PATENTOVÝ SPIS

(19) ČESKÁ REPUBLIKA

ÚŘAD PRŮMYSLOVÉHO VLASTNICTVÍ

(21) Číslo přihlášky: **2007-509**
(22) Přihlášeno: **30.07.2007**
(40) Zveřejněno: **11.02.2009**
(Věstník č. 6/2009)
(47) Uděleno: **14.05.2010**
(24) Oznámení o udělení ve Věstníku: **23.06.2010**
(Věstník č. 25/2010)

(11) Číslo dokumentu:
301 799
(13) Druh dokumentu: **B6**
(51) Int. Cl.:

G06F 17/30	(2006.01)
G06F 19/00	(2006.01)
G06F 21/24	(2006.01)
H04L 9/14	(2006.01)
H04L 12/24	(2006.01)
H04L 12/26	(2006.01)

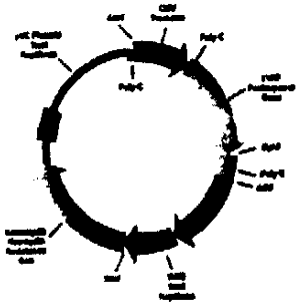
(56) Relevantní dokumenty:

J.Z.PONCE, M.LOEBL, L.KENCL, Packet Content Anonymization by Hiding Words, IEEE INFOCOM, Barcelona, 2006; R.Pang and V.Paxson, A high-level programming environment for packet trace anonymization and transformation, ACM SIGCOMM, 2003.

(73) Majitel patentu:
Kencl Lukáš Dr., Praha 2, CZ
Loebl Martin RNDr. CSc., Praha 5, CZ
Blamey Jenny, Santiago, CL

(72) Původce:
Kencl Lukáš Dr., Praha 2, CZ
Loebl Martin RNDr. CSc., Praha 5, CZ
Blamey Jenny, Santiago, CL

(74) Zástupce:
Ing. Květoslava Kubíčková, Doubravčická 2201, Praha 10, 10000



(54) Název vynálezu:
Způsob úpravy datové informace v systému

(57) Anotace:
Datové informace, reprezentované elektrickými nebo vlnovými signály nebo datové informace v umělých a přírodních databázích a záznamových mediích, například DNA, kódované jako řada znaků, se za účelem maskování a při současném zachování jejich vybrané lokální datové informace, rozdělí ve fyzickém médiu, zejména hardwaru počítače, fyzickém komunikačním kanálu, fyzickém záznamovém médiu či biologickém materiálu, na krátké překrývající se datové úseky. Délka těchto úseků je minimálně stejná nebo delší než je délka vybrané lokální datové informace, a sledovaná lokální datová informace je v krátkých úsecích obsažena celá. Tyto rozdělené krátké úseky vytvoří první skupinu a alespoň k jednomu krátkému úseku první skupiny se přidají data, kódovaná jako vybrané znaky, před nebo za znaky krátkých úseků první skupiny, přičemž vybrané znaky a jejich sled je stejný nebo různý a získaná směs původních krátkých úseků a modifikovaných úseků tvoří druhou skupinu a úseky druhé skupiny se vzájemně spojí v řadu dat. Celý proces lze několikrát opakovat. Předmětem je i použití způsobu úpravy datových informací ke konstrukci systému prvků, které sdílí lokální informace a jejich datových řad za účelem získání nových vlastností systému.

CZ 301799 B6

Způsob úpravy datové informace v systému

Oblast techniky

Vynález se týká úpravy informace za účelem jejího maskování a ochrany obsahu citlivé informace při ponechání pouze její části, sledované lokální datové informace, v systému obsahujícím alespoň jednu informaci. Systém je tvořen prvky a sdílením informací mezi nimi. Každá informace je všeobecně chápána jako datová informace, která je obsažena v přírodních databázích, například genetické datové informace uložené v DNA a vázané na chromatinová vlákna, nebo datová informace uložená v umělých databázích a záznamových médiích.

Datová informace je také reprezentovaná elektrickými nebo vlnovými signály, jako jsou zejména datagramy, pakety či stopy procházející jedním či více body v sítích Internetu v daném časovém úseku, nebo data ve vyhledávačích, nebo data generovaná operačními systémy a s nimi spojenými aplikacemi, nebo data zaznamenaná např. uzlem komunikační sítě nebo vyměňované informace obsažené v sítích operátorů. Datová informace je všeobecně kódovaná jako řada znaků vybraných ze skupiny tvořené písmeny, číslicemi nebo libovolnými symboly. Systém je charakterizován minimálně jednou řadou znaků.

Dosavadní stav techniky

Prvky se rozumí například uživatelé Internetu, různé pracovní skupiny jedné firmy nebo státní instituce, uživatelé stejné aplikace, například počítačového programu, nebo informačního systému, buňky v biologickém systému, například v mnohobuněčném organizmu, prvky mohou být také prostředníci sdílení datových informací, například komunikační kanály a uzly, prvky mohou být také skupiny prvků nebo podsystémy. Datové informace o prvku mohou být například genetická informace buňky nebo elektrické signály reprezentující stav operačního systému v daném okamžiku, nebo například elektrické signály nebo vlnové signály reprezentující identifikační údaje uložené v umělých databázích, například zaznamenání věku a adresy. Sdílením informací se rozumí například internetové komunikace nebo souhrnná komunikační data (např. stopa výměny dat mezi několika uživateli, zaznamenaná např. uzlem komunikační sítě) nebo vyměňované datové informace obsažené v sítích operátorů nebo zveřejnění hromadných dat například o vyhledávání v sítích. Sdílením informaci v biologických systémech se rozumí například zpracování různých fyzických signálů, nebo fragmentů genetické informace obsažené v genomické DNA buňkami, nebo částmi DNA jedné buňky.

Běžné statistické metody sbírají informace o prvcích a jimi sdílené informace nezbavené soukromého obsahu. Tyto metody předpokládají, že informace je sbírána autoritou mající obecnou důvěru, ze sebrané informace nezneužije a také je ochrání před nepřátelskými útoky. Tento předpoklad je nerealistický: jednotlivé prvky často nemají důvod věřit autoritě, která informace sbírá. Proto informace pro sběr nedávají a statistika je nemožná. Někdy se ochrana přístupu k citlivým informacím při zpracování či studiu dalším subjektem řeší dvoustrannými právními dohodami a značnými sankcemi při jejich případném úniku či zneužití. Nevýhodou takového postupu je značná pracnost, nerozšiřitelnost ve větším měřítku, komplikované či nemožné zpřístupnění dalším subjektům, obtížná i nákladná kontrola a vymahatelnost, obtížné a nákladné zabezpečení a značná neúčinnost (zejména ochrana před úniky způsobenými vlastním personálem).

Existují ovšem důležité postupy týkající se systémů, které potřebují pouze část informace. Příkladem takového postupu je detekce virů v internetové komunikaci, kde nebezpečí viru je indikováno tím, že internetová komunikace obsahuje krátký úsek z databáze nebezpečných úseků. Nebezpečí jiného ohrožení může být indikováno častým opakováním stejného krátkého úseku v internetové komunikaci.

Z hlediska postupů týkajících se systémů, které potřebují jen část datové informace, jsou důležité krátké úseky původní datové informace, které nazveme lokální datovou informací. Neuspořádanou skupinu více lokálních datových informací nazveme souborem lokálních datových informací. Výše uvedené vede k potřebě úpravy původní datové informace - kódované jako řada znaků, při které by se její obsah skryl a zároveň by se zachoval relevantní soubor lokálních datových informací.

Upravená datová informace obsahující soubor lokálních datových informací by se mohla sdílet a umožnila by studium a úpravy systému a podpořila by vzájemnou komunikaci mezi prvky. Návrh na úpravu datových informací byl zpracován v práci: Lukáš Kencl, Jose Zamora, Martin Loebl, „Packet Content Anonymization by Hiding Words", Demo na IEEE INFOCOM, Barcelona, Spain, April 2006, pomocí náhodného míchání souboru překrývajících se datových úseků. Ukázalo se nicméně, že tento postup nevede ke skrytí původní datové informace.

Z oblasti biologie je známé, že velká část genomu eukaryotů je tvořena podúseky DNA, které se mnohokrát opakují přesně nebo s malými odchylkami. V počítačové biologii je tento jev identifikován jako hlavní překážka současných metod rekonstrukce delších úseků DNA ze známého souboru kratších překrývajících se úseků. Důvodem je, že jestliže soubor obsahuje velké množství kratších úseků s opakujícími se počátečními nebo koncovými podúseky, existuje nezvládnutelně mnoho možných variací na rekonstrukci delších úseků, které jsou konzistentní s analýzou překrývání.

Podstata vynálezu

Datové informace, potřebné k řešení celé řady problémů různých systémů, jsou zejména z důvodů jejich možného zneužití často nedostupné. Jejich nedostatek nás přinutil zabývat se otázkou úpravy datových informací tak, aby jejich původní obsah se skryl, ale aby jejich lokální datové informace, které by byly předmětem studia systému, zůstaly zachovány.

Uvedené nedostatky týkající se zejména možného zneužití informací odstraňuje úprava datových informací, reprezentovaných elektrickými nebo vlnovými signály nebo datových informací v přírodních a umělých databázích a záznamových médiích, za účelem maskování a ochrany obsahu citlivých datových informací a současném ponechání jejich vybrané části, sledované lokální datové informace v systému obsahujícím alespoň jednu datovou informaci, kde původní úplná datová informace je kódovaná jako řada znaků vybraných ze skupiny tvořené písmeny, číslicemi nebo libovolnými symboly spočívající v tom, že se celá původní datová informace rozdělí ve fyzickém médiu, například v hardwaru počítače, fyzickém komunikačním kanálu, fyzickém záznamovém médiu či biologickém materiálu, na krátké překrývající se datové úseky, přičemž délka těchto krátkých úseků je minimálně stejná nebo delší než je délka sledované lokální datové informace, a sledovaná lokální datová informace je v krátkých úsecích obsažena celá, tyto rozdělené krátké úseky vytvoří první skupinu a alespoň k jednomu krátkému úseku první skupiny se přidají data kódovaná jako vybrané znaky před nebo za znaky krátkých úseků první skupiny, přičemž vybrané znaky a jejich sled je stejný nebo různý a získaná směs původních krátkých úseků a modifikovaných úseků tvoří druhou skupinu a úseky druhé skupiny se vzájemně spojí v řadu dat.

S výhodou se tyto spojené úseky druhé skupiny znovu zpracují podle předcházejícího postupu, přičemž rozdělení na krátké úseky a přidávání vybraných znaků se může provádět stejně nebo jinak a vzniklá další nová druhá skupina se opět spojí v řadu dat a tato procedura se několikrát opakuje za účelem rozšíření opakujících se podúseků, to je stejných posloupností znaků, jejichž počet a složení je náhodné. Opakující se podúseky mohou, ale nemusí obsahovat lokální informaci.

Tímto způsobem upravené datové informace, rozumí se datové řady a/nebo maskované datové řady vytvořené, například vlastním systémem pomocí opakujících se datových podúseků, jak je to známé například z oblasti biologie, lze použít ke studiu příslušného systému a zpracování údajů týkajících se lokální informace. Například se sleduje výskyt specifických krátkých úseků, relativní četnost stejných úseků nebo jejich závislost na čase.

Tímto způsobem upravené datové řady podle vynálezu a/nebo maskované datové řady vytvořené, například vlastním systémem pomocí opakujících se datových podúseků, lze dále použít ke konstrukci a koordinaci složitých systémů.

Tímto způsobem upravené datové řady a/nebo maskované datové řady vytvořené například vlastním systémem, pomocí opakujících se datových podúseků, lze použít pro kontrolu datových informací, například zpracování záznamů, stop, otisků či instancí operačních systémů, nebo počítačových aplikací, z hlediska bezpečnosti, korektností, důvěryhodnosti. Nebo je lze použít pro statistické analýzy, například často používaných instrukcí či datových bloků, například pro navázání bezpečné komunikace či spolupráce, pro jejich migraci či replikaci na jiné výpočetní zařízení či procesor, pro jejich kopírování na jiná zařízení či média za účelem úschovy či pro studium jejich chování v závislosti na čase.

Tímto způsobem upravené datové řady a/nebo maskované datové řady, vytvořené například vlastním systémem pomocí opakujících se datových podúseků, lze bezpečně použít například pro sdílení, zveřejňování, prodej, vyměňování a pro analýzu dat v sítích a vyhledávačích, nebo například stop, otisků, či instancí operačních systémů, počítačových aplikací či zaznamenané komunikace a interakce.

Tímto způsobem upravené datové řady a/nebo maskované datové řady, vytvořené, například vlastním systémem pomocí opakujících se datových podúseků, lze použít pro studium systému za účelem poznání jeho dynamického vývoje a rozpoznání, prevence, využití či izolace útoku, nákazy či jiného pozorovatelného jevu v systému.

S výhodou způsob použití spočívá v tom, že v nejméně jedné upravené a/nebo maskované datové řadě, kódované jako řada znaků se sledují skupiny opakujících se datových podúseků a/nebo jejich umístění v datové řadě a/nebo jejich průběžné změny v čase, a/nebo změny v jejich umístění v datové řadě, přičemž tyto změny mohou být vyvolané přirozeně nebo uměle.

Tento způsob sledování je vhodný k diagnostice abnormálního chování mechanismu maskování datové informace pomocí skupin opakujících se datových podúseků ve sledované části systému a ke koordinaci systému spočívající v prevenci vad systému, například nemocí v biologickém systému, způsobených abnormálním chováním mechanismu maskování informace pomocí skupin opakujících se podúseků v části systému.

Dlouhodobý pozorovatel datové řady upravené podle vynálezu a/nebo maskované datové řady, vytvořené například vlastním systémem, pomocí opakujících se datových podúseků si může po čase pozorování zrekonstruovat část původní datové informace.

Proto se dále s výhodou podle vynálezu průběžně provádí popsaná modifikace a konstrukce systému, která spočívá v tom, že se skupiny opakujících se datových podúseků obsažené v nejméně jedné upravené a/nebo maskované datové řadě, kódované jako řada znaků, průběžně modifikují například přidáním dalších dat, kódovaných jako znaky, k opakujícím se datovým podúsekům na jejich začátek a/nebo na jejich konec, a/nebo se v těchto datových podúsecích vymění skupina dat, kódovaných jako znaky, a/nebo se datové podúseky přemístí na jiné místo v upravené datové řadě a/nebo se do upravené datové řady vloží náhodně další datové podúseky v datové řadě již obsažené.

Tuto konstrukci a modifikaci si každý prvek systému může provádět podle vybraného technologického postupu sám.

Dále předmětem vynálezu je způsob konstrukce systému novou datovou informací, který spočívá v tom, že se ke krátkým úsekům obsahujícím soubor lokálních datových informací z nové datové informace přidají na jejich začátek a /nebo konec datové podúseky ze skupin opakujících se datových podúseků konstruované datové řady a výsledná skupina datových úseků se náhodně vloží, do nejméně jedné konstruované upravené a /nebo maskované datové informace.

Při sledování chování biologických systémů nás napadlo, že skupiny opakujících se podúseků v genomu mohou mít ochrannou maskovací funkci. Snaží se zabránit, aby se delší úseky DNA jednotlivých buněk - prvků biologického systému daly snadno zrekonstruovat ze znalosti kratších úseků.

Dále nás napadlo využít popsané empiricky pozorované těžkosti rekonstrukce řady znaků, která obsahuje skupiny opakujících se podúseků, ze souboru překrývajících se úseků, na návrh postupu, jak upravovat datovou informaci kódovanou jako řada znaků tak, aby se skryla původní datová informace, ale aby soubor lokálních datovaných informací krátkých datových úseků zůstal zachován pro statistická zpracování a jiné důležité postupy týkající se systémů.

Některé z těchto krátkých datových úseků se mohou najít například v databázi indikující počítačové viry. V počítačové biologii se také pozorovala korelace atypického vývoje skupin opakujících se krátkých úseků v DNA a konkrétních nemocí, například diabetes a schizofrenie.

DNA se obecně považuje za depositář genetické informace. Napadlo nás, že DNA, která obsahuje skupiny opakujících se podúseků, se chová jako upravená maskovaná informace. Soubor lokálních datových informací - krátkých úseků řady DNA je dostupný ostatním buňkám v biologickém systému k pozorováním. Napadlo nás, že toto chování DNA se dá využít při sledování a konstrukci biologických systémů ale také při konstrukci, sledování a koordinaci jiných systémů.

Nejdůležitějším přínosem úpravy informace podle vynálezu je maskování původní informace, kódované jako řada znaků, a tím skrytí obsahu a významu původní informace, zatímco relevantní soubor lokálních datových informací zůstane zachován. Prvky systému se nemusí obávat rekonstrukce původní datové informace či zneužití při předávání upravených datových informací respektive datových řad. K úpravě datové informace se použijí různé variace postupu podle vynálezu, které jsou dány k disposici každému prvku. Pomocí těchto různých variací postupu prvky svou původní datovou informaci samy upraví a dají ji k disposici pro studium a úpravy systému.

Výhodou úpravy datové informace podle vynálezu je umožnění různých použití týkajících se upravené datové informace v systémech. Jde například o studium a modelování chování a komunikace účastníků v komunikačních sítích na základě upravené datové řady záznamu komunikace (včetně jejího upraveného obsahu), což lze využít například při vytváření cílené reklamy, při vytváření aplikací pro koučink, analýzu, školení a zefektivnění činnosti jednotlivců a organizací či při optimalizaci provozu či odhalování závad či útoků v síti Internetu, kdy by ve všech těchto případech přítomnost citlivých údajů, obsažených v neupravené datové informaci, znemožňovala použití takovéto aplikace, zejména vyvíjené, provozované či užívané další stranou. Prostřednictvím upravené datové informace-záznamů vyhledávání v automatizovaných vyhledávačích lze studovat a modelovat chování uživatelů a na základě získaných údajů mohou provozovatelé vyhledávačů i další strany optimalizovat činnost vyhledávačů a cílené reklamy. Záznamy komunikace účastníků v sítích a vyhledávání ve vyhledávačích uživateli lze v podobě upravené datové řady podle vynálezu také komerčně nabízet k prodeji třetím stranám, zveřejňovat či vyměňovat mezi provozovateli sítí či vyhledávačů, neboť vysoká výpočetní složitost rekonstrukce původní (neupravené) datové informace zabraňuje zneužití citlivých informací o účastnících či uživate-

lích. Upravené datové informace podle vynálezu lze využít také například pro zpřístupnění databází zdravotních či biologických záznamů za účelem výzkumu či vývoje nových léčiv či léčebných či diagnostických postupů či modelování chování jednotlivců. Dále je lze využít při sdílení upravené datové řady ze záznamových médií (např. zvukové či kamerové systémy) pro případy ochrany bezpečnosti jinou stranou či povinnosti sdílet data se státními bezpečnostními orgány. Upravené datové informace lze také s výhodou využít při kontrole bezpečnosti, korektnosti při důvěryhodnosti jednotlivých instancí operačních systémů či aplikací nad nimi spuštěných, bez odhalení citlivých dat v těchto systémech obsažených.

Dále vynález umožňuje konstrukci a koordinaci systémů, které vedou ke zlepšení funkce systémů. V systému, kde upravená a/nebo maskovaná datová informace každého prvku je volně přístupná a sdílená celým systémem, je možná složitá koordinace umožňující například zabránit šíření nepříznivého jevu, například virů v internetové síti, nebo, jako je například koordinace systému buněk v mnohobuněčném organismu, kde sdílená maskovaná genetická informace každé buňky je obsažena v jejím DNA. Při sledování relevantní lokální datové informace a skupin opakujících se datových podúseků a jejich změn, a/nebo změn jejich umístění, v závislosti na čase, lze například provádět diagnostiku a prevenci chyb v systému nebo v jeho částech, a je také umožněno zjišťování a předpovídání globálních poruch systému způsobených poruchami mechanismu, provádějícího vlastní maskování informací. Lze například usuzovat na zdravotní stav biologických systémů.

Úpravou nové datové informace při jejím vkládáním do konstruované řady dat podle vynálezu lze docílit rychleji kladného výsledku a větší stability výsledného produktu, například v biologických systémech větší odolnosti proti autoimunitním reakcím.

Příklady provedení

Příklad 1 Zpracování předávaných datových informací na Internetu pro umožnění statistiky a studia Internetu.

V sítích Internetu jsou datové informace reprezentované jako elektrické a vlnové signály předávané jako datagramy, čí pakety. Datagramy či pakety jsou kódované řadou znaků určité délky, dané použitým protokolem. Pakety jsou rozděleny na hlavičku (header), obsahující výchozí a cílovou adresu a další kontrolní údaje, a tělo (payload, content), obsahující vlastní vyměňovaná data. Data vyměňovaná mezi dvěma či více uživateli mohou být na Internetu rozdělena do více paketů, které nemusí po Internetu cestovat bezprostředně po sobě, ani stejnou cestou. Záznam paketů procházejících jedním či více body Internetu v daném časovém úseku se obvykle nazývá stopa (trace). Tato stopa, či pouze její část bez hlaviček, představuje předávanou -sdílenou datovou informaci na Internetu a tvoří vstupní - původní datovou informaci kódovanou řadou znaků. Studiem takové stopy lze například odhalit předávaný počítačový virus porovnáním s databází vírů, či rychle se šířící Internetový červ (worm) díky náhlému častému výskytu konkrétních krátkých úseků. Stopa ovšem obsahuje i sdělení soukromé povahy, které operátor sítě nechce či nesmí zaznamenávat, zpřístupňovat či zveřejnit. Ke studiu stopy za účelem odhalení nebezpečného viru může však postačit soubor lokálních datových informací-krátkých úseků.

Úprava vstupní, původní datové informace kódované řadou znaků některou z variant postupu podle vynálezu upraví stopu na upravenou datovou řadu kódovanou řadou znaků, která obsahuje soubor lokálních datových informací z původní stopy, ale původní stopu je z ní prokazatelně výpočetně složité odvodit zejména proto, že upravená stopa obsahuje skupiny opakujících se datových podúseků. Úprava datové informace podle vynálezu může být provedena například podle následující varianty postupu.

Varianta je popsána pro případ, že chceme zachovat všechny lokální datové informace, které mají danou délku. Ale podobně je možno postupovat i v případě že lokální datové informace, které chceme zachovat jsou jiné, například nejsou stejně dlouhé. Označme tedy k délku lokální datové informace, kterou chceme v úpravě zachovat. Variantu postupu popíšeme pomocí procedur, pracujících s řadami znaků, které kódují datové informace.

Začneme popisem několika jednodušších procedur, postup se nakonec z těchto procedur poskládá. Cyklickou řadou míníme řadu-sled znaků, kde ztotožníme začátek a konec.

Procedura S1(s, o, z, Z): vstupem je cyklická řada s, a parametry délky úseků jsou čísla o, z, Z přičemž platí, že délka lokální datové informace k, je menší nebo rovna o, což je menší nebo rovno z, což je menší nebo rovno Z.

Tato procedura se skládá z následujících kroků. Nejprve se řada s rozdělí na úseky B1,... ,Bm tak že délka každého úseku Bi je náhodně vybrána z intervalu (y, Z). Úseky B1,...,Bm tvoří první skupinu. Dále se na začátek každého úseku první skupiny přidá koncový úsek předešlého úseku délky o. Dále se může na konec každého takto vzniklého úseku přidat úsek obsahující jeden až dva znaky, vybraný náhodně z řady s. Takto upravené úseky tvoří druhou skupinu. Nakonec se úseky druhé skupiny seřadí do řady v náhodném pořadí.

Procedura S2(s, z, Z): vstupem je cyklická řada s, a parametry délky úseků jsou čísla z. Z. Tato procedura se skládá z následujících kroků.

Nejprve se řada s rozdělí na úseky B1,... ,Bm tak že délka každého úseku Bi je náhodně vybrána z intervalu (z,Z). Úseky B1,... ,Bm tvoří první skupinu. Dále se na začátek každého úseku první skupiny přidá kopie celého předešlého úseku. Alternativně je například možno přidat jen podstatnou koncovou část předešlého úseku. Alternativně je také například možno ke každému vzniklému úseku na jeho konec přidat krátký úsek vybraný z řady s tak, aby koncový úsek tohoto nově vzniklého úseku se v souboru úseků opakoval. Takto upravené úseky tvoří druhou skupinu. Nakonec se úseky druhé skupiny seřadí do řady v pořadí, které je možno popsat takto: Každý úsek první skupiny se objeví jednou jako počáteční a jednou jako koncový úsek nějakého úseku druhé skupiny. Proto je možno popsat seřazení pomocí vzájemně jednoznačného zobrazení (permutace) p na úsecích první skupiny, tak, že úsek druhé skupiny končící úsekem Bi první skupiny bude v seřazení pokračovat úsekem druhé skupiny začínajícím úsekem Bp(i) první skupiny. Při alternativním provádění procedury se pracuje s podstatnými částmi úseků první skupiny namísto těchto samotných úseků. Při seřazení použijeme libovolnou permutaci, pro kterou dostatečně mnoho (například 10 %) indexů i splňuje p(p(i)-1)= i+1.

Procedura S3(s,o,z,Z): vstupem je cyklická řada s, a parametry délky úseků jsou čísla o,z,Z. Předpokládáme že řada s má tvar výstupu procedury S2, tj. řadu s je možno zapsat jako B1B2Bp(2)B(p(2)+1)... B(r-1)Br, kde r je index splňující p(r)= 1. Nejprve popíšeme, jak rozdělíme řadu s. Pro takový popis stačí, opsat rozdělení, které v řadě s provedeme. V obou výskytech (v řadě s= B1B2Bp(2)B(p(2)+1)...B(r-1)Br), každého úseku Bi první skupiny provedeme stejné rozdělení řady s tak aby úsek Bi byl rozdělen na dva úseky, přičemž počáteční úsek by měl délku alespoň o. Úseky takto rozdělené řady s tvoří novou první skupinu. Dále se na začátek každého úseku nové první skupiny přidá koncový úsek předešlého úseku délky o. Dále se může na konec každého vzniklého úseku přidat úsek krátké délky, vybraný náhodně z řady s. Takto upravené úseky tvoří novou druhou skupinu. Nakonec se úseky nové druhé skupiny seřadí do řady v náhodném pořadí.

Dvě základní varianty postupu úpravy datové informace nyní můžeme popsat jako S3(S2(S1^3 (s,k,k,3k/2),k,3k/2), k,k,3k/2) a S3(S2(s,k,3k/2),k,k,3k/2),

kde cyklická řada s vznikne ze vstupního řetězce ztotožněním jeho konce a počátku, k je délka lokální informace, kterou chceme uchovat, a S1^3 značí tři aplikace procedury S1, ale je možno procedury kombinovat i řadou jiných způsobů, například je možno zvolit jiný počet aplikací procedury S1 než tři.

Každý prvek systému může velmi snadno pomocí těchto variací svoji datovou informací upravit a dát ji k disposici ostatním neboť rekonstrukce citlivé původní datové informace-kódované jako řady znaků, na základě upravené datové řady je prokazatelně výpočetně složitá, protože existuje velké množství možných variací na rekonstrukci delších úseků, které jsou konzistentní s analýzou upravené řady znaků.

Vzniklá upravená datová řada umožňuje studium souboru lokálních informací původní stopy, například studium přibližné četnosti výskytu určitých vzorů-krátkých úseků naznačujících přítomnost viru. Mechanismus umožňuje rovněž zveřejňování, zpřístupnění či výměnu stop mezi prvky systému (operátory, uživateli, provozovateli), které si nemusí navzájem důvěřovat.

Příklad 2 Koordinace systému prvků - uživatelů operačních systémů.

Stavem operačního systémem nazýváme soubor konfigurací, počítače, které spravují nakládání se zdroji v počítači přítomnými, jako jsou například procesor, paměť, pevný disk, provádění programů, aplikace, apod.. Operační systém mj. sleduje stav jednotlivých aplikací nad ním spuštěných, a souborů uložených na pevném disku. V libovolném okamžiku se operační systém nachází v nějakém zcela zaznamenatelném stavu, reprezentovaném elektrickými signály a záznamy na záznamových médiích, které dohromady můžeme nazvat otiskem operačního systému. Otisk obsahuje například stav veškerých aplikací nad tímto operačním systému spuštěných a jím spravovaných a stav i obsah všech souborů daným operačním systémem spravovaných. Tento otisk operačního systému v daném okamžiku se všeobecně kóduje jako řada znaků. V rámci systému tato řada znaků představuje kódování původní datové informace prvku uživatele v daném okamžiku. Systém se vytvoří následujícím způsobem: Prvky uživatelé upraví otisk svého operačního systému například podle varianty popsané v prvním příkladě. Prvek-uživatel pak může zpřístupnit výslednou upravenou datovou řadu na přístupném místě, například na své webové stránce, kódovanou řadou znaků.

Podobně jako se v čase mění stav operačního systému, mění se také v čase jeho otisk, i jeho upravený otisk. Úpravy lze dosáhnout například opětovným prováděním úpravy celého otisku - vždy v určitém časovém okamžiku na příklad podle varianty popsané v prvním příkladě. Je také možné soustředit se pouze na ta data a, která jsou v čase změněná oproti předchozímu otisku-s.

Při konstrukcí nového upraveného otisku v, lze tyto úseky a vložit do konstruované datové řady v, například následujícím způsobem: nově vkládaná data se nejprve upraví například podle varianty popsané v prvním příkladě. Ke každému krátkému datovému úseku souboru takto vzniklých lokálních datových informací-úseků, nebo k celému novému datovému úseku pokud se předešlá úprava neprovádí, se přidají na jejich začátek a /nebo konec opakující se datové podúseky ze skupin opakujících se podúseků již obsažených v konstruované upravené datové řadě v. Takto upravené úseky se vloží náhodně do konstruované datové řady v.

Prvky-uživatelé zpřístupňují upravenou datovou řadu v, kódovanou jako řada znaků, na volně přístupném místě, například na své webové stránce, či na místě s přístupem omezeným například jen pro některé další prvky-uživatele. Tato sdílená upravená datová řada slouží ke studiu systému i jednotlivých prvků zkoumáním údajů týkajících se lokální datové informace v otiscích (například výskyt specifických krátkých datových úseků, relativní četnost stejných datových úseků nebo jejich proměna v závislosti na čase). Toto studium může sloužit například ke kontrole sys-

tému či prvků-operačních systémů z hlediska bezpečnosti, korektnosti, důvěryhodnosti či statistické analýzy např. často používaných instrukcí či datových bloků. Toho lze využít např. při navázání bezpečné komunikace či spolupráce mezi prvky, při jejich migraci či replikaci na jiné výpočetní zařízení či procesor čí při jejich kopírování na jiná zařízení či média za účelem úschovy. Lze toho rovněž využít při konstrukci a studiu celého systému například za účelem poznání jeho dynamického vývoje a rozpoznání a izolace útoku či nákazy v systému. Příkladem systémů, které tak fungují, jsou biologické systémy.

Příklad 3

Pro vložení nových genetických datových informací, například fluorescentních fúzních proteinových produktů, do buněk, jsou používány geneticky vytvořené bakteriální plasmidy a virové vektory. Úpravou genomu pomocí plasmidů lze vytvořit transformované buněčné linie.

Pro vytváření specifických proteinů lze využít specifický plasmid zkonstruovaný ze základní struktury plasmidu pUC. Do tohoto plasmidu se vloží promotér obratlovců vzatý z lidského cytomegaloviru CMV a genetická datová informace, kterou chceme vkládat.

Bylo pozorováno, že vytvořený plasmid a přepis a exprese příslušného proteinu z naší genetické datové informace v buňce, obr. 1, jsou stabilnější, jestliže promotéru předchází krátký opakující se podúsek póly C, za ním následuje promotér, potom opět podúsek póly C, a potom genetická datová informace (například gen) kterou chceme vložit do nově konstruovaného plasmidu. Nakonec na jeho konci následuje další podúsek póly C. Bylo pozorováno, že tento způsob vkládání opakujících se podúseků vede k větší stabilitě příslušného plasmidu a přepisu a exprese příslušného proteinu. Chceme-li vytvářet příslušný protein v eukaryotech, bylo experimentálně pozorováno, že je výhodnější nakonec použít podúsek póly A místo podúseku póly C.

PATENTOVÉ NÁROKY

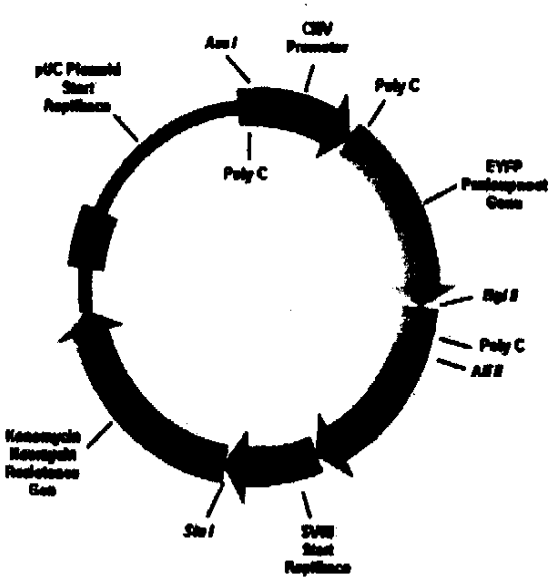
1. Způsob úpravy datových informací, reprezentovaných elektrickými nebo vlnovými signály nebo datových informací v přírodních a umělých databázích a záznamových médích, za účelem maskování a ochrany obsahu citlivých datových informací při současném zachování jejich vybrané části, sledované lokální datové informace, v systému obsahujícím alespoň jednu datovou informaci, kde původní úplná datová informace je kódovaná jako řada znaků vybraných ze skupiny tvořené písmeny, číslicemi nebo libovolnými symboly, **vyznačující se tím**, že se celá původní datová informace rozdělí ve fyzickém médiu, zejména hardwaru počítače, fyzickém komunikačním kanálu, fyzickém záznamovém médiu či biologickém materiálu, na krátké překrývající se datové úseky, přičemž délka těchto úseků je minimálně stejná nebo delší než je délka sledované sbírané lokální datové informace, a sledovaná lokální datová informace je v krátkých úsecích obsažena celá, tyto rozdělené krátké úseky vytvoří první skupinu a alespoň k jednomu krátkému úseku první skupiny se přidají data kódovaná jako vybrané znaky před nebo za znaky krátkých úseků první skupiny, přičemž vybrané znaky a jejich sled je stejný nebo různý a získaná směs původních krátkých úseků a modifikovaných úseků tvoří druhou skupinu a úseky druhé skupiny se vzájemně spojí v řadu dat.

2. Úprava datové informace podle nároku 1 **vyznačující se tím**, že se spojené úseky druhé skupiny znovu zpracují podle nároku 1, přičemž rozdělení na krátké úseky a přidávání vybraných znaků se může provádět stejně nebo jinak a vzniklá další nová druhá skupina se opět spojí v řadu dat a tato procedura se několikrát opakuje za účelem rozšíření opakujících se podúseků.

3. Použití způsobu úpravy datových informací podle nároků 1 a 2 ke konstrukci systému prvků, které sdílí lokální informace a jejich datových řad za účelem získání nových požadovaných vlastností systému.

1 výkres

Obr.1



Konec dokumentu