

**SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE**  
Fakulta elektrotechniky a infromatiky

**Autoreferát dizertačnej práce**

**Analýza a klasifikácie prevádzky komunikačných sietí**

Ing. Milan Zelina

na získanie akademického titulu  
*philosophiae doctor, PhD.*  
v doktorandskom študijnom programe: 9.2.9 Aplikovaná informatika

Bratislava, 2014

Dizertačná práca bola vypracovaná v dennej forme doktorandského štúdia na Ústave informatiky a matematiky, Fakulta elektrotechniky a informatiky, Slovenská technická univerzita v Bratislave

Doktorand: Ing. Milan Zelina

Ústav informatiky a matematiky

Fakulta elektrotechniky a informatiky, STU v Bratislave

Ilkovičova 3, 812 19 Bratislava

Školiteľ: prof. Dr. Ing. Miloš Oravec

Ústav informatiky a matematiky

Fakulta elektrotechniky a informatiky, STU v Bratislave

Ilkovičova 3, 812 19 Bratislava

Oponenti: doc. Ing. Ivan Kotuliak, PhD.

Ústav počítačových systémov a sietí

FIIT STU v Bratislave

Dr. Lukáš Kencl

R&D Centrum pre mobilné aplikácie, Katedra telekomunikačnej techniky

Fakulta elektrotechnická ČVUT v Prahe

Autoreferát bol rozposlaný dňa: .....

Obhajoba dizertačnej práce sa koná dňa ..... o .....h, na Fakulte elektrotechniky a informatiky STU, Ilkovičova 3, 812 19 Bratislava v

.....

prof. RNDr. Gabriel Juhás, PhD.

Dekan FEI STU

## Úvod

Oblasť klasifikácie prevádzky je predmetom záujmu od vzniku internetu a zahŕňa aktívny výskum čo dokazuje aj množstvo publikácií v tejto oblasti [1], [2]. Klasifikácia sieťovej prevádzky sa v súčasnosti využíva najmä na oblasti ako: QoS – kvalita služby, tvarovanie prevádzky (ang. traffic shaping), filtrovanie obsahu (ang. traffic filtering), detekcia a prevencia narušenia (intrusion detection / prevention), riadenie zaťaženia (load balancing), meranie spotreby a mnohé ďalšie.

Existuje viacero prístupov k vykonávaniu klasifikácie prevádzky. Klasifikácia na základe čísla portu je pravdepodobne stále najpoužívanejšia a najjednoduchšia technika na rozlišovanie aplikácií.

Prístupy založené na skúmaní informačného poľa paketu majú vysokú výpočtovú náročnosť a úplne zlyhávajú pri klasifikácii šifrovanej prevádzky.

Preto sa v súčasnosti výskum sústreďuje na klasifikáciu prevádzky použitím štatistických vlastností toku. Pri tomto prístupe sa predpokladá, že externe pozorovateľné atribúty tokov na sieťovej vrstve majú štatistické vlastnosti, ktoré majú dostatočnú mieru variancie na odlíšenie jednotlivých typov aplikácií [1]. V našej práci spravidla chápeme tok ako postupnosť IP paketov, ktoré majú rovnakú päťicu: typ protokolu, zdrojová adresa, zdrojový port, cieľová adresa, cieľový port.

Vzťah medzi triedou aplikácií a jej externe pozorovateľnými štatistickými vlastnosťami bol zaznamenaný napríklad už v [3]. Existuje množstvo štatistík, ktoré môžeme použiť na odlíšenie tokov. V literatúre sa používajú rôzne príznaky a ich vhodnosť sa ohodnocuje metódami selekcie príznakov.

## Ciele dizertačnej práce

Na základe prehľadu súčasného stavu sme si stanovili nasledovné ciele:

- Extrakcia a selekcia (identifikácia) príznakov pre klasifikáciu pomocou metód strojového učenia. Výberom vhodnej množiny príznakov môžeme zlepšiť presnosť klasifikácie a znížiť výpočtovú náročnosť tréningu ako aj samotnej klasifikácie. V rámci tohto cieľa je potrebné:
  - Na základe analýzy množiny príznakov pre klasifikáciu, identifikovať najvýznamnejšie príznaky pre klasifikáciu prevádzky. Z dôvodu reprezentatívnosti, vykonáme ich výber na viacerých dátových množinách. Tieto výsledky využijeme pri návrhu ďalších metód.
  - Určiť výkonnosť jednotlivých metód výberu príznakov.
  - Určiť najvýznamnejšie príznaky pre vybrané aplikácie.

- Navrhnuť metódu na stanovenie vhodných hraníc binov pre využitie rozdelenia dĺžky paketov na klasifikáciu prevádzky. V mnohých publikáciách sa tomuto problému venuje pozornosť, ale voľba rozsahu binov sa často nijako nezdôvodňuje [4], [5]. Iní autori používajú priamo zachytené veľkosti paketov [6], preto je potrebné zistiť nakoľko je tento postup vhodný a či je presná veľkosť paketu dôležitá pri klasifikácii.
  - Metóda nebude založená na empirických odhadoch ale z dostupných veľkostí paketov pre dané dáta sama určí vhodné rozdelenie veľkostí binov.
  - Zistiť nakoľko sú biny vhodné pre použitie na rozdelenie času medzi príchodmi paketov.
  - Vyhodnotiť význam informácie o smere prevádzky.
  
- Navrhnuť metódu pre klasifikáciu sieťovej prevádzky založenú na štatistických vlastnostiach toku schopnú pracovať v reálnom čase.
  - Vypracovať metódu schopnú identifikovať typ aplikácie z prvých niekoľko paketov. Počet paketov, z ktorého je možné presne určiť aplikáciu bude predmetom ďalšej analýzy.
  - Za týmto účelom treba preskúmať viacero algoritmov strojového učenia (ang. Machine Learning - ML) najmä ich presnosť klasifikácie tokov, časovú náročnosť a zložitosť vytvoreného modelu prevádzky.
  - Preskúmať vhodnosť navrhnutých metód na klasifikáciu konkrétnych tried prevádzky. Keďže každá aplikácia má vlastné charakteristiky, treba určiť aplikácie, ktoré je možné týmito metódami klasifikovať a presnosť akú dosahujú.
  
- Pre plnenie uvedených cieľov a celkovo pre vykonávanie experimentov s klasifikáciou prevádzky je potrebná implementácia systému na zachytávanie a klasifikáciu prevádzky, ktorý umožní použitie rôznych metód strojového učenia.

## **Teória a použité metódy**

V literatúre pozorujeme, že pre klasifikáciu sa využívajú charakteristiky založené na vlastnostiach jednotlivých paketoch ako aj celých tokov. Jeden z najobsiahlejších zoznamov príznakov je v [7], kde je uvedených 249 príznakov na odlíšenie tokov. Väčšina súčasných prác vychádza práve z týchto príznakov.

Selekcia príznakov je kľúčový krok spracovania dát pri klasifikácii, keďže odstraňuje redundantné, alebo irelevantné príznaky z dát. Použitie metód výberu príznakov nielen spravidla zlepšuje jeho presnosť, ale aj významne znižuje

výpočtovú náročnosť. Identifikácia čo najmenej množiny príznakov je kľúčovou časťou návrhu akéhokoľvek klasifikátora [8].

Algoritmy pre selekciu príznakov sa najčastejšie rozdeľujú na 2 triedy metód [9]:

- metódy založené na všeobecných charakteristikách dát (ang. filter methods), ďalej ich budeme nazývať filter metódy,
- metódy ohodnocujúce príznaky použitím ML metód (ang. wrapper methods), ktoré následne budú použité na klasifikáciu - ďalej ich budeme nazývať wrapper metódy.

Pre zistenie dôležitosti jednotlivých príznakov a vzťahov medzi nimi sú vhodnejšie filter metódy. Keďže wrapper metódy využívajú na ohodnotenie príznakov konkrétny ML algoritmus, je ich použitie vhodnejšie pri implementácii konkrétnej metódy.

Algoritmy na selekciu príznakov, ktoré neohodnocujú príznaky samostatne vyžadujú použitie algoritmu, ktorý hľadá kandidátske podmnožiny z množiny všetkých príznakov. Často sa používajú napríklad metódy sekvenčného výberu: sekvenčný dopredný výber (SFS - začína s prázdnu množinou príznakov a postupne pridáva príznaky, ktoré najviac zvyšujú úspešnosť klasifikácie), sekvenčný spätný výber (SBS - začína s množinou všetkých príznakov a vyraduje príznaky, ktoré najmenej prispievajú k úspešnej klasifikácii), a pod.

Ďalšie metódy na hľadanie a ohodnotenie jednotlivých podmnožín zahŕňajú použitie rôznych deterministických a náhodných algoritmov ako lúčové vyhľadávanie, genetické algoritmy, simulované žihanie a mnohé ďalšie. Ďalšie detaily o technikách na výber množiny príznakov a hľadanie podmnožín môžu byť nájdené napríklad v prácach [8], [10], [11], [12], [13].

Medzi príznakmi, ktoré sa najčastejšie opakujú v rôznych publikáciách sú najmä veľkosti paketov, čas medzi ich príchodmi a rôzne, z nich odvodené príznaky [1]. Veľkosti paketov často nevstupujú priamo do procesu strojového učenia, ale môžu sa deliť na niekoľko intervalov, tzv. binov [14], [15], [4], [16].

Samozrejme neexistuje „ideálny“ počet binov a voľba rôznych intervalov môže v dátach odhaliť rôzne charakteristiky.

V literatúre sa vyskytujú rôzne pokusy, ako určiť optimálny počet a šírku binov, avšak tieto metódy zvyčajne robia rôzne predpoklady o rozdelení dát. Často sa biny jednoducho zvolia empiricky [5], [4], bez zdôvodnenia konkrétnej voľby. Takisto je možné použiť na odhad hraníc binov štatistické odhady [17].

Vo štvrtej kapitole popisujeme ML algoritmy použité pri experimentoch s klasifikáciou prevádzky.

Model neurónu sa skladá zo synáps charakterizovaných svojimi váhami  $w_1, w_2, \dots, w_n$ , ktoré v prvom kroku prenasobia vstupný signál  $x_1, x_2, \dots, x_n$ . Takto prenasobený vstupný signál je potom privedený do sumátora, ktorý spočítava

vstupné signály váhované synaptickými váhami. Hodnotu sumy sa nazýva úroveň vnútornej aktivity neurónu. Túto hodnotu používame ako vstup do aktivačnej funkcie. Výstupom aktivačnej funkcie je samotný výstup neurónu. Štandardný rozsah výstupu aktivačnej funkcie je  $\langle 0, 1 \rangle$  alebo  $\langle -1, 1 \rangle$ . Model neurónu obsahuje aj tzv. prah (ang. threshold)  $\Theta_k$ , ktorý má za úlohu znížiť vstup do aktivačnej funkcie. Ak je potrebné vstup do aktivačnej funkcie naopak zvýšiť, používa záporne zobrazený prah a nazýva sa predpätie (ang. bias).

Viacvrstvé perceptróny (ang. Multilayer perceptron) sa skladajú z neurónov s architektúrou popísanou vyššie. Najčastejšie používanými aktivačnými funkciami sú sigmoidálna (napr. logistická funkcia), ktorou vieme zabezpečiť nelinearitu na výstupe neurónovej siete. Dôležité je aj to, aby táto nelinearita bola hladká, to znamená všade diferencovateľná [18]. Skryté neuróny viacvrstvého perceptrónu nie sú časťou vstupnej, ani výstupnej vrstvy. Práve vďaka skrytým neurónom je viacvrstvý perceptrón vhodný na riešenie veľmi komplexných úloh. Pojem úplné prepojenie znamená, že každý neurón je spojený so všetkými neurónmi z predchádzajúcej vrstvy.

Neurónové siete s radiálnymi bázovými funkciami pracujú na inom princípe ako viacvrstvý perceptrón [18], [19], [20]. Namiesto sumátora, ktorý sa používa v klasickom modeli neurónu, je tu použitá vzdialenostná norma  $\| \cdot \|$  (väčšinou euklidovská). Daná norma určuje vzdialenosť vstupného vektora a váh daného neurónu, ktoré môžeme nazvať aj centrami aktivačných funkcií. Následne je táto vzdialenosť privedená ako vstup do aktivačnej funkcie. Aktivačná funkcia musí byť navrhnutá tak, aby pri vstupe 0 mala maximálny výstup. Je to dané tým, že ak dostaneme vstup do aktivačnej funkcie rovný 0, znamená to, že vstup do siete je identický s nastavenými váhami. Čiže máme vstup, ktorý sa zhoduje s centrom niektorej bázovej funkcie. Čím ďalej sme od 0, tým menšiu hodnotu má daný výstup.

Trénovanie RBF neurónových sietí prebieha v troch krokoch:

Prvý krok je nastavenie centier bázových funkcií (skrytých prvkov). Jednoduchým riešením je vybrať niekoľko bodov z daných vstupov a použiť ich ako centrá. Ďalšie riešenie tohto problému je rovnomerne rozdeliť centrá vo vstupnom priestore. Podobné prístupy však v určitých prípadoch nie sú žiadúce. Niekedy je dôležité, aby centrá vystihovali štruktúru vstupných dát. Na riešenie tejto úlohy existuje mnoho spôsobov (algoritmov) [18].

Druhý krok je voliteľný. Ide o nastavenie prídavných parametrov RBF siete, ak existujú. Vo väčšine prípadov sa jedná o nastavenie parametra  $\sigma$ , ktorý určuje šírku gaussovskej funkcie, ktorá je v RBF sieťach najpoužívanejšia. Tento parameter riadi radiálnu oblasť okolo centra. Všeobecne sa parametre  $\sigma$  hľadajú minimalizáciou chybovej funkcie.

V treťom kroku učenia sa veľmi rýchlym procesom nastavujú váhy. Jeho ukončením sa končí aj samotné trénovanie siete. Všetky parametre sú potom nastavené a sieť je pripravená na simulovanie.

Ďalším algoritmom často využívaným na klasifikáciu prevádzky je rozhodovací strom. Výhoda rozhodovacích stromov oproti ostatným metódam strojového učenia spočíva v tom, že rozhodovacie stromy reprezentujú pravidlá, ktoré je možno vyjadriť prirodzeným jazykom a sú jednoducho pochopiteľné. Najdôležitejším kritériom pri algoritme rozhodovacieho stromu je výber vhodného príznaku v každom rozhodovacom uzle stromu. Cieľom je vybrať taký atribút, ktorým získame najvyššiu úspešnosť klasifikácie. Na meranie vhodnosti konkrétneho atribútu sa používa štatistická vlastnosť nazývaná informačný zisk. Toto meranie sa uskutočňuje pri výbere z kandidátskych atribútov v každom kroku rastu stromu a udáva mieru do akej atribút rozdeľuje tréningové príklady. Zisk je očakávaným zmenšením entropie spôsobeným známou hodnotou atribútu A. Inými slovami vyjadruje informáciu poskytovanú o konkrétnej hodnote atribútu, danej hodnotou iného atribútu A. Je to teda počet bitov ušetrených pri zakódovaní hodnoty ľubovoľného člena S pri známej hodnote atribútu A.

V našich experimentoch sme použili časť prevádzky z dátovej množiny UNIBS [21] obsahujúcej viac ako 15 tisíc tokov a vlastnú zachytenú prevádzku. Používame podobnú kompozíciu prevádzky ako väčšina súčasných prístupov a skladá sa z najmä z HTTP, P2P prevádzky rôznych aplikácií, mailové protokoly, VoIP prevádzky.

## **Dosiahnuté výsledky**

Cieľom dizertačnej práce bola klasifikácia sieťovej prevádzky. Sústredili sme sa najmä na problém včasnej detekcie, ktorý nechápeme len ako klasifikáciu z prvých paketov, ale aj z ľubovoľného sub-toku, ktorý je súčasťou toku.

Kľúčovým krokom je selekcia príznakov. Identifikovali sme veľkosť paketov a čas medzi ich príchodmi ako najdôležitejšie pre odlišenie zvolených aplikácií.

Veľkosti paketov nevstupujú priamo do procesu strojového učenia, ale delia sa do binov. Hľadanie optimálnej voľby binov bolo ďalším cieľom dizertačnej práce. Preskúmali sme rovnomerné rozdelenie intervalu veľkostí, štandardné empirické voľby, štatistické metódy pre určovanie binov a navrhli sme 2 metódy pre voľbu binov.

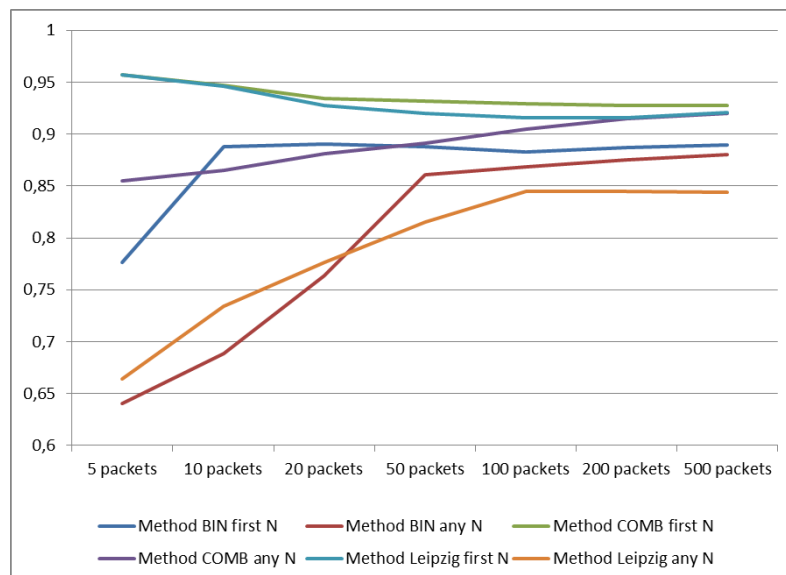
Prvá metóda COG je založená na hľadaní hmotného stredu vo vektore počtosti jednotlivých veľkostí a dosiahli sme ňou najvyššiu úspešnosť klasifikácie zmiešanej prevádzky.

Druhá metóda PEAKS je založená na hľadaní vrcholov v histograme veľkostí. Táto metóda sa osvedčila najmä pri identifikácii aplikácie záujmu, keďže zachováva veľkosti dôležité pre konkrétnu aplikáciu.

Počet potrebných binov a dôležitosť informácie o smere prevádzky sme určili experimentálne.

Tretím cieľom práce bolo navrhnúť metódu pre včasnú detekciu aplikácií zo zmiešanej prevádzky.

Navrhli sme metódu BIN, ktorá odlišuje aplikácie na základe rozdelenia veľkostí paketov. Skúmame počet paketov potrebný na dobrú výkonnosť tejto metódy a to zo začiatku toku ako aj počas jeho behu. Použitím malého množstva paketov zo začiatku toku sa nám pre väčšinu metód podarilo dosiahnuť úspešnosť 70 – 80% a pre sub-toky zachytené počas trvania toku 65 – 75%. Úspešnosť klasifikácie sa stabilizuje po približne 50 zachytených paketoch. Na Obr. 1 vidíme porovnanie našich metód BIN a COMB s referenčnou metódou Leipzig, pre klasifikáciu zo subtokov rôznej veľkosti z prvých paketov toku a z ľubovoľného miest počas behu toku.



**Obr. 1 Porovnanie metód BIN a COMB s metódou Leipzig**

Pre zvýšenie úspešnosti klasifikácie sme navrhli metódu COMB, ktorá kombinuje klasifikáciu na základe rozdelenia veľkostí paketov s inými príznakmi. Všeobecne sú metódy úspešné pri klasifikácii tokov z prvých niekoľko paketov a metóda COMB v tomto prípade dosahuje úspešnosť viac ako 95%.

V porovnaní s inými metódami sme dosiahli dobré výsledky najmä pri klasifikácii z ľubovoľného sub-toku. Použitím malého množstva paketov pre klasifikáciu z ľubovoľného sub-toku sme metódou COMB o takmer 20% lepšie výsledky ako použitím binov, alebo metódy Leipzig [22], ktorú sme zvolili pre porovnanie.



## Sumarizácia publikovaných prínosov a ďalšie smer výskumu

Hľadaniu optimálneho rozdelenia veľkostí paketov do binov sme sa venovali najmä v prácach [A3], [A4], [A12]. Rôzne metódy selekcie príznakov sme implementovali v [A2], [A5], [A9]. Viacero prístupov k problému včasnej detekcie sme publikovali v [A3], [A6], [A7], [A8], [A10], [A11].

Problematika klasifikácie sieťovej prevádzky je veľmi široká téma, existuje k nej viacero prístupov a vyžaduje znalosti z množstva oblastí od znalosti sieťovej infraštruktúry až po metódy strojového učenia. Existuje viacero smerov, ktorými by sa dali aj ďalej vylepšovať naše metódy, od rôznych iných typov predspracovania, až po využitie zhlukovacích metód pre detekciu neznámych tried.

Téma klasifikácie sieťovej prevádzky je nepochybne zaujímavou oblasťou výskumu s veľkým praktickým využitím. Predložená dizertačná práca predstavuje parciálny príspevok k riešeniu tejto problematiky. Snažili sme sa zachytiť najdôležitejšie otázky spracovávanej problematiky, ktorý by mohli byť základom podnietenia záujmu o skúmanú oblasť.

## Vybraná literatúra

- [1] T. Nguyen a G. Armitage, A Survey of Techniques for Internet Traffic Classification using Machine Learning, zv. 10, Communications Surveys & Tutorials, IEEE, 2008, pp. 56 - 76.
- [2] A. Dianotti, A. Pescape a K. Claffy, „Issues and future directions in traffic classification,“ *IEEE Network*, zv. 26, %1. vyd.1, pp. 35 - 40, 2012.
- [3] V. Paxson, Empirically derived analytic models of wide-area TCP connections, *IEEE/ACM Transactions on Networking* vol. 2, 1994, p. 316–336.
- [4] G. Munz, H. Dai, L. Braun a G. Carle, „TCP Traffic Classification Using Markov Models,“ 2010.
- [5] C. Trivedi, M. Chow, A. Nilsson a H. Trussell, „Classification of Internet Traffic using Artificial Neural Networks,“ NC State University, Raleigh, NC, USA, 2002.
- [6] A. Este, F. Gringoli a L. Salgarelli, „Support Vector Machines for TCP traffic classification,“ *Elsevier Computer Networks*, zv. 53, %1. vyd.14, pp. 2476-2490, 2009.
- [7] A. Moore a Z. D., Discriminators for use in flow-based classification, Cambridge, UK: Intel Research, 2005.
- [8] I. Witten a F. Eibe, Data Mining - practical machine learning tools and techniques, Elsevier, Morgan Kaufman publishing, 2005.

- [9] L. Ladha a T. Deepa, „Feature Selection Methods and Algorithms,“ *International Journal on Computer Science and Engineering*, zv. 3, %1. vyd.5, pp. 1787-1797, 2011.
- [10] M. Hall a G. Holmes, „Benchmarking attribute selection techniques for discrete class data mining,“ *IEEE Transactions on Knowledge and Data Engineering*, zv. XV, %1. vyd.6, pp. 1437-1447, 2003.
- [11] D. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Boston, MA, USA: Addison-Wesley Longman Publishing Co., 1989.
- [12] R. Kohavi a G. H. John, „Wrappers for feature subset selection,“ *Artificial Intelligence*, zv. 97, %1. vyd.1-2, pp. 237-324, 1997.
- [13] P. H. Winston, *Artificial intelligence (2nd ed.)*, Boston, MA, USA: Addison-Wesley Longman Publishing Co., 1984.
- [14] C. Wright, F. Monrose a G. Masson, „HMM Profiles for Network Traffic Classification,“ 2004.
- [15] H. J. Trussell, A. A. Nilsson, P. M. Patel a Y. Wang, „Characterization, Estimation and Detection of Network Application Traffic,“ 2005.
- [16] G. Münz, H. S., L. Braun a G. Carle, „Improving Markov-based TCP Traffic Classification,“ rev. *17th GI/ITG Conference on Communication in Distributed Systems*, 2011.
- [17] A. Botta, A. Pescapé a G. Ventre, „On the statistics of QoS Parameters over Heterogenous Networks,“ *Elsevier European Journal of Operational Research*, zv. 191, %1. vyd.3, pp. 1075-1088, 2005.
- [18] M. Oravec, J. Polec, S. Marchevský a kol., *Neurónové siete pre číslicové spracovanie signálov*, Bratislava: Faber, 1998.
- [19] S. Haykin, *Neural Networks - A Comprehensive Foundation*, New York: Macmillan College Publishing Company, 1994.
- [20] K. Hlaváčková, R. Neruda, „Radial Basis Function Networks,“ *Neural Network World*, zv. 1, pp. 93-102, 1993.
- [21] L. Salgarelli a F. Gringoli, „UNIBS dataset,“ The Telecommunication Networks group at UNIBS, 2014. [Online]. Available: <http://www.ing.unibs.it/ntw/tools/traces/>.
- [22] M. Finsterbusch, J. Muller a C. Richter, „Parameter Estimation for Heuristic-based Internet Traffic Classification,“ *The Seventh International Conference on Internet Monitoring and Protection*, 2012.

## **Zoznam vedeckých projektov**

- Návrh pokročilých metód biometrického rozpoznávania na základe obrazov tváre a dúhovky, grant vedeckej grantovej agentúry VEGA 1/0529/13 (zodpovedný riešiteľ M. Oravec, FEI STU), 2013-2016
- Návrh metód analýzy a klasifikácie pre biometrické rozpoznávanie obrazov ľudských tvárí a prevádzku komunikačných sietí, grant vedeckej grantovej agentúry VEGA 1/0214/10 (zodpovedný riešiteľ M. Oravec, FEI STU), 2010-2011

## **Zoznam publikácií autora**

### **Vedecké monografie vydané v domácich vydavateľstvách**

[A1] Miloš Oravec, Jarmila Pavlovičová, Luboš Omelina, Matej Féder, Jozef Ban, Ján Mazanec, Miloslav Valčo, Milan Zelina: Metódy strojového učenia na extrakciu príznakov a rozpoznávanie vzorov, 2. diel : rozpoznávanie tvárí v biometrii, 1. vyd., Bratislava: FELIA, s.r.o., 2013, 179 s, ISBN 978-80-971512-0-1.

### **Vedecké práce v zahraničných časopisoch**

[A2] Milan Zelina, Miloš Oravec: Feature Selection for Application Recognition in Communication Networks, In: AD ALTA : Journal of Interdisciplinary Research, ISSN 1804-7890, Vol. 1, Iss. 1, 2011, s. 115-117, vyžiadaný príspevok.

[A3] Milan Zelina, Miloš Oravec: Classification of Internet Applications using Multiple Features in Sub-Flows, 2014.

### **Vedecké práce v domácich časopisoch**

[A4] Milan Zelina, Miloš Oravec: Estimation and Reduction of Bins in Traffic Classification,

In: EE časopis pre elektrotechniku a energetiku, ISSN 1335-2547, Roč. 16, mimoriadne č.: ELOSYS, Trenčín, 5.-8.10. 2010, s. 53-57.

### **Vedecké práce v domácich recenzovaných vedeckých zborníkoch, monografiách**

[A5] Milan Zelina, Miloš Oravec: Feature Selection and Classification of Internet Traffic,

In: IP Networking 1 - Theory and Practice, Žilina: EDIS, 2011, ISBN 978-80-554-0494-3, s. 46-50.

### **Publikované príspevky na zahraničných vedeckých konferenciách**

[A6] Milan Zelina, Miloš Oravec: A Method for Early Detection of Internet Applications,

In: Redžúr 2012: Proceedings of 6th International Workshop on Multimedia and

Signal Processing, April 11, 2012, Vienna, Austria. Bratislava: Nakladateľstvo STU, 2012, ISBN 978-80-227-3686-2, s. 1-4.

[A7] Milan Zelina, Miloš Oravec: Early Detection of Internet Applications, In: Proceedings of 9th IEEE International Conference VSACKÝ CÁB 2011, Brno University of Technology, 2011, ISBN 978-80-214-4319-8, s. 157-160.

[A8] Milan Zelina, Miloš Oravec: Early Detection of Network Applications Using Neural Networks.

In: Proceedings ELMAR-2011: 53rd International Symposium ELMAR-2011, 14-16 September 2011, Zadar, Croatia, Zadar: Croatian Society Electronics in Marine, 2011, ISBN 978-953-7044-12-1, s. 161-164.

### **Publikované príspevky na domácich vedeckých konferenciách**

[A9] Milan Zelina, Dávid Hrbatý: Early Detection and Classification of Internet Applications using Neural Networks, In: Redžúr 2013: proceedings; 7th International Workshop on Multimedia and Signal Processing; Smolenice, Slovakia, 1. Máj 2013, Bratislava: Nakladateľstvo STU, 2013, ISBN 978-80-227-3921-4, s. 39-43.

[A10] Milan Zelina, Miloš Oravec: Decision Trees for Recognition of Internet Applications,

In: EE časopis pre elektrotechniku a energetiku, ISSN 1335-2547, Roč. 17, mimoriadne č.: ELOSYS, Trenčín, 11.-14.10.2011, 2011, s. 234-237.

[A11] Milan Zelina, Miloš Oravec: Representation of Traffic in Communication Networks,

In: ELITECH'10 : 12th Conference of Doctoral Students, Bratislava, Slovak Republic, 26.5.2010, Bratislava: STU v Bratislave, 2010, ISBN 978-80-227-3303-8, CD-Rom, ocenenie za najlepšiu prezentáciu.

### **Skriptá a učebné texty**

[A12] Miloš Oravec, Matej Féder, Milan Zelina: Strojové učenie a neurónové siete: učebné texty, 1. vydanie, Bratislava: RT Systems, 2013, ISBN 978-80-970519-5-2.