc2198c6 9383a8b6 2b65f996 702af76f
M_0'	2dd31d1 c4eee6c5 69a3d69 5cf9af98 7b5ca2f ab7e4612 3e580440 897ffbb8 634ad55 2b3f409 8388e483 5a41f125 e8255108 9fc9cdf7 72bd1dd9 5b3c3780
M_1'	313e82d8 5b8f3456 d4ac6dae c619c936 34e253dd fd03da87 6633902 a0cd48d2 42339fe9 e87e570f 70b654ce 1e0d2880 bc2198c6 9383a8b6 ab65f996 702af76f
H	8d5e7019 61804e08 715d6b58 6324c015
H'	79054025 255fb1a2 6e4bc422 aef54eb4

Slika 9. Tablica sa sudarima za MD5 algoritam

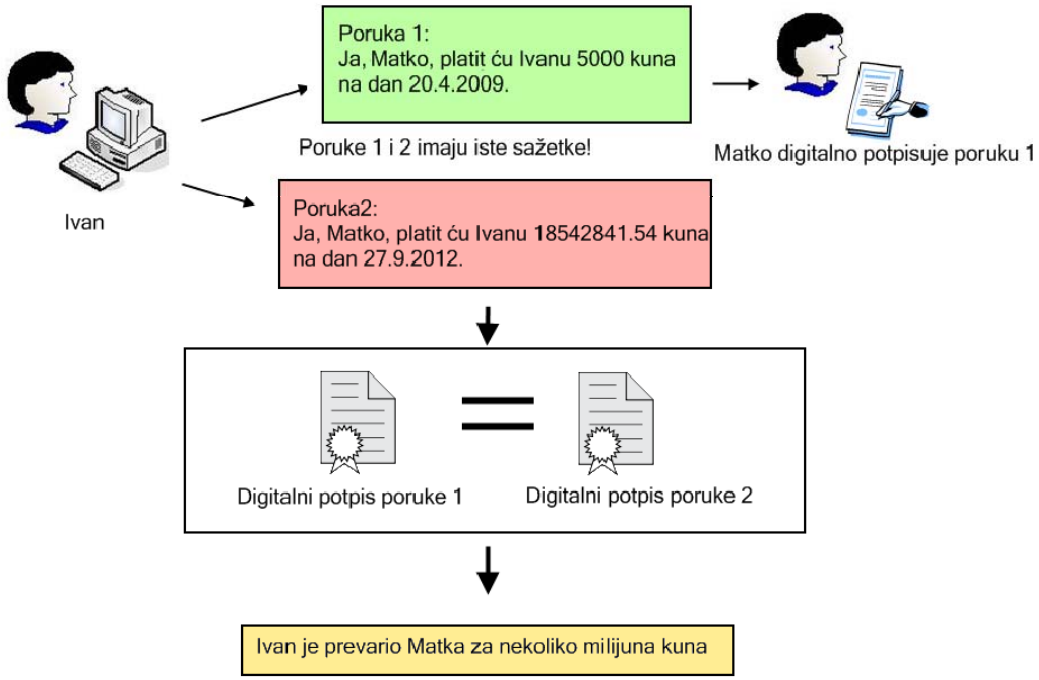
Dva sudara počinju istim prvim 512-bitnim blokom. Ako je zadan prvi blok koji zadovoljava određene uvjete, jednostavno je pronaći mnogo drugih blokova (M_1, M_1') koji dovode do sudara.

Nakon što se otkriju takvi parovi blokova, moguće ih je iskoristiti za lažno potpisivanje dokumenata. Zbog iterativnog načina rada algoritma za računanje sažetka, dva bloka podataka za koje je pronađen sudar mogu se ugraditi u mnogo veće dokumente. Pri tome treba biti zadovoljen uvjet da su dijelovi dokumenta koji se nalaze ispred dijela u kojem dolazi do sudara (prefiks) jednaki. Isti uvjet vrijedi i za dijelove dokumenta koji se nalaze iza dijela u kojem postoji sudar (sufiks). Kada su spomenuti uvjeti zadovoljeni računanjem sažetka cijelih dokumenata dobije se sudar.

U metodi otkrivanja sukoba odabranim prefiksom, dijelovi dokumenta koji prethode bloku u kojem dolazi do sudara mogu biti različiti. Sufiksi ipak moraju biti jednaki.

6.2. Narušavanje integriteta poruke

Napadač može sudar u MD5 algoritmu iskoristiti za narušavanje integriteta poruke. Obično stvori dvije poruke s istom vrijednosti sažetka. Jedna od poruka obično izgleda valjana ili bezopasna. Neka je napadač (Ivan) otkrio da poruka „Ja, Matko, platit ću Ivanu 5000 kuna na dan 20.4.2009.“ ima istu vrijednost sažetka kao poruka „Ja, Matko, platit ću Ivanu 18542841.54 kuna na dan 27.9.2012.“. Ivan bi tada mogao Matka nagovoriti da digitalno potpiše prvu poruku. Ivan tada može tvrditi da je Matko zapravo potpisao drugu poruku i dokazati to pokazivanjem da Matkov digitalni potpis odgovara drugoj poruci.



Slika 10. Prevara upotrebom sukoba

6.3. Napad na PKI sustav

Znanstvenici Alexander Sotirov, Marc Stevens, Jacob Appelbaum, Arjen Lenstra, David Molnar, Dag Arne Osvik, i Benne de Weger objavili su u prosincu 2008. godine kako je moguće iskoristiti sudare u MD5 algoritmu za lažiranje certifikata koje izdaje certifikacijska ustanova ili certifikator (CA). To znači da napadač može ugroziti sigurnost gotovo bilo koje web stranice, uključujući i stranice banaka. Neki od ugroženih PKI sustava su *RSA Data Security*, *Verisign* (Japan), *Thawte*, *FreeSSL*, *Rapid SSL*, i *TC TrustCenter* (Njemačka). Od 30 000 sakupljenih certifikata na Internetu otprilike njih 9000 koristi MD5 algoritam za digitalno potpisivanje.

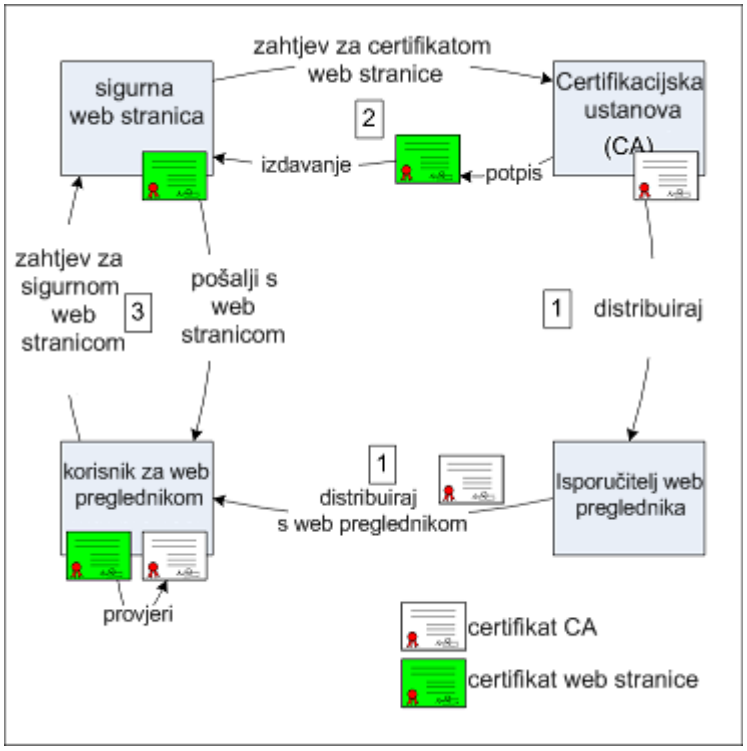
Dakle, otkrivena je ranjivost PKI sustava na Internetu koji se koristi za izdavanje digitalnih certifikata za zaštićene web stranice. Scenarij napada uključuje uspješno stvaranje certifikata certifikacijske ustanove ili certifikatora (CA). Stvoreni certifikat omogućuje lažno predstavljanje na bilo kojoj stranici na Internetu, uključujući i stranice banaka koje koriste HTTPS protokol.

Napad iskorištava ranjivost u MD5 kriptografskoj funkciji sažimanja koja omogućuje stvaranje različitih poruka s istom vrijednosti MD5 sažetka, odnosno iskorištava mogućnost postojanja sudara. Istraživanja sudara MD5 algoritma između 2004. i 2007. godine pokazala su da je moguće zlouporabiti MD5 algoritam u digitalnim potpisima. Rad objavljen u prosincu 2008., dostupan na web adresi http://www.win.tue.nl/hashclash/rogue-ca/ dokazuje da se barem jedan od teoretskih scenarija može praktično primijeniti.

Ako napadač koristi metodu napada s čovjekom u sredini (eng. man-in-the-middle attack), lažni certifikat kojeg posjeduje jamčit će da je veza sigurna te da je „druga strana“ ona čije ime stoji u certifikatu, iako je certifikat lažan. Žrtva napada nema nikakve indikacije da je veza koju koristi zapravo nesigurna jer će se u URL nizu pojaviti oznaka https:\\, slika lokota će se pokazati u web pregledniku i pojavit će se poruka „Ovaj je certifikat valjan“ i pripada odgovarajućoj instituciji.

Napad započinje slanjem zahtjeva za valjanim certifikatom web stranice komercijalne certifikacijske ustanove u koju „imaju povjerenje“ svi uobičajeni web preglednici. Kako je zahtjev valjan, certifikator potpisuje certifikat i vraća ga napadaču. Kako bi napad bio uspješan potrebno je izabrati PKI sustav koji koristi MD5 sažetak u potpisivanju certifikata. Zahtjev za certifikatom je posebno oblikovan da uzrokuje MD5 sudar u nekom drugom certifikatu. Taj drugi certifikat nije certifikat web stranice, već je certifikat jednog od posrednih certifikatora. On se može iskoristit za potpisivanje proizvoljnih certifikata neke web stranice. Kako su MD5 sažetci valjanog certifikata i lažnog certifikata certifikacijske ustanove jednaki, digitalni se potpis komercijalnog certifikatora može jednostavno kopirati na lažni certifikat CA. Na taj način lažni certifikat postaje važeći.

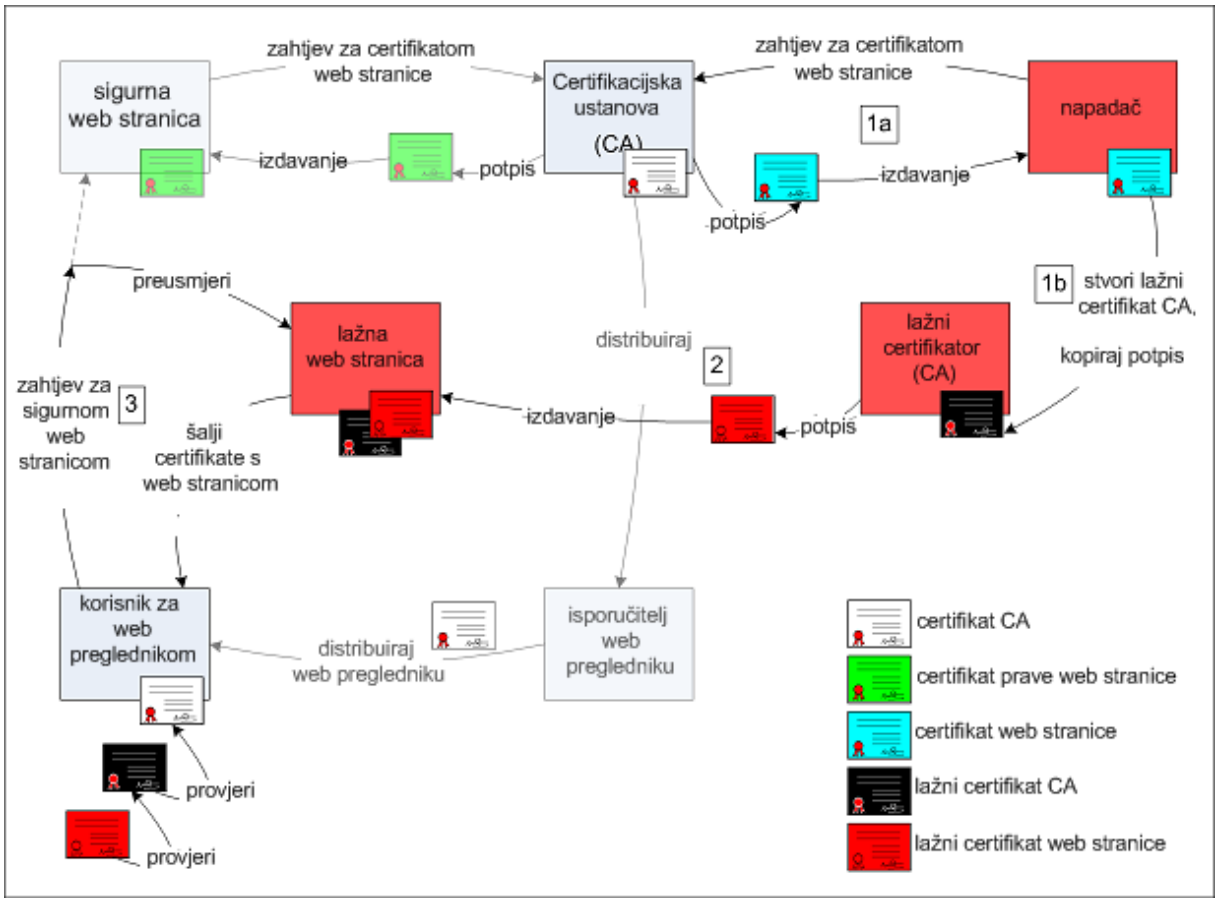
Slijedi objašnjenje i dijagram koji opisuje proces izdavanja certifikata web stranica:



Slika 11. Dijagram izdavanja certifikata web stranica

1. Certifikacijska ustanova (CA) izdaje svoje certifikate (bijele boje na slici) i prenosi ih preko isporučitelja web pregledniku. Ovi certifikati se nalaze na popisu povjerljivih certifikata na računalu korisnika. To znači da će svi certifikati koje je izdala ta certifikacijska ustanova biti prihvaćeni.
2. Tvrtka koja želi da njezina web stranica bude sigurna kupuje certifikate web stranica (zelene boje na slici) od certifikacijske ustanove. Taj je certifikat potpisao certifikator i on jamči identitet web stranice korisnicima.
3. Kada korisnik želi posjetiti sigurnu web stranicu, web preglednik prvo šalje upit web poslužitelju za certifikatom. Ako se digitalni potpis može provjeriti certifikatom certifikacijske ustanove s popisa na računalu korisnika, tada se prihvaća certifikat web stranice.

Slijedeća slika opisuje scenarij napada koji se može koristiti za oponašanje postojeće web stranice:



Slika 12. Scenarij napada na PKI sustav

1. a) Preuzimanje valjanog certifikata od komercijalne certifikacijske ustanove (označen plavom bojom na slici)

   b) Stvaranje lažnog certifikata CA (označen crnom bojom na slici). Ima potpuno jednak digitalni potpis kao i certifikat web stranice. Zbog toga izgleda kao da ga je izdao certifikator (koji ga zapravo nije vidio).
2. Tada se certifikatu web stranice (označen crvenom bojom na dijagramu) s identitetom prave web stranice pridružuje drugi javni ključ i potpisuje ga lažni CA. Stvara se kopija originalne stranice, stavlja na web poslužitelja i postavlja se na njega certifikat lažne web stranice.
3. Kada korisnik želi posjetiti sigurnu web stranicu, web preglednik će tražiti na internetu originalni web poslužitelj. Napadom preusmjeravanja komunikacija web preglednika preusmjeruje se prema lažnom web poslužitelju. Taj lažni poslužitelj predstavlja svoje certifikate korisniku zajedno s certifikatom lažnog CA. Web preglednik prihvaća lažni certifikat CA jer njegov digitalni potpis provjerava certifikat CA koji se nalazi na popisu povjerljivih certifikata (eng. trust list) na računalu korisnika. Korisnik neće primijetiti da nešto nije u redu.

Utvrđivanje valjanosti certifikata koje provode web preglednici može biti loše i zlonamjerni napadači mogu pratiti ili mijenjati podatke koji se šalju zaštićenim web stranicama. Web stranice banaka posebno su ugrožene. S lažnim certifikatom certifikacijske ustanove napadači mogu izvesti *phishing* napade koje je nemoguće razotkriti. Infrastruktura certifikacijske ustanove namijenjena je sprečavanju spomenutog napada te nijedan certifikator ne bi smio koristiti MD5 algoritam za stvaranje digitalnih potpisa. Kako se PKI sustav primjenjuje i u drugim područjima osim Interneta, mogući su i drugi scenariji napada. Neki od njih uključuju ugrožavanje sigurnosti elektroničke pošte, autentikacijskih kodova te druga područja koja koriste certifikate, digitalne potpise i kriptiranje javnim ključevima.

7. Posljedice ranjivosti MD5 algoritma i mjere zaštite

MD5 algoritam više ne pruža dovoljnu zaštitu integriteta podataka. Kako se još uvijek koristi u nekim PKI sustavima, napadači mogu upotrebom tehnika za otkrivanje sukoba u MD5 algoritmu stvarati lažne certifikate i digitalne potpise. Stvaranjem lažnih certifikata certifikacijske ustanove napadači mogu izvesti *phishing* napade koje je nemoguće otkriti te ugroziti web stranice banaka, a time i njihov rad. Osim spomenutih napada koji mogu imati drastične posljedice za korisnike zaštićenih web stranica, ugroženi su i svi sustavi koji koriste MD5 algoritam. Na primjer, sudari mogu predstavljati problem za sustave koji koriste kodove za autentikaciju poruke. Napadač otkrivanje sudara može iskoristiti za stvaranje dva programa s istim sažetkom. Jedan je program bezopasan, dok je drugi zlonamjeran. Ako netko digitalno potpiše bezopasni program, napadač ga može zamijeniti sa onim zlonamjernim i nanijeti štetu žrtvi napada.

Kako bi se izbjegli opisani scenariji napada potrebno je koristiti druge algoritme sažimanja kod kojih još uvijek nisu otkriveni sudari. Neki od njih su RIPEMD-160, SHA-512, SHA-256, WHIRLPOOL, itd.

Urednik sigurne web stranice može otkriti koji se algoritam sažimanja koristi u certifikatima klikom miša na ikonu lokota u web pregledniku kada uređuje sigurnu web stranicu. Ako certifikat web stranice ili posredne certifikacijske ustanove koristi za digitalno potpisivanje MD5 algoritam, tada taj algoritam treba zamijeniti sigurnijim, kao što je npr. SHA-256.

Osim promjene upotrebe algoritma za računanje sažetka poruke mogu se koristiti EV (eng. Extended Validation) certifikati. To su certifikati koji nadograđuju postojeći format certifikata, ali pružaju dodatni sloj zaštite. Taj je sloj zaštite definiran u procesu posebno oblikovanom za utvrđivanje i provjeru identiteta entiteta (fizičke osobe, uređaji, kao što su poslužitelji i usmjerivači (eng. router), programi, odnosno sve što može biti identificirano certifikatom) koji posjeduje certifikat. Kako bi se osigurala besprijekornost procesa, definirane mjere opoziva omogućuju brzo i efektivno opozivanje pogrešno izdanih ili korištenih certifikata. Proces izdavanja EV certifikata strogo je definiran u EV smjernicama (eng. guidelines) koje određuju sve korake koje certifikacijska ustanova mora napraviti prije nego što izda certifikat. Ti koraci su:

- provjera postojanja legalnog, fizičkog i funkcionalnog entiteta
- provjera da identitet entiteta odgovara službenim zapisima
- provjera da entitet ima isključivo pravo koristiti domenu određenu EV certifikatom
- provjera da entitet ima ispravno autorizirano izdavanje EV certifikata

EV certifikati su dostupni za sve tipove poslova, uključujući i vladina tijela te korporacije. Drugi niz smjernica su „EV Audit“ smjernice koje određuju kriterije prema kojima CA treba biti ispitan prije nego izda EV certifikate. Ispitivanje se treba ponavljati godišnje kako bi se osigurala besprijekornost procesa izdavanja.

EV certifikate podržavaju svi glavni web preglednici. Spomenuti certifikati ne smiju koristiti MD5 algoritam te su uz navedene smjernice sigurni od napada koji uključuje lažiranje certifikata.

8. Zaključak

Funkcije za računanje sažetka imaju široku primjenu u kriptografiji. MD5 algoritam je jedan od algoritama koji se često koriste u PKI sustavima, za digitalno potpisivanje dokumenata, u kodovima za autentikaciju itd. Već je 1993. godine otkriveno da bi algoritam mogao postati nesiguran, a u razdoblju od 2004. godine do 2008. godine različitim primjerima to je i potvrđeno. Različitim tehnikama otkrivanja sudara u MD5 algoritmu dokazano je da njegova upotreba može ugroziti zaštićene sustave. Posebno su ugroženi PKI sustavi koji još uvijek koriste MD5 algoritam. Jedina mjera zaštite je upotreba boljeg i sigurnijeg algoritma za računanje sažetka poruke. Sigurniji algoritmi su oni kod kojih još uvijek nisu pronađeni sukobi i ireverzibilni su. Neki takvi algoritmi su RIPEMD-160, SHA-256, SHA-512, WHIRLPOOL i drugi Osim upotrebe sigurnijih algoritama PKI sustavi bi trebali koristiti EV certifikate kako bi se osigurali od lažiranja certifikata. Neka istraživanja ukazuju da uklanjanje MD5 algoritma iz uporabe neće biti lako. Prikupljanjem certifikata na Internetu ustanovilo se da je postotak korisnika MD5 certifikata na sigurnim Web stranicama oko 3%. Ako se gleda statistika po sjednicama, broj korisničkih sjednica na koje bi utjecalo ukidanje MD5 algoritma je oko 6%. Ovi se postoci možda ne čine značajnima, ali upućuju na to da će ukidanje MD5 certifikata utjecati na velik broj korisnika.

Otkrivanjem novih metoda napada na MD5 algoritam ustanovljeno je da više nije siguran za upotrebu i da ga se zbog toga treba ukinuti bez obzira na utjecaj njegovog uklanjanja na korisnike.

9. Reference

[1] http://en.wikipedia.org/wiki/Cryptographic_hash_function , o funkcijama za računanje sažetka poruke

[2] http://os2.zemris.fer.hr/algoritmi/hash/2002_hofman/index.htm , MD5 algoritam

[3] http://tools.ietf.org/html/rfc1321 , MD5 algoritam

[4] http://en.wikipedia.org/wiki/MD5 , MD5 algoritam

[5] http://www.win.tue.nl/hashclash/rogue-ca/, lažiranje certifikata, 30. prosinac 2008.

[6] http://en.wikipedia.org/wiki/Digital_signature, digitalni potpis

[7] http://en.wikipedia.org/wiki/Birthday_paradox, paradoks rođendana

[8] http://www.cryptography.com/cnews/hash.html, posljedice ranjivosti funkcija za računanje sažetka

[9] How to Break MD5 and Other Hash Functions, Xiaoyun Wang, Hongbo Yu, Shandong University, Kina, 2004

[10] Collisions for the compression function of MD5, Bert den Boer, Antoon Bosselaers, srpanj 1993.

[11] Some thoughts on Collision Attacks in the Hash Functions MD5, SHA-0 and SHA-1, Praveen Gauravaram, William Millan, Juanama Gonzalez Neito, Information Securty Institute, Australia

[12] http://dev.chromium.org/developers/md5-certificate-statistics, statistika upotrebe MD5 certifikata

[13] http://www.cabforum.org/, EV certifikati