SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE

FAKULTA INFORMATIKY A INFORMAČNÝCH TECHNOLÓGIÍ

VEDECKÁ RADA FAKULTY INFORMATIKY A INFORMAČNÝCH TECHNOLÓGIÍ STU

Ing. Radovan Semančík

Autoreferát dizertačnej práce

Revidovaná architektúra World Wide Web-u

(Revised World Wide Web Architecture)

na získanie vedecko-akademickej hodnosti philosophiae doctor

vo vednom odbore:

25-21-9 Počítačové prostriedky a systémy

Bratislava, december 2008

Dizertačná práca bola vypracovaná v externej forme doktorandského štúdia o po uplynutí času určeného na štúdium v súlade s § 12 ods. 1 písm. b) Vyhlášky č. 131/1997 Z.z. Na Fakulte informatiky a informačných technológií Slovenskej technickej univerzity v Bratislave.

Predkladateľ: Ing. Radovan Semančík
Ústav počítačových systémov a sietí
FIIT Slovenská technická univerzita v Bratislave
Ilkovičova 3, 842 16 Bratislava

Školiteľ: doc. Ing. Margaréta Kotočová, PhD.
Ústav počítačových systémov a sietí
FIIT Slovenská technická univerzita v Bratislave
Ilkovičova 3, 842 16 Bratislava

Oponenti: prof. Ing. Pavol Horváth, PhD.
Centrum výpočtovej techniky, STU, Bratislava

prof. Ing. Jiří Šafařík, PhD.
Fakulta aplikovaných věd, ZČU, Plzeň

Ing. Martin Chovanec, PhD.
Ústav výpočtovej techniky, TU, Košice

Autoreferát bol rozoslaný dňa 20.4.2009.

Obhajoba dizertačnej práce sa koná: 20.5.2009 o 11:00 pred komisiou pre obhajobu dizertačnej práce vymenovanou predsedom spoločnej odborovej komisie prof. Ing. Liberios Vokorokos, PhD. vo vednom odbore Počítačové prostriedky a systémy 25-21-9 na Fakulte informatiky a informačných technológií Slovenskej technickej univerzity, Ilkovičova 3, 842 16 Bratislava v miestnosti D220.

Predseda spoločnej odborovej komisie:

prof. Ing. Liberios Vokorokos, PhD.
Fakulta elektrotechniky a informatiky
Technická univerzita v Košiciach
Letná 9, 042 00 Košice

Obsah

Úvod

Interakcie s počítačovými systémami sú dôležitou súčasťou našich životov. Počítačové systémy často ukladajú a spracúvajú údaje, ktoré popisujú fyzické osoby. Napriek tomu veľké množstvo nereflektuje vo svojom návrhu osobité črty takýchto informácií.

Dnešný World Wide Web (WWW) je postavený na predpoklade, že väčšina informácií, ktoré sa pomocou neho šíria, sú verejne dostupné. Taktiež sa pri návrhu WWW očakávalo, že len relatívne málo ľudí bude publikovať informácie a väčšina ich bude čítať. V súčasnosti ale tieto predpoklady pomaly prestávajú platiť. Veľa aplikácií postavených na mechanizmoch WWW robia Web *zapisovateľným*. Wiki, blogy, aplikácie na zdieľanie fotografií a videí umožňujú návštevníkom meniť uloženú informáciu, komentovať ju a podobne. Pôvodná architektúra WWW nepredvídala, že veľká časť informácií prístupných cez WWW bude dynamických a že prístup k nim bude neverejný alebo len poloverejný.

Táto práca bola motivovaná snahou navrhnúť architektonickú vrstvu na podporu správy údajov o používateľoch architektonicky umiestnenou nad mechanizmami WWW. Tento prístup sa však zmenil ako dôsledok analýzy súčasného stavu architektúry WWW. Analýza odhalila množstvo architektonických nedostatkov v základnom návrhu princípov WWW. Ukázalo sa, že základné koncepty pre podporu práce s údajmi o používateľoch musia byť zahrnuté do architektúry WWW. Výsledná práca popisuje motiváciu pre tento prístup, identifikuje architektonické problémy súčasnej architektúry WWW a na konceptuálnej úrovni navrhuje jej zlepšenie.

1 Súčasný stav

Autentifikácia pomocou hesiel je neadekvátna pre vysoko distribuované systémy [1]. Použitie jednorazových hesiel [2] [3] alebo systémov založených na výzve a odpovedi (challenge-response) [4] [5] mierne zlepšujú bezpečnostné charakteristiky, v globálnom nasadení však neprinášajú podstatné zlepšenie. Riešenia založené na infraštruktúre verejných kľúčov (PKI) [6] [7] sú nasaditeľné aj v prostredí distribuovaných systémov, problémom však zostáva overenie autenticity kľúča. Tento problém sa prejavuje aj v protokole HTTPS [8], najčastejšie používanej aplikácií PKI princípov v prostredí WWW. Tento protokol dokáže overiť len meno servera, ktorý poskytuje informácie na základe jeho kvalifikovaného doménového mena (FQDN), neposkytuje však žiadne informácie o autenticite informácie ktorá sa pomocou neho prenáša. Použitie PKI na identifikáciu ľudí sa v praxi neosvedčilo, o čom svedčí aj nízka adopcia digitálnych certifikátov v Európskych krajinách [9]. Systémy založená na princípoch PKI sú všeobecne nevhodné na vyjadrovanie dynamických informácií alebo informácií, ktoré môžu ohroziť súkromie používateľov (ako napríklad údaje o používateľoch) [10]. Zlepšením by mohol byť kryptografický systém nazvaný „Digital Credentials“ [11]. Aj keď základné kryptografické mechanizmy tohto systému sú dokumentované, technické detaily

navrhovaného riešenia (ako napríklad protokoly) sú neznáme, preto nie je možné jeho prínos objektívne zhodnotiť.

Systémy Internetovej identity [12] [13] [14] [15] [16] [17] [18] [19] [20] [21] sa pokúšajú presunúť údržbu používateľských prístupových účtov na dedikované systémy a tým znížiť počet účtov ktoré musia používatelia Internetu udržiavať. Zdrojový systém udržiava používateľské účty. Pri prístupe používateľa k službe na cieľovom systéme zdrojový systém oznamuje na cieľový systém, že používateľ je autentifikovaný a v prípade potreby aj jeho atribúty. Systémy Internetovej identity sú systémy jednotného prihlasovania založené na „proxy“ mechanizmoch [22]. Systémy Internetovej identity trpia množstvom nedostatkov. Zdrojový systém sa vie triviálnym spôsobom vydávať za ktoréhokoľvek z jeho používateľov [22]. Zdrojový systém vie sledovať prihlasovanie sa používateľa na cieľových systémoch. Globálne identifikátory používané niektorými systémami môže byť použité na koreláciu aktivít používateľa na niekoľkých cieľových systémoch. Niektoré systémy sú náchylné na takzvané „phishing“ útoky [23]. Táto náchylnosť je často podporovaná nesprávnymi architektonickými rozhodnutiami pri návrhu systémov [24]. Systémy Internetovej identity vyžadujú buď priamy vzťah dôvery medzi zdrojovým a cieľovým systémom, alebo predpokladajú, že tento vzťah je implikovaný základnými štruktúrami Internetu (napr. DNS systémom). Priamy vzťah má však silne obmedzenú škálovateľnosť a implikovaný vzťah sa nedá považovať za dostatočne dôveryhodný.

Niekoľko autorov [25] [26] rozoberá konceptuálne základy identity a anonymity v súvislosti s informačnými technológiami. Zvlášť pozoruhodná je práca [27] popisujúca spoločnú terminológiu v tejto oblasti. Literatúra taktiež popisuje tradičný koncept súkromia [28], ktorý je však nepostačujúci pre dnešnú informačnú dobu. Problémy tradičného vnímania súkromia výborne popisuje Solove [29]. Navrhuje architektúru na ochranu súkromia založenú na dokumente „Fair Information Practices“ [30], ktorý bol jedným zo základov smerníc OECD v tejto oblasti a následne aj smerníc EU pre ochranu osobných údajov [31]. Resnick et al. [32] popisujú ad-hoc interakcie v prostredí Internetu vykonávané medzi účastníkmi bez predchádzajúceho dlhodobého vzťahu. Navrhujú ako riešenie problémov dôvery je navrhovaný distribuovaný reputačný systém. Pozitívne vplyvy podobného mechanizmu popisuje aj Axelrod [33]. Windley et al. [34] pozorujú, že reputácia je v prirodzenom rozpore so súkromím. Sabater a Sierra [35] poskytujú klasifikovaný prehľad množstva reputačných systémov. Friedman and Resnick [36] poskytujú zhodnotenie reputačného mechanizmu založeného na teórií hier.

Architektúra WWW bola v čase jeho vzniku založená len na krátkom dokumente [37]. Princípy a mechanizmy súčasného WWW vznikli v deväťdesiatych rokoch a boli čiastočne dokumentované vo forme špecifikácií URI [38] a HTTP 1.1 [39]. Architektúra bola v tom období vedená architektonickým štýlom REST [40], ktorý bol však popísaný až v roku 2000. Počas vývoja WWW až do roku 2004 neexistoval žiaden ucelený dokument popisujúca architektúru WWW. Takýto dokument [41] bol

vytvorený až retrospektívne. Aj napriek tomu, že architektonický štýl REST bol použitý na riadenie vývoja WWW, sám autor priznáva [40], že architektúra WWW je len v hrubých rysoch zosúladená s týmto štýlom a že architektonické inkonzistencie stále pretrvávajú.

Súčasné WWW prehliadače nezobrazujú žiadnu informáciu o dôveryhodnosti prezeraného obsahu. Jediné informácie ktoré by mohli byť spadať do tejto kategórie sú informácie protokolu HTTPS. Avšak tento protokol sa nezaoberá dôveryhodnosťou poskytovanej informácie, ale len overením mena servera z ktorého informácia pochádza. Tento nedostatok vedie k mohutnému rozšíreniu takzvaných „phishing“ útokov [42].

Internet je celosvetové komunikačné médium spájajúce asi 1,5 miliardy používateľov [43]. Je nerealistické očakávať, že podstatná časť používateľov Internetu bude udržiavať medzi sebou dlhodobé vzťahy s každým človekom, s ktorým prídu na Internete do styku. Viac pravdepodobné je, že veľká časť vzťahov bude krátkodobých a komunikácia bude veľmi často prebiehať medzi partnermi, ktorí sa priamo nepoznajú. V takom prípade bude potrebná metóda, ktoré pomôže používateľom vyhodnotiť dôveryhodnosť osoby s ktorou komunikujú. Takýto mechanizmus však momentálne nie je k dispozícii.

2 Ciele práce

Cieľom práce je zlepšenie architektúry World Wide Web-u, zamerané najmä na konceptuálne vrstvy architektúry. Práca sleduje nasledujúce ciele:

- Definovať očakávania a ciele, ktoré má návrh architektúry World Wide Web-u naplniť tak, aby zodpovedali potrebám súčasnosti a predpovedateľnej budúcnosti.
- Vyhodnotiť stav, konzistenciu a vhodnosť súčasnej architektúry World Wide Web-u vzhľadom na ciele špecifikované v predchádzajúcom kroku.
- Identifikovať a rozobrať principiálne problémy súčasnej WWW architektúry, zamerané najmä na prácu s údajmi o používateľoch a na vyhodnocovanie dôveryhodnosti informácií dostupných pomocou WWW.
- Navrhnúť zlepšenia WWW architektúry zamerané najmä na konceptuálnu úroveň.
- Overiť navrhnutú architektúru predvedením, že nová architektúra dokáže zvládnuť situácie, ktorých realizácia je problematická v súčasnom stave WWW.

Jeden z dôležitých implicitných cieľov práce je udržať snahu o zlepšenie architektúry WWW na realizovateľnej úrovni. Ucelený a overiteľný výsledok obmedzený len na konceptuálnu úroveň bol uprednostnený pred potencionálne nekoncepčným riešením len jedného aspektu architektúry. Revidovanie a zlepšenie architektúry WWW je nesporne náročná úloha. Preto bola táto úloha rozdelená na

menšie, zvládnuteľné kroky. Táto práca sa zameriava na prvý krok: zlepšenie architektúry WWW a odstránenie základných chýb na konceptuálnej úrovni. Tomuto zámeru zodpovedá aj definícia cieľov práce.

3 Model

Prakticky použiteľná architektúra sa nedá spoľahlivo navrhnúť bez predchádzajúceho porozumenia základným princípom prostredia v ktorom má operovať. Na tento účel bol vytvorený model, ktorý sa snaží popisovať interakcie dvoch svetov: sveta ľudských bytostí (nazývaného *reálny priestor*) a sveta počítačov (nazývaného *kyberpriestor*). Model rozoberá spôsob, ktorým ľudia komunikujú s počítačmi a naopak, pričom sa najmä zameriava na spoľahlivosť a dôveryhodnosť informácií. Model sa tiež zaoberá anonymitou a identitou používateľov.

Interakcia medzi reálnym priestorom a kyberpriestorom je možná len pomocou *terminálových zariadení*. Tieto zariadenia sú entitami, ktoré participujú v oboch svetoch a preto dokážu transformovať údaje medzi nimi. Počítačový monitor alebo klávesnica sú príkladmi terminálových zariadení. Okrem interakcie cez terminálové zariadenia žiadna iná priama interakcia medzi priestormi nie je možná. Ani entity z reálneho priestoru ani entity kyberpriestoru nevedia overiť, či terminálové zariadenie pracuje podľa očakávania, pretože nemôžu priamo pozorovať cudzí priestor. Je možné korelovať informácie z niekoľkých terminálových zariadení a tak testovať ich funkčnosť. Tento prístup môže posilniť vieru v správnu činnosť zariadenia, neumožňuje však správnu činnosť zariadenia dokázať. Na základe týchto argumentov môžeme sformulovať nasledujúce tvrdenie:

> Prechod cez hranicu reálneho priestoru a kyberpriestoru je vždy subjektívny.

Entita, ktorá prijíma údaje z cudzieho priestoru cez terminálové zariadenie, sa musí sama rozhodnúť, nakoľko sú poskytnuté údaje relevantné a spoľahlivé, keďže nie je možné tieto informácie priamo overiť. Vnímanie týchto informácií závisí na ich interpretácií, predpojatosti a vnímavosti entít, ich viere alebo prednastavenom správaní (programovaní). Preto považujeme tieto informácie za subjektívne. Keďže všetky informácie, ktoré sa nachádzajú v kyberpriestore majú svoj (priamy alebo nepriamy) zdroj v reálnom priestore, môžeme formulovať nasledujúce tvrdenie:

> Akákoľvek informácia prichádzajúca z kyberpriestoru je subjektívna.

Dôveryhodnosť zdroja je dôležitým faktorom pre vyhodnotenie spoľahlivosti informácie, ktorá z neho pochádza. Dôveryhodnosť subjektívnej informácie nie je možné spoľahlivo vyhodnotiť, ak nie je známy jej zdroj. Preto môžeme formulovať nasledujúce tvrdenie:

> Zdroj informácie v kyberpriestore je rovnako dôležitý ako samotný obsah informácie.

Nie je nutné, aby bola známa plná reálna identita zdroja informácie. Vhodné množstvo informácií o zdroji nutné na vyhodnotenie dôveryhodnosti informácie je postačujúce a tieto informácie o zdroji by mali byť (priamo či nepriamo) prenášané spolu s primárnou informáciou.

Osoby (entity z reálneho priestoru), ktoré používajú počítačové systémy sú v kyberpriestore reprezentované pomocou dátových štruktúr. Tieto štruktúry sú zložené zo subjektívnych informácií a sú len nekompletnou reprezentáciou reálnych osôb. V tejto práci budeme používať pojem *subjekt* (subject) na popis osôb a podobných entít z reálneho priestoru a pojem *persóna* (persona) pre ich kyberpriestorové reprezentácie:

> **Subjekt** je vedomá entita z reálneho priestoru vedená slobodnou vôľou.
>
> **Persóna** je kyberpriestorová dátová štruktúra, ktorá reprezentuje niektoré aspekty subjektu alebo kyberpriestorovej entity subjektom ovládanej. Tieto aspekty sú reprezentované vo forme súboru charakteristík so strojovo-čitateľnými hodnotami.

Keďže persóny sú kyberpriestorové dátové štruktúry, z predchádzajúceho textu vyplýva že persóny môžu byť len subjektívnou reprezentáciou reálnych osôb. Persóna obvykle vzniká ako dátová štruktúra popisujúca konkrétnu osobu. Po vytvorení záznamu však spojitosť medzi osobou a persónou nemusí byť zrejmá. Preto je potrebný mechanizmus na vyhodnocovanie spojitosti medzi osobami a persónami.

Definujeme koncept svetovej množiny (world set) ako množiny všetkých subjektov ktoré by mohli potencionálne prichádzať do úvahy ako zdroje informácií nachádzajúcich sa v persónach:

$$W = \{S_1, S_2, S_3, \ldots, S_n\}$$

Základný koncept identity definujeme pomocou pravdepodobnostných mechanizmov:

> **Pravdepodobnosť identity** je (Bayesovská) pravdepodobnosť, že daná persóna popisuje daný subjekt. Je zapisovaná ako
>
> $i_{P,S}$
>
> kde P je persóna a S je subjekt.

Pravdepodobnosť identity, podobne ako ostatné pravdepodobnostné hodnoty v tejto práci, je subjektívna vzhľadom na konkrétneho pozorovateľa. Keďže persóna popisuje práve jeden subjekt, súčet pravdepodobností identity konkrétnej persóny pre všetky subjekty v svetovej množine sa musí rovnať jednej:

$$\sum_{k=1}^{n} i_{P,S_k} = 1$$

Môžeme definovať náhodnú premennú I_P, kde svetová množina W reprezentuje možné stavy a pravdepodobnosti týchto stavov sú zhodné s pravdepodobnosťou identity pre konkrétnu persónu:

$$P(I_P = S) = i_{P,S}\ ,\quad \forall S \in W$$

Náhodná premenná I_P reprezentuje subjekty, ktoré persóna P môže popisovať.

Môžeme definovať množinu anonymity (anonymity set) ako pravdepodobnostnú variáciu definície poskytnutej v [26]:

> Pre konkrétnu persónu, **množina anonymity** (zapisovaná AS_P) je množina všetkých subjektov, pre ktoré platí, že pravdepodobnosti identity pre danú persónu a subjekt je väčšia ako nula:
>
> $$AS_P = \{S:\ S \in W\ \wedge\ i_{P,S} > 0\}$$

Ďalej definujeme mieru anonymity (anonymity ratio), metriku užitočnú pre vyhodnocovanie anonymity:

> **Miera anonymity** persóny vzhľadom na svetovú množinu a pozorovateľa je relatívna neistota korešpondencie persóny a subjektu, definovaná ako
>
> $$ar(P) = \frac{H(I_P)}{H_{max}}\ ,$$
>
> kde $H(I_P)$ je entropia [44] náhodnej premennej I_P, definovaná ako
>
> $$H(I_P) = -\sum_{k=1}^{n} i_{P,S_k} \log_2(i_{P,S_k})\ ,$$
>
> a H_{max} je maximálna entropia náhodnej premennej s n stavmi
>
> $$H_{max} = \log_2(n)\ .$$

Miera anonymity rovná jednej znamená úplnú anonymitu: pozorovateľ nevie nijako rozlíšiť, ktorý subjekt zo svetovej množiny by mohol byť zdrojom pre persónu. Miera anonymity rovná nule znamená žiadnu anonymitu: pozorovateľ si je istý, ktorý subjekt je zdrojom pre persónu. V praxi môže byť zaujímavé vyhodnocovať identity namiesto anonymity, preto definujeme aj opačný koncept:

> **Miera identity** persóny vzhľadom na svetovú množinu je pravdepodobnostný opak miery anonymity. Je definovaná ako:
>
> $$ir(P) = 1 - ar(P)\ .$$

Presné hodnoty miery anonymity a identity môže byť veľmi zložité zistiť v praxi, ak je to vôbec realizovateľné. Preto definujeme koncepty podobné identite a

anonymite ale použiteľné v kyberpriestorových informačných systémoch: analógia a heterológia persón.

Analogické persóny popisujú rovnaký subjekt. Identita takého subjektu nemusí byť známa pre vyhodnotenie analógie.

Persóny sú **analogické** práve vtedy keď popisujú ten istý subjekt.

Takáto deterministická definícia analógie je však problematická. Určiť analógiu s úplnou istotou môže byť nemožné. Preto obdobne ako v prípade identity a anonymity, definujeme pravdepodobnostnú verziu analógie:

Pravdepodobnosť analógie je (Bayesovská) pravdepodobnosť, že persóny sú analogické. Je zapisovaná ako

$$l_{P_1, P_2} \quad ,$$

kde P_1 a P_2 sú persóny.

Pre úplnosť definujeme aj opak analógie:

Persóny sú **heterologické** práve vtedy, keď popisujú rôzne subjekty.

Pravdepodobnosť heterológie je (Bayesovská) pravdepodobnosť, že persóny sú heterologické. Je zapisovaná ako

$$h_{P_1, P_2} \quad ,$$

kde P_1 a P_2 sú persóny.

Persóny môžu buď popisovať ten istý subjekt alebo rôzne subjekty. Z toho vyplýva, že

$$l_{P_1, P_2} + h_{P_1, P_2} = 1 \quad .$$

Keďže koncepty analógie a heterológie môžu byť vyhodnotené z informácií, ktoré sú k dispozícii v kyberpriestore, za istých okolností môžu byť použité ako praktické aproximácie identity a anonymity v kyberpriestorových informačných systémoch.

4 Ciele a metódy návrhu

Správna architektúra musí zohľadňovať požiadavky a zvyky ľudí, ktorým má slúžiť. Našim cieľom je navrhnúť architektúru, ktorá bude podporovať prostredie efektívnej spolupráce, ktoré bude vzbudzovať pozitívny sieťový efekt [45]. Také prostredie by malo podporovať spoluprácu medzi akýmikoľvek entitami v sieti. Prostredie by nemalo by byť obmedzené len na spoluprácu pomocou distribučných kanálov, kde niekoľko silných entít sprostredkováva väčšinu komunikácie v sieti. Sú rozobraté tri potencionálne prostredia: anarchické, autoritatívne a prostredie zodpovednosti. Anarchické prostredia sú prekážkou pre spoluprácu. V neriadenom prostredí niekoľko silných entít bude využívať väčšinu ostatných, stále ich oslabujúc. Spolupráca v takom prostredí je neefektívna, pretože nikomu nie je možné

dôverovať. Autoritatívne prostredia môžu byť efektívne v malom, je však zložité ich škálovať bez rizika neefektivity vyplývajúcej z byrokracie. Preto sú autoritatívne prostredia nevhodné pre Internet. V prostredí zodpovednosti majú ľudia slobodu na prácu a najmä spoluprácu, ľudia však musia prebrať zodpovednosť za svoje činy. Kľúč je v rovnováhe medzi protichodnými silami a v mechanizme ktorý dokáže udržiavať takú rovnováhu. Našim cieľom je architektúra ktorá dokáže podporovať takého prostredie zodpovednosti.

5 Architektúra Internetu

Vývoj Internetu je aktivita veľkej a rozmanitej skupiny ľudí. Vývoj protokolov a aplikácií nie je riadený centralizovanou autoritou. Niekoľko organizácií ovplyvňuje vývoj Internetu, žiadna z nich však vývoj úplne neriadi. Vývoj Internetu a príbuzných technológií je preto silne evolučný. Žiaden dokument nepopisuje architektonické princípy Internetu, pretože žiadne neexistujú [46]. Keďže architektúra Internetu nie je silne koordinovaná, architektonické inkonzistencie sú nevyhnutné.

World Wide Web vznikol začiatkom deväťdesiatych rokov ako distribuovaný hypertextový systém. Neskôr sa vyvinul na hypermediálny systém a dnes je považovaný za globálny „informačný priestor“ [41]. Architektúra WWW bola vedená architektonickým štýlom REST [40]. Tento štýl sa zakladá na klient-server princípoch, kde komunikácia musí byť inicializovaná klientom. Súčasné požiadavky však vyžadujú aj komunikáciu iniciovanú serverom, čo je v praxi riešené nesystematickými mechanizmami ako RSS [47] alebo neefektívnym čakaním na udalosť v prípade AJAX aplikácií [48]. Architektonický štýl vyžaduje, aby server bol bezstavový [40]. Toto obmedzenie môže platiť, len ak sú WWW prostriedky (*resource*) statické. REST však umožňuje meniť stav prostriedkov, čo odporuje požiadavke na bezstavovosť. Súčasné požiadavky na „zapisovateľný Web“ však jasne odhaľujú tento architektonický nedostatok. REST obsahuje obmedzenie na jednotné rozhranie, ktoré by zjavne malo byť definované špecifikáciami URI [38], HTTP [39] a HTML [49]. Tieto špecifikácie však umožňujú veľmi veľkú mieru rozšíriteľnosti a sú silne zamerané na syntax a len minimálne na sémantiku rozhraní. Takýto prístup obmedzuje interoperabilitu a môže byť len ťažko považovaný za silnú definíciu rozhrania.

Koncepty prostriedku (resource) a jednotného identifikátora prostriedku (uniform resource identifier, URI) sú základnými prvkami architektúry WWW. Ale aj napriek tomu je k dispozícií len veľmi vágna definícia prostriedku [41]. Je zrejmé, že aj objekt z reálneho priestoru môže byť prostriedkom. Prostriedky sú identifikované pomocou URL a vlastníci URI by mali poskytnúť reprezentáciu prostriedku [41]. Z toho vyplýva, že aj objekty z reálneho priestoru by mali mať reprezentáciu v kyberpriestore poskytnutú vlastníkom URI. Taká reprezentácie je však vždy subjektívna. Nie je žiadna záruka, že vlastník URI je aj vlastníkom objektu z

reálneho priestoru, ktorý URI identifikuje. Preto je možné, že takáto reprezentácia nebude správna alebo bude dokonca škodlivá. Tento problém bol identifikovaný [50] skupinou technickej architektúry World Wide Web konzorcia (W3C TAG). Bolo navrhnuté riešenie [51], ktoré nedovoľuje poskytnúť reprezentáciu prostriedku, ktorý nie je „informačným prostriedkom“. Aj keď W3C TAG deklaruje, že architektúra zostala konzistentná, niektoré problémy stále pretrvávajú. Najmarkantnejším problémom je zavedenie závislosti WWW architektúry a konceptu URI na protokole HTTP, pričom iný dokument [38] vyžaduje nezávislosť konceptu URI od protokolov.

Vágna definícia prostriedku väčšinou nespôsobuje problémy, ak je príjemcom informácie inteligentná osoba. Problém s interpretáciou nastáva pri automatickom spracovaní informácie, napríklad pri zavedení reputačných systémov. Napríklad v prípade, ak používateľ vyjadril negatívny názor o prostriedku, ktorého reprezentácia je digitálna fotografia osoby. Automatický systém nedokáže rozlíšiť, či negatívny názor sa týka kvality kódovacieho algoritmu, schopností fotografa zachytiť model, vzhľadu osoby na fotografií alebo jej duševných vlastností. Doporučenie vyhýbať sa URI aliasom robí celú situáciu zložitejšou, najmä keď W3C TAG za istých okolností doporučuje [52] [53] používať rôzne URI v situáciach, ktoré by sa dali považovať za rôzne reprezentácie rovnakého prostriedku. Ďalšia nekonzistencia sa týka menných priestorov jazyka XML [54], ktorý využívajú URI. Plné mená objektov v menných priestoroch však používajú takzvané kvalifikované meno (QName), ktoré už nie je URI, čo je v rozpore s princípmi architektúry WWW. Aj keď architektúra WWW [41] vyžaduje mapovanie medzi kvalifikovanými menami a URI, neposkytuje žiaden mechanizmus a ani odporúčanie pre také mapovanie.

W3C TAG tvrdí, že URI identifikátory využívajúce schému *http* podporujú perzistenciu identifikátorov tak, ako je to len v prakticky možné [53]. Perzistencia takýchto identifikátorov však závisí od perzistencie pridelenia DNS mena, ktoré sa v nich nachádza. Pridelenie DNS mena je vhodné pre strednodobú perzistenciu v prípade etablovaných organizácií. Avšak je veľmi náročné a nepohodlné pre individuálnych používateľov získať vlastné DNS meno. Navyše W3C doporučuje používať čitateľné mená v URI [55] [56], čo tiež obmedzuje perzistenciu. Preto je otázka perzistencie *http* URI identifikátorov sporná. Takéto URI by sa mali považovať za adresy prostriedkov, nie za ich identifikátory.

Súčasná architektúra WWW predpokladá, že informácia vždy prichádza priamo z autoritatívneho zdroja. Protokol HTTPS využíva tento predpoklad na zabezpečenie prenosu informácie cez sieť. Tento prístup však zlyháva v prípadoch, keď je informácie masívne a dynamicky distribuovaná alebo replikovaná na rôznych miestach.

Sémantický web [57] je navrhovaný koncept, ktorý stava na princípoch WWW. Primárnym cieľom sémantického webu nie je distribúcia a linkovanie dokumentov pre priame potreby ľudí. Cieľom sú štruktúrované informácie, ktoré môžu byť automaticky spracovávané. Je predpoklad, že objekty budú reprezentované pomocou

jazykov založených na XML, ako napríklad RDF [58]. Oponenti [59] sémantického web-u popisujú množstvo prekážok, ktoré bude potrebné prekonať pri jeho nasadení. Väčšina prekážok sa týka nespoľahlivosti takto dostupných informácií a problematického určenia ich dôveryhodnosti.

6 Navrhovaná architektúra

Navrhovaná architektúra je založená na novom architektonickom štýle RRSS, ktorý je silne inšpirovaný štýlom REST [40]. Štýl RRSS kombinovaný s ďalšími štýlmi a ohraničeniami vytvára základ pre navrhovanú architektúru WWW.

6.1 Architektonický štýl RRSS

Architektonický štýl RRSS definuje základné ohraničenia štruktúrovania a reprezentácie informácií v kyberpriestore. Štýl berie do úvahy zdroj týchto informácií ako ich neoddeliteľnú súčasť. RRSS je konceptuálny štýl ktorý definuje štyri základné prvky: prostriedok (resource), reprezentácia (representation), zdroj (source) a sémantika (semantics).

Prostriedok (resource) je definovaný nepriamo, pomocou jeho vlastností:

> **Prostriedok** je kyberpriestorová entita.
>
> **Prostriedok** môže reprezentovať entity z reálneho priestoru ako aj entity z kyberpriestoru.
>
> Prostriedok je **úplná, samostačná a konzistentná** reprezentácia objektu alebo konceptu.
>
> Stav prostriedku je **dynamický a nestály**. Nie je možné predpokladať, že stav prostriedku nasleduje akýkoľvek stavový model, ak to nie je explicitne špecifikované dodatočnou informáciou.

Prostriedok sa prezentuje ostatným entitám vo forme reprezentácií (resource representation):

> Prostriedok môže byť reprezentovaný v kyberpriestore ľubovoľným počtom **reprezentácií prostriedku**.

Veľké množstvo reprezentácií a ich formátov môže negatívne ovplyvniť interoperabilitu. Preto doporučujeme množinu povinných reprezentácií:

> Malý počet **dobre definovaných a stabilných formátov reprezentácií prostriedku** by mal byť štandardizovaný a povinný. Prostriedok by mal poskytnúť aspoň jednu reprezentáciu v štandardnom formáte.

Informácie, ktoré tvoria prostriedok, prechádzajú cez hranicu reálneho priestoru a kyberpriestoru. Preto považujeme prostriedok za subjektívny:

> Prostriedok je vždy **subjektívny**.

Pri subjektívnej informácií je zdroj informácie dôležitý:

Identifikácia **zdroja prostriedku** je integrálna súčasť prostriedku.

Podľa modelu predstaveného v kapitole 3, zdroj prostriedku persóna.

Sémantický popis prostriedku je potrebný na upresnenie významu prostriedku:

> Prostriedok môže byť sémanticky popísaný pomocou štandardizovaného, počítačovo-čitateľného formátu. **Sémantický popis** je povinný pre všetky prostriedky a musí byť súčasťou akéhokoľvek rozhrania ktoré je použité na prístup k prostriedku.

6.2 Identifikácia prostriedkov

Architektonický štýl RRSS cielene neobsahuje žiadne ohraničenie ktoré sa týkajú identifikácie prostriedkov. Identifikácia prostriedkov je definovaná ako samostatná sada ohraničení:

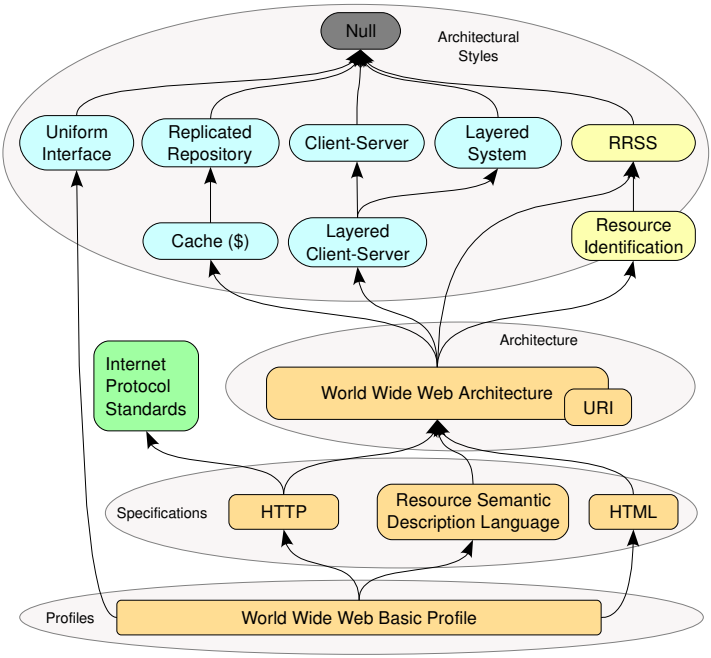
> Každý prostriedok má pridelený **identifikátor**, ktorý jednoznačne a konzistentne identifikuje prostriedok v systéme.
>
> Identifikátor prostriedku musí identifikovať najviac jeden prostriedok.
>
> Identifikátor prostriedku ktorý bol pridelený jednému prostriedku už nesmie byť pridelený inému prostriedku.
>
> Schéma identifikátoru prostriedku nesmie závisieť na žiadnej konštrukcií alebo koncepte z implementácie identifikátora.
>
> Sémantika identifikátora nesmie byť viditeľná klientským aplikáciam.

6.3 Revidovaná architektúra World Wide Web-u

Architektúru WWW navrhujeme rozdeliť na niekoľko úrovní abstrakcie. Takéto rozdelenie uľahčí porozumenie výslednej architektúre a zlepší viditeľnosť architektonických konceptov. Jednotlivé vrstvy riešia rôzne požiadavky na dynamiku architektonický konceptov a interoperabilitu výsledných produktov. Navrhujeme štyri vrstvy abstrakcie:

- **Architektonické štýly** tvoria najabstraktnejšiu vrstvu architektúry. Táto vrstva definuje základné architektonické ohraničenia indukujúce želané vlastnosti v nasledujúcich vrstvách. Architektonický štýl RRSS, ohraničenia identifikácie prostriedkov spolu s ďalšími štýlmi a ohraničeniami popísanými v [40] sú súčasťou tejto vrstvy. Táto vrstva je všeobecná, nie je obmedzená len na WWW. Očakáva sa, že koncepty v tejto vrstve budú stabilné a budú sa len veľmi málo meniť.
- **Architektúra WWW** je vrstva obsahujúca architektonické ohraničenia, pravidlá a odporúčania, ktoré definujú základné princípy architektúry WWW. Princípy definované v tejto vrstve sú fundamentálne pre WWW a očakáva sa že sa budú platné dlhú dobu a budú sa meniť len pomaly. Táto vrstva je popísaná nižšie.

- **Špecifikácie protokolov** sú zlúčené v samostatnej vrstve, ktorá upresňuje a ďalej ohraničuje koncepty WWW architektúry. Obsahuje špecifikácie komunikačných protokolov, ale aj dátových formátov a ďalších rozhraní. Očakáva sa, že špecifikácie protokolov sa budú prispôsobovať novým požiadavkám a že niekoľko rôznych protokolov s rôznymi charakteristikami môže existovať pre ten istý účel. Vrstva protokolov je popísaná len okrajovo.
- **WWW Profily** sú umiestnené v najkonkrétnejšej vrstve architektúry. Táto vrstva presne definuje sadu protokolov, dátových formátov a rozhraní potrebných pre správnu činnosť WWW. Táto vrstva definuje mechanizmus pre interoperabilitu a predpokladá sa, že sa bude meniť veľmi často. WWW profily sú v tejto práci spomínané len okrajovo.



Obr. 1: Architektonický diagram navrhovanej architektúry

Architektonická vrstva WWW je kompozíciou architektonických štýlov z abstraktnej vrstvy architektonických štýlov s pridaním dodatočných architektonických ohraničení, ako je zobrazené na obrázku 1. Jedným z dôležitých

ohraničení je identifikácia prostriedku, pretože tvorí základ hypermediálneho princípu WWW:

> Prostriedky sú identifikované pomocou identifikátorov s pevnými syntaktickými pravidlami, nazývanými **jednotné identifikátory prostriedkov** (Uniform Resource Identifier, URI).

Úplná definícia syntaxe a sémantiky URI by mala byť súčasťou architektúry WWW. Existujúce nevyhovujúce URI schémy by mali byť vyradené z prevádzky alebo považovaná za adresy prostriedkov (Uniform Resource Locator, URL).

Žiadne ohraničenie neobmedzuje počet identifikátorov priradených k prostriedku, čo umožňuje tvorbu aliasov. Je však potrebný mechanizmus na zistenie ekvivalencie identifikátorov. Pre tento účel navrhujeme kanonický identifikátor:

> Každému prostriedku musí byť v každej chvíli priradený práve jeden kanonický **kanonický jednotný identifikátor prostriedku**.

Porovnaním kanonických identifikátorov prostriedku je možné zistiť ekvivalenciu identifikátorov.

Architektúra WWW definuje reprezentáciu prostriedku konkrétnejšie:

> **Reprezentácia prostriedku** je pole bajtov ktoré tvorí obsah reprezentácie spolu s meta-dátami, ktoré popisujú formát reprezentácie a ďalšie aspekty reprezentácie.

Ďalej definujeme prístupový protokol, ktorý je implikáciou architektonického štýlu klient-server a adresu reprezentácie, ktorá je adresovací mechanizmus pre prístupový protokol.

> Reprezentácia prostriedku môže byť získaná zo servera pomocou **prístupového protokolu.**
>
> Reprezentácia prostriedku je adresovaná pomocou **adresy reprezentácie prostriedku.**

Architektonický štýl RRSS vyžaduje sémantický popis prostriedku, preto definujeme sémantický popis ako povinnú reprezentáciu prostriedku:

> **Sémantický popis** prostriedku je povinnou reprezentáciou prostriedku.

Podľa modelu popísaného v kapitole 3 pokladáme zdroj prostriedku za persónu. Môžeme výhodne definovať:

> **Zdroj prostriedku** je prostriedok.
>
> **Identifikácia zdroja prostriedku** by mala byť povinný prvok v sémantickom popise prostriedku.

Globálny charakter URI identifikátorov by pri nesprávnom použití mohol ohroziť súkromie používateľov. Vhodný stupeň súkromia môže byť dosiahnutý používaním pseudonymov namiesto primárnych identifikátorov zdroja.

6.4 Overenie architektúry

Navrhovaná architektúra je popísaná najmä na konceptuálnej úrovni a je značne abstraktná. Bežné prístupy pre validáciu architektúry, ako napríklad prototypovanie, nie sú v tomto prípade efektívne použiteľné. Na overenie architektúry bolo preto použité overovanie pomocou scenárov. Pri hodnotení nedostatkov súčasnej WWW architektúry bolo uvedených niekoľko scenárov, na ktorých bola demonštrovaná zložitosť riešenie niektorých problémov v súčasnej WWW architektúre. Rovnaká sada scenárov bola použitá aj na overenie architektúry, popisujúc, ako je možné problémy riešiť pri nasadení navrhovanej architektúry. Každý scenár obsahuje zoznam dôležitých elementov navrhovanej architektúry využitých pri riešení scenára. Týmto prístupom sme overili, že navrhovaná architektúra neobsahuje žiadne zjavne redundantné časti.

7 Prínosy práce

Cieľom práce bolo zhodnotenie nedostatkov súčasnej WWW architektúry a navrhnutie zlepšení na konceptuálnej úrovni. Motiváciou bolo podporiť spoluprácu na Internete vytvorením vhodného prostredia. Práca sa zameriavala najmä na najdôležitejší element Internetu: používateľov. Skúmali sme základy interakcie reálneho priestoru a kyberpriestoru a poskytli model, ktorý popisuje tieto interakcie na vysokej úrovni abstrakcie. Model definuje pojem persóny ako kyberpriestorovej reprezentácie subjektov z reálneho priestoru. Popisuje koncepty ako anonymita a identita a poskytuje nové koncepty analógie a heterológie, použiteľné priamo v kyberpriestore.

Práca identifikuje množstvo nedostatkov a architektonických nekonzistencií súčasnej architektúry WWW. Existujúca WWW architektúra slabo zohľadňuje dynamiku informácií. Slabá definícia rozhraní WWW, nekonzistentná aplikácia základných princípov, vágna definícia významu prostriedkov, závislosť na konkrétnych protokoloch a nevhodný bezpečnostný mechanizmus boli identifikované ako hlavné problémy súčasnej architektúry.

Model a analýza súčasných problémov sú východiskom pre návrh zlepšenej architektúry. Navrhnutý je nový architektonický štýl RRSS, ktorý na vysokej úrovni abstrakcie definuje základné princípy a konštrukcie pre reprezentáciu údajov. Tento architektonický štýl je použitý ako základ novej architektúry WWW. Navrhovaná architektúra je rozdelená do štyroch vrstiev s rôznou úrovňou abstrakcie. Dve najabstraktnejšie vrstvy sú detailne popísané, naplňujúc cieľ definovať architektúru na konceptuálnej úrovni. Navrhovaná architektúra presnejšie definuje koncept prostriedku, jeho identifikátora, zdroja (ktorý je sám o sebe prostriedkom) a sémantického popisu prostriedku. Práca rozoberá základný prístup k dôveryhodnosti

informácií a zachovaniu súkromia používateľov. Dve konkrétne architektonické úrovne sú okrajovo spomenuté a v hrubých rysoch popísané, nie sú však primárnym cieľom tejto práce. Práca poskytuje nasledujúce prínosy:

- Model popisujúci interakcie medzi reálnym priestorom a kyberpriestorom, zameraný najmä na reprezentáciu údajov o osobách v kyberpriestore.
- Mechanizmus pre vyhodnocovanie anonymity a identity založený na navrhnutom modeli.
- Zhodnotenie nedostatkov súčasnej WWW architektúry. Práca popisuje vnútorné nekonzistencie architektúry, ako aj nedostatky odhalené aplikovaním navrhnutého modelu.
- Definíciu nového architektonického štýlu RRSS na reprezentáciu informácií v kyberpriestore, zohľadňujúc ich potencionálny zdroj a cieľ v reálnom priestore.
- Aplikáciu architektonického štýlu RRSS spolu s ďalšími architektonickými ohraničeniami na architektúru WWW. Výsledkom je návrh novej architektúry WWW popísaný na konceptuálnej úrovni.

Záver

Internet a World Wide Web sú revolučné technológie. Tieto technológie však nevznikli v súčasnej forme, ale postupne sa vyvinuli. Tento evolučný prístup dáva dobré výsledky, ak potrebné zmeny sú malé a ľahko implementovateľné. Evolučný prístup ale zlyhal v prípade veľkých štrukturálnych zmien, ako napríklad zmeny nutné na podporu autenticity údajov, spolupráce a súkromia používateľov.

Až donedávna bola väčšina informácií na Internete určená pre verejné použitie a značne statická. Tento prístup sa však v poslednej dobe mení. Veľká časť informácií je neverejná a dynamická. Organizácia Internetu na jednotlivé sídla už nie je jediné možné usporiadanie. Mechanizmy dynamickej distribúcie údajov ako napríklad „peer-to-peer“ siete zasahujú silne do základných princípov WWW architektúry.

Navrhli sme vylepšenú architektúru WWW, ktorá by mala pomôcť riešiť niektoré principiálne problémy. Navrhnutá architektúra bola vedená modelom interakcií medzi reálneho priestoru a kyberpriestoru. Architektúra bola navrhnutá v snahe zlepšiť prostredie pre spoluprácu na Internete. Snaha na pretvorení WWW je však len na začiatku. Táto práca uložila základnú konceptuálnu vrstvu novej architektúry, na ktorej je možné vystavať kompletnú praktickú architektúru pridávaním chýbajúcich detailov.

Literatúra

[1] SCHNEIER, B.: *Applied Cryptography, Protocols, Algorithms, and Source Code in C.* John Wiley & Sons, 1996. ISBN: 0-471-11709-9.

[2] HALLER, N.: *The S/Key One-Time Password System.* In: ISOC Symposium on Network and Distributed Systems, 1994.

[3] HALLER, N. et al.: *A One-Time Password System.* RFC 2289, 1998.

[4] MENEZES, A., VAN OORSCHOT, P., VANSTONE, S.: *Handbook of Applied Cryptography.* CRC Press, 1996. ISBN: 0849385237.

[5] SIMPSON, W.: *PPP Challenge Handshake Authentication Protocol (CHAP).* RFC 1994, 1996.

[6] *Information Technology - Open Systems interconnection ....* ITU-T Recommendation X.509, 2000.

[7] DIERKS, T., ALLEN, C.: *The TLS Protocol Version 1.0.* RFC 2246, 1999.

[8] RESCORLA, E.: *HTTP Over TLS.* RFC 2818, 2000.

[9] HOEPNER, P. (Ed.): *Study on PKI and Certificate Usage in Europe 2006.* Fraunhofer Institute FOKUS, 2006.

[10] CLARKE, R.: *The Fundamental Inadequacies of Conventional Public Key Infrastructure.* In: Proceedings of ECIS'2001 conference, Bled, Slovenia, 2001.

[11] BRANDS, S.: *Rethinking Public Key Infrastructures and Digital Certificates.* MIT Press, 2000. ISBN: 0-262-02491-8.

[12] CANTOR, S., KEMP, J. (Ed.): *Liberty Protocols and Schema Specification.* Liberty Alliance Project Specification, 2003.

[13] CANTOR, S., KEMP, J., CHAMPAGNE, D. (Ed.): *Liberty Bindings and Profiles Specification.* Liberty Alliance Project Specification, 2003.

[14] MALER, E. et al. (Ed.): *Assertions and Protocol for the OASIS Security Assertion Markup Language (SAML) v1.1.* OASIS, 2003.

[15] BAJAJ, S. et al.: *Web Services Federation Language (WS-Federation).* BEA, IBM, Microsoft, RSA Security, Verisign, 2003.

[16] BAJAJ, S. et al.: *WS-Federation: Passive Requestor Profile.* BEA, IBM, Microsoft, RSA Security, Verisign, 2003.

[17] NANDA, A.: *Identity Selector Interoperability Profile V1.0.* Microsoft, 2007.

[18] *OpenID Authentication 2.0 – Final*, 2007. http://openid.net/specs/openid-authentication-2_0.html

[19] CANTOR, S. et al.: *Shibboleth Architecture, Protocols and Profiles.* 2005. http://shibboleth.internet2.edu/shibboleth-documents.html

[20] HARDT, D.: *The Simple eXtensible Identity Protocol (SXIP) Reference.* 2004. https://sxip.net/archive/specs/sxip-reference.pdf

[21] ERNST, J.: *Light-Weight Identity*. NetMesh Inc., 2005. http://lid.netmesh.org/docs/NetMesh-LID.pdf

[22] PASHALIDIS, A., MITCHELL, C.: *A taxonomy of single sign-on systems.* In: Information Security and Privacy, ACISP, 2003.

[23] SLOT, M.: *Beginner's guide to OpenID phishing.*

http://marcoslot.net/apps/openid/

[24] LAURIE, B.: *OpenID: Phishing Heaven*, 2007.
http://www.links.org/?p=187

[25] ABELSON, H., LESSIG, L.: *Digital Identity in Cyberspace.* White Paper Submitted for 6.805/Law of Cyberspace: Social Protocols, 1998.

[26] PFITZMANN, A., KÖHNTOPP, M.,: *Anonymity, Unobservability, Pseudonymity, and Identity Management A Proposal for Terminology.* In: Designing Privacy Enhancing Technologies, International Workshop on Design Issues in Anonymity and Unobservability, 2000.

[27] PFITZMANN, A., HANSEN, M.: *Anonymity, Unlinkability, Undetectability, Unobservability,Pseudonymity, and Identity Management –A Consolidated Proposal for Terminology v0.31.* 2008.
http://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.31.pdf

[28] WARREN, S., BRANDEIS, L.: *The Right to Privacy.* In: Harvard Law Review, Vol. IV, No. 5, 1890.

[29] SOLOVE, D.: *The Digital Person: Technology and Privacy in the Information Age.* NYU Press, 2004. ISBN: 0814798462.

[30] *Records, Computers and the Rights of Citizens.* Report of the Secretary's Advisory Committee on Automated Personal Data Systems, U.S. Department of Housing, Education, and Welfare, 1973.

[31] *Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.* European Parliament and the Council of European Union, 1995.

[32] RESNICK, P. et al.: *Reputation Systems.* In: Communications of the ACM, 43(12), 2000.

[33] AXELROD, R.: *Evolution Of Cooperation.* Basic Books, New York, 1984.
ISBN: 0465021220.

[34] WINDLEY, P. et al.: *Using reputation to augment explicit authorization.* In: Proceedings of the 2007 ACM workshop on Digital identity management, 2007.

[35] SABATER, J., SIERRA, C.: *Review on computational trust and reputation models.* In: Artificial Intelligence Review, Vol. 24(1), 2005.

[36] FRIEDMAN, E., RESNICK, P.: *The Social Cost of Cheap Pseudonyms.* In: Journal of Economics and Management Strategy, Vol. 10, 2001.

[37] BERNERS-LEE, T.: *Information Management: A Proposal.* 1990.
http://www.w3.org/History/1989/proposal.html

[38] BERNERS-LEE, T., et al.: *Uniform Resource Identifier (URI): Generic Syntax.* RFC 3986, 2005.

[39] FIELDING, R., et al.: *Hypertext Transfer Protocol – HTTP/1.1.* RFC 2616, 1999.

[40] FIELDING, R.: *Architectural Styles and the Design of Network-based Software Architectures.* University of California, Irvine, 2000. Dizertačná práca.

[41] JACOBS, I., WALSH, N. (Ed.): *Architecture of the World Wide Web, Volume One.* W3C Recommendation, 2004.
http://www.w3.org/TR/2004/REC-webarch-20041215/

[42] DHAMIJA, R., TYGAR, J. D., HEARST, M.: *Why Phishing Works.* In: Proceedings of

CHI 2006 Conference on Human Factors in Computing Systems, 2006.

[43] *Internet Usage Statistics, The Internet Big Picture.* 2008. http://www.internetworldstats.com/stats.htm

[44] SHANNON, C.E.: *A Mathematical Theory of Communication.* In: Bell System Technical Journal, vol. 27, 1948.

[45] ECONOMIDES, N.: *The Economics of Networks.* In: Brazilian Electronic Journal of Economics, vol. 1(0), 1997.

[46] CARPENTER, B. (Ed.): *Architectural Principles of the Internet.* RFC 1958, 1996.

[47] *RSS 2.0 Specification*. 2003. http://cyber.law.harvard.edu/rss/rss.html

[48] GARRETT, J.: *Ajax: A New Approach to Web Applications.* 2005. http://www.adaptivepath.com/ideas/essays/archives/000385.php

[49] RAGGETT, D., LE HORS, A., JACOBS, I. (Ed.): *HTML 4.01 Specification.* W3C Recommendation, 1999.

[50] BERNERS-LEE, T.: *What do HTTP URIs Identify?*. 2002. http://www.w3.org/DesignIssues/HTTP-URI.html

[51] BERNERS-LEE, T.: *What HTTP URIs Identify.* 2005. http://www.w3.org/DesignIssues/HTTP-URI2.html

[52] RAMAN, T.V. (Ed.): *On Linking Alternative Representations To Enable Discovery And Publishing*. 2006. http://www.w3.org/2001/tag/doc/alternatives-discovery.html

[53] THOMPSON, H.S., ORCHARD, D.: *URNs, Namespaces and Registries.* 2006. http://www.w3.org/2001/tag/doc/URNsAndRegistries-50

[54] BRAY, T. et al. (Ed.): *Extensible Markup Language (XML) 1.1 (Second Edition).* W3C Recommendation, 2006. http://www.w3.org/TR/xml11/

[55] MENDELSOHN, H., WILLIAMS, S. (Ed.): *The use of Metadata in URIs.* 2006. http://www.w3.org/2001/tag/doc/metaDataInURI-31-20061204.html

[56] BERNERS-LEE, T.: *Cool URIs don't change*. 1998. http://www.w3.org/Provider/Style/URI.html

[57] BERNERS-LEE, T., et al.: *The Semantic Web.* In: Scientific American Magazine, May 17, 2001.

[58] KLYNE, G., CARROLL, J. (Ed.): *Resource Description Framework (RDF): Concepts and Abstract Syntax.* W3C Recommendation, 2004. http://www.w3.org/TR/rdf-concepts/

[59] DOCTOROW, C.: *Metacrap: Putting the torch to seven straw-men of the meta-utopia.* 2001. http://www.well.com/~doctorow/metacrap.htm

Vlastné publikácie súvisiace s predmetom práce

SEMANČÍK, R.: *Basic Properties of the Persona Model*. In: Computing and Informatics, Vol. 26, 2007. pp. 105-121.

SEMANČÍK, R.: *Choosing the best Identity Management Technology for Your Business*. In: Proceedings of InfoSecOn Conference, Cavtat, Croatia, 2006.

SEMANČÍK, R.: *Internet Single Sign-On Systems.* In: BIELIKOVÁ, M. (Ed.): Proceedings of IIT.SRC 2005: Student Research Conference in Informatics and Information Technologies, Bratislava, Slovakia, pp. 116-123, 2005.

SEMANČÍK, R.: *Cesta k digitální identite*. In: Data Security Management, Vol. 5, 2003. pp 16-18.

SEMANČÍK, R.: *Digital Identity as a Basis for Internet Security Infrastructure*. In: Proceedings of 2nd International Conference on Emerging Telecommunications Technologies and Applications, 2003. pp. 175-178.

SEMANČÍK, R.: *Elektronický podpis v praxi*. In: AT&P Journal 9,3, marec 2002. pp. 65-67.

Summary

Internet and World Wide Web were developed in evolutionary fashion as the requirements of the users and usage patterns were constantly changing. The recent requirements on the dynamics of World Wide Web, increased interactivity and effectiveness of cooperation introduce yet another paradigm shift for the Internet. These requirements go deeper into the basic principles and therefore they must be reflected directly into the architecture of World Wide Web. A model that can be used to describe the interactions of physical and virtual worlds is provided for better understanding of the forces behind these changes. The model is focused especially on the persona - representations of physical person in the virtual world. It is shown how anonymity of a person can be estimated by examination of virtual persona. The target environment for the Internet and World Wide Web is discussed; motivated by a desire to better understand requirements of the users. The extremes of anarchical and authoritarian environments are discussed, resulting in a proposal of a middle-ground approach. The aspects of privacy and trust are discussed as well, proposing a mechanism based on reputation systems. The described model and the discussion of desired environment is used to evaluate the architecture of the Internet and World Wide Web. Inconsistencies of World Wide Web architecture are identified and described in detail and architectural improvements are proposed to solve the problems. The improvements are based on the described model, with the goal of supporting the desired environment as it is discussed in this work. The improvements are described in a form of a new architectural styles and constraints, especially the RRSS architectural style. The proposed architecture is divided into several layers of abstraction for easier understanding and maintainability: the layers of architectural styles, architecture, specifications and profiles. The layer of architectural styles is described in detail, the architectural layer is described roughly and only the outline of the other two layers is provided. The work concludes with proposal of specific changes to the World Wide Web architectural guidelines and an outline of the migration path from current architecture to the proposed architecture.