Curriculum Vitae

Personal information

Name:	**Tomasz Bujlow**
Address:	Leipzig, Germany
Mobile:	+48 504 971 103
E-mail:	tomasz@bujlow.com
Citizenship:	European Union, Polish
Date of birth:	August 05, 1984
Home page:	http://tomasz.bujlow.com
LinkedIn:	http://www.linkedin.com/in/tomaszbujlow
Google Scholar:	http://scholar.google.com/citations?user=WvFturoAAAAJ

Summary

Currently, I am working on the development of Network Intelligence (NI) software solutions, which involve traffic classification, analysis, and complete decoding of detected protocols and applications. These solutions are characterized by high performance for core network links with speeds up to 100 Gbit/s and faster. They use various technologies (e.g., Deep Packet Inspection, behavioral, heuristic, and statistical analysis) to reliably detect network protocols, applications, and services, and extract metadata, in real time. I am a daily user of G Suite, Atlassian Software (JIRA, Confluence) and GIT. From the time of taking this position, I am working with Agile development methodologies including SCRUM. During that time, I identified many aspects of SCRUM that are critical from the quality and development productivity points of view. I also served as a customer support developer channel, which made me able to better understand how the customers see and use our software and what are their priorities for the product development and maintenance. As sharing the technical knowledge is my passion, I organized multiple training workshops related to computer networks and network traffic analysis.

I obtained my PhD in Classification and Analysis of Computer Network Traffic from Aalborg University in Denmark on June 6, 2014. My PhD project was co-financed and co-supervised by Bredbånd Nord A/S, a regional electricity and Internet provider. Due to this industrial collaboration, I learned how to collect and understand customer requirements, present high-level concepts and results to the company management, and structure the work in order to reach both the scientific and industrial goals on time.

I was the founder and developer of nDPIng - the next generation open-source computer network traffic classification tool, which aims in consistent real-time traffic identification on multiple levels: transport layer protocol, all application level protocols, type of content, service provider, and content provider. I was also the principal investigator in the Volunteer-Based System for Research on the Internet project, which was focused on designing and developing a system, which is able to provide detail data about applications used in the Internet. This information can be used for obtaining the knowledge which applications are most frequently used in the network, providing the users some basic statistics about their Internet connection usage (for example, for which kinds of applications their connection is used the most), creating scientific profiles of traffic generated by different applications or different groups of applications.

I used to work independently and cover the entire development process, from architecture, design, implementation and customer feedback up to bug fixing. Apart from the nDPIng and Volunteer-Based System for Research on the Internet projects, I fully authored 2 industrial projects. Web-Based Client for InDesign Server uses web-based techniques and tools in collaboration with a headless version of InDesign Server, which is controlled by scripts produced by the designed web interface, to render InDesign documents in the real-time. The Efficient Invoicing Solution with Offline Synchronization Capabilities project was concentrated on creating an invoicing system for a mining company, which is characterized by a significant fraction of features differing from other systems already existing in the market. The designed and implemented system was in use in around 30 departments of OPA-LABOR during 4 years, successfully satisfying all the requirements set in this project.

I am quick in learning new technologies (e.g., programming languages, development platforms and frameworks) and using the new knowledge and skills in practice, which allows me to easily switch between different IT-related fields. I put the ability to solve problems with the help of the Internet, books, or other people above the encyclopedic knowledge (e.g., knowing by heart the syntax of a particular programming language or an already existing and documented algorithm).

During my PhD, I was a visiting researcher at Universitat Politècnica de Catalunya (UPC) in Barcelona, Spain, where I was working together with the Broadband Communications Research Group on the comparison of Deep Packet Inspection Tools for traffic classification. I was also visiting ntop in Pisa, Italy (collaboration on nDPI) and TELECOM Sudparis in Evry, France (collaboration on traffic classification in 802.11). I am an author of 4 journal articles, 8 conference papers, and 3 technical reports on the topics related to traffic monitoring and analysis. Two of my papers got awards as top 7 % and top 5 %, respectively. Since 2011, I gave 11 presentations in seminars and guest lectures at Aalborg University in Denmark, TELECOM Sudparis in France, University of Pisa in Italy, Polytechnic University of Turin in Italy, RWTH University in Germany, Universitat Politècnica de Catalunya in Spain, IDA House of Engineers in Denmark, and Albena Resort in Bulgaria. I am a reviewer of 15 articles submitted to different journals and conferences.

During my postdoctoral research, I investigated techniques used for tracking users' activity online. Many content providers and online retailers collect large amounts of personal information from their users when browsing the web. The large scale collection and analysis of personal information constitutes the core business of most of these companies, which use this information for lucrative purposes, such as online advertising and price discrimination. However, most mechanisms used to track users and collect personal information are not well known or intentionally obfuscated. The main objective was to uncover these mechanisms and understand how they collect, analyze, store and (possibly) sell this information.

I am also a holder of 2 language certificates: TOEFL iBT (98/120) and Prøve i Dansk 3 (9/12).

Does that sound interesting? If yes, you are welcome to contact me, as, currently, I am looking for new job opportunities worldwide! I am open to almost any form of employment - I can work as a full-time company employee as well as a contracted project-based consultant. However, I would like to be able to work at least half of the time remotely from home (in a way according to the needs of both the company and me).

Work experience

	Position	Employer	Start date	End date
9	Senior Developer DPI	ipoque GmbH	06/2015	—
8	Founder and Developer of nDPIng	Own open-source project	03/2014	04/2015
7	Postdoctoral Investigator	Universitat Politècnica de Catalunya	10/2014	03/2015
6	PhD Student	Aalborg University	12/2010	12/2013
5	Visiting PhD Student	Universitat Politècnica de Catalunya	01/2013	04/2013
4	PHP Software Developer, Project Leader	cbit / Imento	08/2008	09/2010
3	Home Delivery Assistant	Morgendistribution Danmark	11/2007	01/2008
2	Wireless Network Specialist (internship)	Proximetry Poland	07/2007	08/2007
1	C++ Software Developer (internship)	OPA-LABOR	04/2007	08/2007

9

Period	June 2015 – Present
Occupation or position held	Senior Developer DPI
Activities and responsibilities	Development of Deep Packet Inspection (DPI) / Network Intelligence (NI) solutions: R&S® Protocol and Application Decoding Engine (PADE), R&S®PACE 2 (learn more: https://ipoque.com/products/pace).
Name of employer	ipoque GmbH
Address of employer	Augustusplatz 9, 04109 Leipzig, Germany
Phone	+49 341 594030
Fax	+49 341 594030 19
E-mail	info.ipoque@rohde-schwarz.com
Details	R&S®PACE 2 is the next generation software library that identifies thousands of protocols, applications, and services, and provides deeper insight into application attributes (e.g., real-time performance metrics). R&S®PACE 2 combines the power of the Protocol and Application Classification Engine (PACE) and decoding engine (PADE), and is also capable of advanced metadata extraction. This solution is characterized by high performance for core network links with speeds up to 100 Gbit/s and faster. It uses various technologies (e.g., Deep Packet Inspection, behavioral, heuristic, and statistical analysis) to reliably detect network protocols, applications, and services, and extract metadata, in real time. Key performance indicators are calculated for deeper insight.

The decoding results of R&S®PACE 2 provide the deepest information about the current connection. R&S®PACE 2 extracts all important and relevant metadata from a number of network classification results with a configurable level of detail to suit different use cases. For example, it is possible to decompress HTTP payload and reconstruct all images or videos from internet sites. The depth of information required can be flexibly adjusted to provide just the actual data needed. Internal aggregators gather decoding information from certain decoders and bundle them into classes. For example, even if an email connection takes a long time, the full session decoding information still provides all of the data in one single place. The decoding feature of R&S®PACE 2 is especially useful in network security applications, e.g, the playback of VoIP calls, websites and chat sessions, or gathering upload and download statistics of various documents.

8

Period	March 2014 – April 2015
Occupation or position held	Founder and Developer of nDPIng
Activities and responsibilities	Development of the next generation computer network traffic classification tool
Type of activity	Own open-source project
Initial development location	Computer Science Department, University of Pisa, Pisa, Italy
Status	Development stage
Accessible (SVN)	https://svn.ntop.org/svn/ntop/trunk/nDPIng/
Details	The aim of this unique project is to bring new quality to the field of traffic classification by providing the results on many levels. The clear, unambiguous identification of network flows is meant to be ensured by various classification techniques combined into a single tool. The following information is intended to be given for each flow inspected by the classifier: transport layer protocol, all the application-layer protocols, type of the content, service provider, and content provider. Look at the *Projects* section for a detailed description.

7

Period	October 2014 – March 2015
Occupation or position held	Postdoctoral Investigator
Activities and responsibilities	Research on online users' privacy
Name of employer	Universitat Politècnica de Catalunya, Department of Computer Architecture, Broadband Communications Research Group
Address of employer	Jordi Girona, 1–3, 08034 Barcelona, Spain
Phone	+34 93 401 69 82
Fax	+34 93 401 70 55
E-mail	pbarlet@ac.upc.edu
Details	It is widely known that content providers and online retailers (e.g., Google, Facebook and Amazon) collect large amounts of personal information from their users when browsing the web. The large scale collection and analysis of personal information constitutes the core business of most of these companies, which use this information for lucrative purposes, such as online advertising and price discrimination. However, most mechanisms used to track users and collect personal information are still unknown. Our main objective is to uncover these mechanisms and understand how they collect, analyze, store and (possibly) sell this information. Personal information in the web can be voluntarily given by the user (e.g., by filling web forms) or it can be collected indirectly without their explicit knowledge through the analysis of the IP headers, HTTP requests, queries in search engines, or even by using JavaScript and Flash programs embedded in web pages. Among the collected data, we can find information of technical nature (e.g., the browser in use) and also more sensible information (e.g., the geographical location or the visited web pages). The webmail services are also known for scanning and processing user's e-mails, even if they are received from a user who did not allow any kind of message inspection. In order to track their users, online services use various methods. The most popular techniques are the use of different kinds of browser cookies, fingerprinting the user in the background, or suggesting (or requiring) the user to fill in a profile, so the web identity can be further extended by associating it with the real user's identity.

We investigate whether the services are using other, unexpected mechanisms to track user activity, as if the network of contacts of a user and their interests are used to build the profile of the users, and what impact it has on their privacy. We also analyze if online services collect information when users are logged off to a service using cookies or user fingerprints and later combine this information with their online profiles when they log in. We investigate the ability of web services to follow the users' activity in the *private browsing* mode and analyze special privacy-focused search engines. We test their capabilities and compare them with the standard search engines. In another front, we investigate the impact of user tracking in price discrimination. Product pricing can be based on the geographical location of the user but also on the user profiles sold by online services.

6

Period	December 2010 – December 2013
Occupation or position held	PhD Student
Activities and responsibilities	Classification and analysis of computer network traffic
Name of employer	Aalborg University, Department of Electronic Systems, Networking and Security Section
Address of employer	Fredrik Bajers Vej 7, 9220 Aalborg Øst, Denmark
Phone	+45 9940 8616
Fax	+45 9940 9840
E-mail	netsec@es.aau.dk
Details	Our objective: to evaluate the performance of various applications in a high-speed Internet infrastructure. 1. We performed substantial testing of widely used DPI classifiers (PACE, OpenDPI, L7-filter, NDPI, Libprotoident, and NBAR) and assessed their usefulness in generating ground-truth, which can be used as training data for Machine Learning Algorithms (MLAs). 2. Because the existing methods (DPI, port-based, statistical) were shown to not be sufficient, we built our own host-based system (VBS) for collecting and labeling of network data. The packets are grouped into flows, which are labeled by the process name obtained from the system sockets. Look at the *Projects* section for a detailed description. 3. We assessed the usefulness of C5.0 MLA in the classification of computer network traffic. We showed that the application-layer payload is not needed to train the C5.0 classifier, defined the sets of classification attributes and tested various classification modes. 4. We showed how to use our VBS tool to obtain per-flow, per-application, and per-content statistics of traffic in computer networks. Furthermore, we created two datasets composed of various applications, which can be used to assess the accuracy of different traffic classification tools. The datasets contain full packet payloads and they are available to the research community as a set of PCAP files and their per-flow description in the corresponding text files. 5. We designed and implemented our own system for multilevel traffic classification, which provides consistent results on all of the 6 levels: Ethernet, IP protocol, application, behavior, content, and service provider. The system is able to deal with unknown traffic, leaving it unclassified on all the levels, instead of assigning the traffic to the most fitting class. Our system was implemented in Java and released as an open-source project. 6. Finally, we created a method for assessing the Quality of Service in computer networks.

5

Period	January 2013 – April 2013
Occupation or position held	Visiting PhD Student
Activities and responsibilities	Comparison of Deep Packet Inspection tools for traffic classification
Name of employer	Universitat Politècnica de Catalunya, Department of Computer Architecture, Broadband Communications Research Group
Address of employer	Jordi Girona, 1–3, 08034 Barcelona, Spain
Phone	+34 93 401 69 82
Fax	+34 93 401 70 55
E-mail	pareta@ac.upc.edu

Details	The outcomes were thoroughly described in a technical report Comparison of Deep Packet Inspection (DPI) Tools for Traffic Classification, which is shown below in the Publications section. 1. We created a dataset of 10 different applications (eDonkey, BitTorrent, FTP, DNS, NTP, RDP, NETBIOS, SSH, HTTP, RTMP), which is available to the research community. It contains 1 262 022 flows captured during 66 days. The dataset is available as a bunch of PCAP files containing full flows including the packet payload, together with corresponding text files, which describe the flows by providing all the necessary details, including the corresponding application name, start, and end timestamps based on the system sockets. 2. We tested the accuracy of several Deep Packet Inspection tools (PACE, OpenDPI, L7-filter, NDPI, Libprotoident, and NBAR) on our dataset. To test NBAR, we needed to replay the packets to the Cisco router and process the Flexible NetFlow logs. The other tools were tested directly as libraries by a special software, which was reading packets from the PCAP files and providing the packets to the classifiers.

4	
Period	August 2008 – September 2010
Occupation or position held	PHP Software Developer, Project Leader
Activities and responsibilities	Development of Imento product. My own project: Web-Based Client for InDesign Server (look at the *Projects* section for a detailed description)
Name of employer	cbit / Imento
Address of employer	Cikorievej 20A, 5220 Odense SØ, Denmark
Phone	+45 4098 1417
E-mail	clausbolund@me.com

3	
Period	November 2007 – January 2008
Occupation or position held	Home Delivery Assistant
Activities and responsibilities	Providing delivery of products to customers or designated locations dispatched from the Central Delivery Depot
Name of employer	Morgendistribution Danmark
Address of employer	Fjordsgade 11, 1. sal., 5000 Odense C, Denmark
Contact	none known – bankruptcy

2	
Period	July 2007 – August 2007
Occupation or position held	Wireless Network Specialist (internship)
Activities and responsibilities	Designing and developing Quality of Service measurement software for Wireless Local Area Networks, WiMax testing
Name of employer	Proximetry Poland
Address of employer	Roździeńskiego 91, 40-203 Katowice, Poland
Phone	+48 322 580 682
Fax	+48 322 000 333
E-mail	mossysek@gmail.com

1	
Period	April 2007 – August 2007
Occupation or position held	C++ Software Developer (internship)
Activities and responsibilities	Development of an application used for creating, managing and printing invoices. This program was in use in around 30 departments of OPA-LABOR during 4 years. Look at the *Projects* section for a detailed description
Name of employer	OPA-LABOR
Address of employer	Wyzwolenia 22, 41-103 Siemianowice Śląskie, Poland
Phone	+48 322 281 340 - 102
Fax	+48 322 201 131
E-mail	opa@opalabor.pl

Certificates

5

Certified	October 2014
Validity	October 2016
Title	TOEFL iBT | certificate
Scores	Total: 98/120 (82 %), Reading: 28/30, Listening: 22/30, Speaking: 23/30, Writing: 25/30
Subjects/skills covered	English language knowledge
Issuing institution	Educational Testing Service, USA

4

Certified	December 2012
Validity	unlimited
Title	Bevis for Prøve i Dansk 3 | certificate
Subjects/skills covered	Danish language knowledge
Issuing institution	Ministry of Education (Undervisningsministeriet), Denmark

3

Certified	November 2010
Re-certified	October 2013
Validity	October 2016
Title	Cisco Certified Network Professional (CCNP) | certificate
Subjects/skills covered	Administration of LAN, WLAN, and WAN computer networks
Issuing institution	Cisco Systems, USA

2

Certified	September 2006
Re-certified	October 2013
Validity	October 2016
Title	Cisco Certified Network Associate (CCNA) | certificate
Subjects/skills covered	Administration of LAN, WLAN, and WAN computer networks
Issuing institution	Cisco Systems, USA

1

Certified	March 2007
Validity	unlimited
Title	English Language Certificate for Applicants for The International Association for the Exchange of Students for Technical Experience (IAESTE) Training
Subjects/skills covered	English language knowledge
Issuing institution	Silesian University of Technology (Politechnika Śląska), Poland

Education

3

Period	December 2010 – June 2014
Degree	Doctor of Philosophy (PhD) | diploma
Thesis title	Classification and Analysis of Computer Network Traffic
Main supervisor	Jens Myrup Pedersen, Aalborg University, Aalborg, Denmark
Co-supervisor	Tahir Riaz, Aalborg University, Aalborg, Denmark
Co-supervisor	Pere Barlet-Ros, Universitat Politècnica de Catalunya (UPC), Barcelona, Spain
University	Aalborg University, Aalborg, Denmark

2

Period	September 2007 – June 2009
Degree	Bachelor of Computer Engineering | diploma
Field of study	Computer Engineering, Faculty of Engineering
University	University of Southern Denmark (Syddansk Universitet), Odense, Denmark

1

Period	October 2003 – October 2008
Degree	Master of Science in Engineering | diploma
Field of study	Computer Engineering, specialty: Databases, Computer Networks and Computer Systems
University	Silesian University of Technology (Politechnika Śląska), Gliwice, Poland

Professional training

3	**Period**	October 2009 – June 2010
	Title	CCNP 1 Building Scalable Cisco Internetworks (BSCI), v. 5.0 CCNP 2 Implementing Secure Converged Wide Area Networks (ISCW), v. 5.0 CCNP 3 Building Cisco Multilayer Switched Networks (BCMSN), v. 5.0 CCNP 4 Optimizing Converged Cisco Networks (ONT), v. 5.0
	Name of organization	Cisco Networking Acadamy
2	**Period**	October 2005 – February 2007
	Title	MS SQL Server Managing and Maintaining a Microsoft Windows Server 2003 Environment Implementing and Supporting Microsoft Windows XP Professional
	Name of organization	Silesian University of Technology (Politechnika Śląska) & Microsoft Corporation
1	**Period**	October 2005 – September 2006
	Title	CCNA 1 Networking Basics, v. 3.1 CCNA 2 Routers and Routing Basics, v. 3.1 CCNA 3 Switching Basics and Intermediate Routing, v. 3.1 CCNA 4 WAN Technologies, v. 3.1
	Name of organization	Cisco Networking Acadamy

Languages

Language	Reading	Writing	Speaking
Polish	Native	Native	Native
English	Advanced	Advanced	Advanced
Danish	Intermediate	Elementary	Elementary
German	Elementary	Elementary	Elementary

Driver's licenses

Category	Valid for	Valid from
European B	Motor vehicles	December 2002
European AM	Mopeds	December 2002

Other skills and competences

Academic skills

Research, experimentation, supervision, teaching, scientific writing, LaTeX, typesetting

Computer networks

Network monitoring, traffic analysis and classification, Deep Packet Inspection (DPI)
Routing protocols (RIP, OSPF, and BGP) & switching
TCP/IP stack, HTTP, SSL, DNS

Databases

Planning, designing, implementing, troubleshooting and securing databases
SQLITE, MSSQL, MySQL, and PostgreSQL database servers, SQL programming

Software development

Scrum agile framework
C/C++, Python, Java, SQL, PHP, JavaScript, AJAX, and InDesign Server programming
Deep Packet Inspection, DNS inspection, BGP analysis, Autonomous Systems matching, client-server applications, raw sockets, system sockets monitoring, network protocol decoding and classification
Atlassian software (JIRA, Confluence), GIT, Gerrit, Jenkins

Operating systems

Windows and Linux operating systems
Implementing, securing and troubleshooting Linux routers, including wireless routers

Internet and Web Services

HTML, JavaScript, DHTML, PHP languages, and AJAX technology
Managing and troubleshooting WWW servers and websites, Internet portals, databases, and control panels (e.g. cPanel)
E-mail, WWW, DNS, instant messaging, P2P technology, Windows and Linux firewalls, and proxies

Grants and scholarships

3 **Period** December 2010 – December 2013
Description PhD scholarship, grant no. 8–10100
Providers Aalborg University, Denmark
Bredbånd Nord A/S, Denmark
European Regional Development Fund (ERDF)

2 **Period** January 2013 – April 2013
Description Research grant for 3-month stay at Universitat Politècnica de Catalunya (UPC) in Barcelona, Spain
Provider Aalborg University, Denmark

1 **Period** September 2007 – June 2008
Description ERASMUS (European Region Action Scheme for the Mobility of University Students) student grant.
Destination: Syddansk Universitet (University of Southern Denmark), Odense, Denmark
Provider European Union

Memberships

3 **Period** February 2017 – Present
Organization A Million Happy Cats Association (Stowarzyszenie Milion Szczęśliwych Kotów)
Description Non-Governmental Organization (NGO) in Szczecin, Poland | web

2 **Period** April 2012 – December 2015
Organization Institute of Electrical and Electronics Engineers (IEEE)
Description Member #92273053

1 **Period** November 2012 – December 2013
Organization PhD Network at Aalborg University (PAU)
Description Board Member of the official association of PhD Students at Aalborg University

Distinctions and awards

2

Date	February 2012
Description	Certificate of Outstanding Paper Award. Top 7% of 597 submissions to the ICACT 2012 conference
Awarder	Global IT Research Institute, Republic of Korea

1

Date	February 2012
Description	Distinguished group of 5% best papers presented at TELFOR 2011
Awarder	TELFOR Journal Editor, Serbia

Projects

Scientific projects

4

Period	October 2014 – March 2015
Title	Architecture with Knowledge of the Environment for the Future Internet (Arquitectura con Conocimiento del Entorno de la Futura Internet)
Role	Project Investigator
Project code	K00530
Funding entity	Ministry of Economy and Competitiveness (Ministerio de Economía y Competitividad), Spain
Funding entity code	EUIN2013-51199
Budget	25 000.00 EUR
Start date	2014-01-01
End date	2015-07-31
Scientific coordinator	Josep Solé Pareta, Universitat Politècnica de Catalunya, Spain

3

Period	March 2014 – April 2015
Title	nDPIng – Next Generation Traffic Classification Library
Role	Principal Investigator
Project code	nDPIng
Funding entity	None
Start date	2014-03-01
End date	Undefined
Accessible (SVN)	https://svn.ntop.org/svn/ntop/trunk/nDPIng/
Details	The aim of this unique project is to bring new quality to the field of traffic classification by providing the results on many levels. The results obtained from nDPIng are easy to be accounted and they are given as: protocol (beginning from TCP/UDP, then going into higher levels), content type, service provider (the well-known name of the remote host , e.g., *Facebook* for web browser flows from Facebook), and content provider (content delivery network: cdn, e.g., Akamai or Google). Examples of the results provided in the non-verbose mode: - *proto: TCP->SSL_with_certificate->POP3S, service: Google* – an encrypted POP3 session with a Google mail server. - *proto: TCP->SSL_with_certificate, service: Twitter* – an encrypted connection to a Twitter server. - *proto: TCP->FTP_Data, content: JPG* – a file-transfer FTP session, which carries a JPG image. - *proto: TCP->SSL_with_certificate->Dropbox, cdn: Dropbox* – an encrypted Dropbox session (the application is Dropbox) with the Dropbox server. - *proto: TCP->SSL_with_certificate, cdn: Dropbox* – an encrypted session with a Dropbox server, while the application is unknown (it can be a web browser connection). - *proto: TCP->HTTP, content: WebM, service: YouTube, cdn: Google* – a flow from YouTube coming from Google server, which transports WebM movie. - *proto: TCP->HTTP, service: Google, cdn: Google* – an HTTP flow from Google, obtained from the Google server. There is a possibility to obtain the domain names which are associated with the service and content providers – see the example application attached to the project.

2

Period	January 2011 – December 2013
Title	Volunteer-Based System for Research on The Internet
Role	Principal Investigator
Project code	VBS
Funding entities	Aalborg University, Denmark Bredbånd Nord, Denmark European Regional Development Fund (ERDF)
Start date	2011-01-01
End date	Undefined
Accessible	http://vbsi.sourceforge.net
Details	This project is focused on designing and developing a system, which is able to provide detail data about applications used in the Internet. This information can be used for obtaining the knowledge which applications are most frequently used in the network, providing the users some basic statistics about their Internet connection usage (for example, for which kinds of applications their connection is used the most), creating scientific profiles of traffic generated by different applications or different groups of applications, etc. The developed Volunteer-Based system has the client-server architecture. Clients are installed among machines belonging to volunteers, while the server is installed on the computer located in the premises of the data collecting entity. Each client registers information about the data passing computer's network interfaces. Captured packets are grouped into flows. A flow is defined as a group of packets which have the same local and remote IP addresses, local and remote ports, and using the same transport layer protocol. For every flow, the client registers: anonymized identifier of the client, start timestamp of the flow, anonymized local and remote IP addresses, local and remote ports, transport protocol, anonymized global IP address of the client, and name of the application associated with that flow. The name of the application is taken from the system sockets. For every packet, the client additionally registers: direction, size, state of all TCP flags (for TCP connections only), time in microseconds elapsed from the previous packet in the flow, and type of transmitted HTTP content. We do not inspect the payload – the type of the HTTP content is obtained from the HTTP header, which is present in the first packet carrying this specific content. One HTTP flow (for example a connection to a web server) can carry multiple files: HTML documents, JPEG images, CSS stylesheets, etc. Thanks to that ability implemented in our VBS, we are able to split the flow and separate particular HTTP contents. The data collected by VBS are stored in a local file and periodically sent to the server. The task of the server is to receive the data from clients and to store them into the MySQL database. This open source tool is released under *GNU General Public License v3.0* and published as a SourceForge project. Both Windows and Linux versions are available. VBS is designed to collect the traffic from numerous volunteers spread around the world and, therefore, with a sufficient number of volunteers the collected data can provide us with a good statistical base.

1

Period	December 2010 – March 2013
Title	Collaborating Living Labs
Role	Project Member
Funding entity	NordForsk, Norway
Start date	2010-08-01
End date	2013-03-31
Scientific coordinator	Mari Linn Larsen, University of Stavanger, Norway
Accessible	http://www.coll-livinglab.org
Details	Compare Testlab in Karlstad, NettOp at the University of Stavanger, and CNP at Aalborg University, are three living labs for development of new ICT-services, infrastructure and media by means of involving users (i.e. end users as well as companies). The industrial partners Ipark (Stavanger Innovation Park), ICTNORCOM, and the Greater Stavanger Development will present real cases to which users will be invited to co-create and test ICT services. The aim of this project is to build on and improve the work of existing Living Labs and generate knowledge on how to innovate new services, media and infrastructure in Living Labs in three different Nordic countries.

Other projects

3

Period	November 2014 – April 2015
Title	Deep Packet Inspection API Standardization
Role	Project Member
Funding entity	None (collaborative open-source project)
Start date	2014-11-10
End date	Undefined
Industrial coordinator	Franck Baudin, Qosmos, France
Accessible	http://groups.google.com/d/forum/dpi-api-standardization-group
Details	This project aims at defining a standard Deep Packet Inspection API that most DPI implementations will support. In order to achieve this goal, the API will be released under an open license. This will promote DPI libraries interchange, so that it will be possible to plug/unplug implementations as needed. The standardization group consists of developers of both the commercial and open-source DPI software.

2

Period	February 2009 – September 2010
Title	Web-Based Client for InDesign Server
Role	Project Leader, Principal Software Developer
Funding entity	Imento, Denmark
Start date	2009-02-01
End date	2010-09-30
Industrial coordinator	Claus Bolund Pedersen, Imento, Denmark
Details	The goal of this project was to design and implement a new module for Imento – a web-based system for creating fliers and advertisements, which is in use by many well-known companies in Denmark, e.g., 727, Cosmographic, Lidl, Spar, Bong, Nordal, Intersport, Bygma, and Tempur. The system consists of a media bank and a product database, which are used to store all the information about the products needed by the customers. The task of the module being the subject of this project was to allow easy production of real advertisements, in the inDesign and PDF formats, using the web-based Imento interface. The built solution uses web-based techniques and tools (e.g., HTML, JavaScript, jQuery, and AJAX) in collaboration with a headless version of inDesign Server, controlled by scripts produced by the web interface. At first, the user is able to choose a template used for building the advertisement. Then, the website turns into an environment known from drawing and painting applications, where the user can use existing snippets (per-product graphical templates) to build multi-pages multi-layer document by dragging and dropping the selected objects. The information about the products (e.g., images, prices, and descriptions) are automatically imported from the database and rendered in the document in the real-time. The user is able to save the document and return to it later. The document can be saved in the inDesign format or exported to PDF.

1

Period	April 2007 – August 2009
Title	An Efficient Invoicing Solution with Offline Synchronization Capabilities
Role	Project Leader, Principal Software Developer
Project code	Faktury2007
Funding entity	OPA-LABOR, Poland
Budget	4 000.00 EUR
Start date	2007-04-01
End date	2009-08-31
Industrial coordinator	Tadeusz Gruszka, OPA-LABOR, Poland

Details The project was concentrated on creating an invoicing system for a mining company, which will be characterized by a significant fraction of features differing from other systems already existing in the market. These requirements are imposed due to a very specific way how the company works and makes its revenue. The company consists of main headquarters and more than 30 departments in different geographical locations. The tariffs used by the particular departments are different and should be able to be created and entered into the system only in the main headquarters, while both the main headquarters and the departments should be able to use the tariffs for invoicing purposes. Additionally, the departments are allowed to create custom invoices, which are not based on tariffs, but they must be properly marked to be checked into the headquarters. The departments cannot directly print any invoices; this ability is reserved for the headquarters. The departments had only dial-up Internet connection and, therefore, the tariffs and generated invoices needed to be synchronized between the headquarters and departments using small files distributed by e-mails. Additionally, the headquarters needed to have abilities to edit any invoice or to create a memo. The designed and implemented system was in use in around 30 departments of OPA-LABOR during 4 years, successfully satisfying all the requirements set in this project.

Publications

Books

1

Authors Tomasz Bujlow
Title Classification and Analysis of Computer Network Traffic
Pages 1–262
Publisher Networking & Security, Department of Electronic Systems, Aalborg University
Date June 2014
ISBN 978-87-71520-30-9
Accessible Publisher's version (DOI: none) | Author's version (free of charge)
Abstract Traffic monitoring and analysis can be done for multiple different reasons: to investigate the usage of network resources, adjust Quality of Service (QoS) policies in the network, log the traffic to comply with the law, or create realistic models of traffic for academic purposes. The core activity in this area is traffic classification, which is the main topic of this thesis.
We introduced the already known methods for traffic classification (as by using transport layer port numbers, Deep Packet Inspection (DPI), statistical classification) and assessed their usefulness in particular areas. Statistical classifiers based on Machine Learning Algorithms (MLAs) were shown to be accurate and at the same time they do not consume a lot of resources and do not cause privacy concerns. However, they require good quality training data. We performed substantial testing of widely used DPI classifiers and assessed their usefulness in generating ground-truth, which can be used as training data for MLAs. Because the existing methods were shown to not be capable of generating the proper training data, we built our own host-based system for collecting and labeling of network data, which depends on volunteers. Afterwards, we designed and implemented our own system for traffic classification based on various statistical methods, which provides consistent results on all of the 6 levels: Ethernet, IP protocol, application, behavior, content, and service provider. Finally, we contributed to the open source community by improving the accuracy of nDPI traffic classifier. The thesis also evaluates the possibilities of using various traffic classifiers in order to assess the per-application QoS level.

Articles in journals

4

Authors Tomasz Bujlow, Valentín Carela-Español, Josep Solé-Pareta, and Pere Barlet-Ros
Title A Survey on Web Tracking: Mechanisms, Implications, and Defenses
Journal Proceedings of the IEEE
ISSN 0018–9219 (print), 1558–2256 (electronic)
Volume 105
Number 8
Pages 1476–1510
Publisher IEEE
Date March 2017

Accessible Publisher's version (DOI: 10.1109/JPROC.2016.2637878) | Author's version (free of charge)

Abstract Privacy seems to be the Achilles' heel of today's web. Most web services make continuous efforts to track their users and to obtain as much personal information as they can from the things they search, the sites they visit, the people they contact, and the products they buy. This information is mostly used for commercial purposes, which go far beyond targeted advertising. Although many users are already aware of the privacy risks involved in the use of internet services, the particular methods and technologies used for tracking them are much less known. In this survey, we review the existing literature on the methods used by web services to track the users online as well as their purposes, implications, and possible user's defenses. We present five main groups of methods used for user tracking, which are based on sessions, client storage, client cache, fingerprinting, and other approaches. A special focus is placed on mechanisms that use web caches, operational caches, and fingerprinting, as they are usually very rich in terms of using various creative methodologies. We also show how the users can be identified on the web and associated with their real names, e-mail addresses, phone numbers, or even street addresses. We show why tracking is being used and its possible implications for the users. For each of the tracking methods, we present possible defenses. Some of them are specific to a particular tracking approach, while others are more universal (block more than one threat). Finally, we present the future trends in user tracking and show that they can potentially pose significant threats to the users' privacy.

3

Authors Tomasz Bujlow, Valentín Carela-Español, and Pere Barlet-Ros
Title Independent Comparison of Popular DPI Tools for Traffic Classification
Journal Computer Networks
ISSN 1389–1286
Volume 76
Number 0
Pages 75–89
Publisher Elsevier B.V.
Date January 2015
Accessible Publisher's version (DOI: 10.1016/j.comnet.2014.11.001) | Author's version (free of charge)

Abstract Deep Packet Inspection (DPI) is the state-of-the-art technology for traffic classification. According to the conventional wisdom, DPI is the most accurate classification technique. Consequently, most popular products, either commercial or open-source, rely on some sort of DPI for traffic classification. However, the actual performance of DPI is still unclear to the research community, since the lack of public datasets prevent the comparison and reproducibility of their results. This paper presents a comprehensive comparison of 6 well-known DPI tools, which are commonly used in the traffic classification literature. Our study includes 2 commercial products (*PACE* and *NBAR*) and 4 open-source tools (*OpenDPI*, *L7-filter*, *nDPI*, and *Libprotoident*). We studied their performance in various scenarios (including packet and flow truncation) and at different classification levels (application protocol, application and web service). We carefully built a labeled dataset with more than 750 K flows, which contains traffic from popular applications. We used the Volunteer-Based System (VBS), developed at Aalborg University, to guarantee the correct labeling of the dataset. We released this dataset, including full packet payloads, to the research community. We believe this dataset could become a common benchmark for the comparison and validation of network traffic classifiers. Our results present *PACE*, a commercial tool, as the most accurate solution. Surprisingly, we find that some open-source tools, such as *nDPI* and *Libprotoident*, also achieve very high accuracy.

2

Authors Tomasz Bujlow, Sara Ligaard Nørgaard Hald, Tahir Riaz, and Jens Myrup Pedersen
Title A Method for Evaluation of Quality of Service in Computer Networks
Journal ICACT Transactions on the Advanced Communications Technology (ICACT-TACT)
ISSN 2288–0003 (Online)
Volume 1
Number 2
Pages 17–25
Publisher Global IT Research Institute (GiRI)
Date July 2012
Accessible Publisher's version (DOI: none) | Author's version (free of charge)

Abstract Monitoring of the Quality of Service (QoS) in high-speed Internet infrastructures is a challenging task. However, precise assessments must take into account the fact that the requirements for the given quality level are service-dependent. The backbone QoS monitoring and analysis requires processing of large amounts of data and the knowledge about the kinds of applications, which generate the traffic. To overcome the drawbacks of existing methods for traffic classification, we proposed and evaluated a centralized solution based on the C5.0 Machine Learning Algorithm (MLA) and decision rules. The first task was to collect and to provide to C5.0 high-quality training data divided into groups, which correspond to different types of applications. It was found that the currently existing means of collecting data (classification by ports, Deep Packet Inspection, statistical classification, public data sources) are not sufficient and they do not comply with the required standards. We developed a new system to collect the training data, in which the major role is performed by volunteers. Client applications installed on volunteers' computers collect the detailed data about each flow passing through the network interface, together with the application name taken from the description of system sockets. This paper proposes a new method for measuring the level of Quality of Service in broadband networks. It is based on our Volunteer-Based System to collect the training data, Machine Learning Algorithms to generate the classification rules and the application-specific rules for assessing the QoS level. We combine both passive and active monitoring technologies. The paper evaluates different possibilities of the implementation, presents the current implementation of the particular parts of the system, their initial runs and the obtained results, highlighting parts relevant from the QoS point of view.

1

Authors Tomasz Bujlow, Kartheepan Balachandran, Sara Ligaard Nørgaard Hald, Tahir Riaz, and Jens Myrup Pedersen
Title Volunteer-Based System for Research on the Internet Traffic
Journal TELFOR Journal
ISSN 1821–3251 (Print), 2334–9905 (Online)
Volume 4
Number 1
Pages 2–7
Publisher TELFOR
Date September 2012
Accessible Publisher's version (DOI: none) | Author's version (free of charge)
Abstract To overcome the drawbacks of the existing methods for traffic classification (by ports, Deep Packet Inspection, statistical classification), a new system was developed, in which the data are collected and classified directly by clients installed on machines belonging to volunteers. Our approach combines the information obtained from the system sockets, the HTTP content types, and the data transmitted through network interfaces. It allows to group packets into flows and associate them with particular applications or the types of service. This paper presents the design and implementation of our system, the testing phase and the obtained results. The performed threat assessment highlights potential security issues and proposes solutions in order to mitigate the risks. Furthermore, it proves that the system is feasible in terms of uptime and resource usage, assesses its performance and proposes future enhancements. We released the system under *The GNU General Public License v3.0* and published it as a SourceForge project called *Volunteer-Based System for Research on the Internet*.

Conference papers

8

Authors Luca Deri, Maurizio Martinelli, Tomasz Bujlow, and Alfredo Cardigliano
Title nDPI: Open-Source High-Speed Deep Packet Inspection
Publication Proceedings of the 10th International Wireless Communications & Mobile Computing Conference 2014 (IWCMC 2014)
Pages 617–622
Organization IEEE
Place Nicosia, Cyprus
Date August 2014
Accessible Publisher's version (DOI: 10.1109/IWCMC.2014.6906427) | Author's version (free of charge)

Abstract Network traffic analysis was traditionally limited to packet header, because the transport protocol and application ports were usually sufficient to identify the application protocol. With the advent of port-independent, peer-to-peer, and encrypted protocols, the task of identifying application protocols became increasingly challenging, thus creating a motivation for creating tools and libraries for network protocol classification. This paper covers the design and implementation of nDPI, an open-source library for protocol classification using both packet header and payload. nDPI was extensively validated in various monitoring projects ranging from Linux kernel protocol classification, to analysis of 10 Gbit traffic, reporting both high protocol detection accuracy and efficiency.

7

Authors Valentín Carela-Español, Tomasz Bujlow, and Pere Barlet-Ros
Title Is our Ground-Truth for Traffic Classification Reliable?
Publication Proceedings of the 15th Passive and Active Measurement Conference (PAM 2014), Proceedings Series: Lecture Notes in Computer Science 8362
Pages 98–108
Organization Springer International Publishing Switzerland
Place Los Angeles, USA
Date March 2014
Accessible Publisher's version (DOI: 10.1007/978-3-319-04918-2_10) | Author's version (free of charge)
Abstract The validation of the different proposals in the traffic classification literature is a controversial issue. Usually, these works base their results on a ground-truth built from private datasets and labeled by techniques of unknown reliability. This makes the validation and comparison with other solutions an extremely difficult task.
This paper aims to be a first step towards addressing the validation and trustworthiness problem of network traffic classifiers. We perform a comparison between 6 well-known DPI-based techniques, which are frequently used in the literature for ground-truth generation. In order to evaluate these tools we have carefully built a labeled dataset of more than 500 000 flows, which contains traffic from popular applications. Our results present *PACE*, a commercial tool, as the most reliable solution for ground-truth generation. However, among the open-source tools available, *NDPI* and especially *Libprotoident*, also achieve very high precision, while other, more frequently used tools (e.g., *L7-filter*) are not reliable enough and should not be used for ground-truth generation in their current form.

6

Authors Tomasz Bujlow and Jens Myrup Pedersen
Title Obtaining Application-Based and Content-Based Internet Traffic Statistics
Publication Proceedings of the 6th International Conference on Signal Processing and Communication Systems (ICSPCS'12)
Pages 1–10
Organization IEEE
Place Gold Coast, Queensland, Australia
Date December 2012
Accessible Publisher's version (DOI: 10.1109/ICSPCS.2012.6507984) | Author's version (free of charge)
Abstract Understanding Internet traffic is crucial in order to facilitate the academic research and practical network engineering, e.g. when doing traffic classification, prioritization of traffic, creating realistic scenarios and models for Internet traffic development etc. In this paper, we demonstrate how the Volunteer-Based System for Research on the Internet, developed at Aalborg University, is capable of providing detailed statistics of Internet usage. Since an increasing amount of HTTP traffic has been observed during the last few years, the system also supports creating statistics of different kinds of HTTP traffic, like audio, video, file transfers, etc. All statistics can be obtained for individual users of the system, for groups of users, or for all users altogether. This paper presents results with real data collected from a limited number of real users over six months. We demonstrate that the system can be useful for studying the characteristics of computer network traffic in application-oriented or content-type-oriented way, and is now ready for a larger-scale implementation. The paper is concluded with a discussion about various applications of the system and the possibilities of further enhancements.

5

Authors Jens Myrup Pedersen and Tomasz Bujlow
Title Obtaining Internet Flow Statistics by Volunteer-Based System
Publication Proceedings of the Fourth International Conference on Image Processing & Communications (IP&C 2012), Image Processing & Communications Challenges 4, AISC 184
Pages 261–268
Organization Springer Berlin Heidelberg
Place Bydgoszcz, Poland

Date September 2012
Accessible Publisher's version (DOI: 10.1007/978-3-642-32384-3_32) | Author's version (free of charge)
Abstract In this paper, we demonstrate how the Volunteer Based System for Research on the Internet, developed at Aalborg University, can be used for creating statistics of Internet usage. Since the data are collected on individual machines, the statistics can be made on the basis of both individual users and groups of users, and as such be useful also for segmentation of the users into groups. We present results with data collected from real users over several months; in particular we demonstrate how the system can be used for studying flow characteristics - the number of TCP and UDP flows, average flow lengths, and average flow durations. The paper is concluded with a discussion on what further statistics can be made, and the further development of the system.

4

Authors Tomasz Bujlow, Tahir Riaz, and Jens Myrup Pedersen
Title Classification of HTTP Traffic Based on C5.0 Machine Learning Algorithm
Publication Proceedings of the Fourth IEEE International Workshop on Performance Evaluation of Communications in Distributed Systems and Web-based Service Architectures (PEDISWESA 2012)
Pages 882–887
Organization IEEE
Place Cappadocia, Turkey
Date July 2012
Accessible Publisher's version (DOI: 10.1109/ISCC.2012.6249413) | Author's version (free of charge)
Abstract Our previous work demonstrated the possibility of distinguishing several kinds of applications with accuracy of over 99 %. Today, most of the traffic is generated by web browsers, which provide different kinds of services based on the HTTP protocol: web browsing, file downloads, audio and voice streaming through third-party plugins, etc. This paper suggests and evaluates two approaches to distinguish various types of HTTP content: distributed among volunteers' machines and centralized running in the core of the network. We also assess the accuracy of the global classifier for both HTTP and non-HTTP traffic. We achieved accuracy of 94 %, which supposed to be even higher in real-life usage. Finally, we provided graphical characteristics of different kinds of HTTP traffic.

3

Authors Tomasz Bujlow, Tahir Riaz, and Jens Myrup Pedersen
Title A Method for Assessing Quality of Service in Broadband Networks
Publication Proceedings of the 14th International Conference on Advanced Communication Technology (ICACT)
Pages 826–831
Organization IEEE
Place Phoenix Park, PyeongChang, Korea
Date February 2012
Accessible Publisher's version (DOI: none) | Author's version (free of charge)
Abstract Monitoring of Quality of Service (QoS) in high-speed Internet infrastructure is a challenging task. However, precise assessments must take into account the fact that the requirements for the given quality level are service-dependent. Backbone QoS monitoring and analysis requires processing of large amount of the data and knowledge of which kind of application the traffic belongs to. To overcome the drawbacks of existing methods for traffic classification we proposed and evaluated a centralized solution based on C5.0 Machine Learning Algorithm (MLA) and decision rules. The first task was to collect and provide C5.0 high-quality training data, divided into groups corresponding to different types of applications. It was found that currently existing means of collecting data (classification by ports, Deep Packet Inspection, statistical classification, public data sources) are not sufficient and they do not comply with the required standards. To collect training data a new system was developed, in which the major role is performed by volunteers. Client applications installed on their computers collect the detailed data about each flow passing through the network interface, together with the application name taken from the description of system sockets. This paper proposes a new method for measuring the Quality of Service (QoS) level in broadband networks, based on our Volunteer-Based System for collecting the training data, Machine Learning Algorithms for generating the classification rules and application-specific rules for assessing the QoS level. We combine both passive and active monitoring technologies. The paper evaluates different implementation possibilities, presents the current implementation of particular parts of the system, their initial runs and obtained results, highlighting parts relevant from the QoS point of view.

2

Authors Tomasz Bujlow, Tahir Riaz, and Jens Myrup Pedersen
Title A Method for Classification of Network Traffic Based on C5.0 Machine Learning Algorithm

Publication Proceedings of ICNC'12: 2012 International Conference on Computing, Networking and Communications (ICNC): Workshop on Computing, Networking and Communications
Pages 244–248
Organization IEEE
Place Maui, Hawaii, USA
Date February 2012
Accessible Publisher's version (DOI: 10.1109/ICCNC.2012.6167418) | Author's version (free of charge)
Abstract Monitoring of the network performance in a high-speed Internet infrastructure is a challenging task, as the requirements for the given quality level are service-dependent. Therefore, the backbone QoS monitoring and analysis in Multi-hop Networks requires the knowledge about the types of applications forming the current network traffic. To overcome the drawbacks of existing methods for traffic classification, usage of C5.0 Machine Learning Algorithm (MLA) was proposed. On the basis of the statistical traffic information received from volunteers and C5.0 algorithm, we constructed a boosted classifier, which was shown to have the ability to distinguish between 7 different applications in the test set of 76,632–1,622,710 unknown cases with average accuracy of 99.3–99.9 %. This high accuracy was achieved by using high quality training data collected by our system, a unique set of parameters used for both training and classification, an algorithm for recognizing flow direction and the C5.0 itself. The classified applications include Skype, FTP, torrent, web browser traffic, web radio, interactive gaming and SSH. We performed subsequent tries using different sets of parameters and both training and classification options. This paper shows how we collected accurate traffic data, presents arguments used in classification process, introduces the C5.0 classifier and its options, and finally, evaluates and compares the obtained results.

1 **Authors** Tomasz Bujlow, Kartheepan Balachandran, Tahir Riaz, and Jens Myrup Pedersen
Title Volunteer-Based System for Classification of Traffic in Computer Networks
Publication Proceedings of the 19th Telecommunications Forum TELFOR 2011
Pages 210–213
Organization IEEE
Place Belgrade, Serbia
Date November 2011
Accessible Publisher's version (DOI: 10.1109/TELFOR.2011.6143528) | Author's version (free of charge)
Abstract To overcome the drawbacks of existing methods for traffic classification (by ports, Deep Packet Inspection, statistical classification) a new system was developed, in which the data are collected from client machines. This paper presents design of the system, implementation, initial runs and obtained results. Furthermore, it proves that the system is feasible in terms of uptime and resource usage, assesses its performance and proposes future enhancements.

Technical reports

4 **Authors** Tomasz Bujlow, Valentín Carela-Español, Josep Solé Pareta, and Pere Barlet-Ros
Title Web Tracking: Mechanisms, Implications, and Defenses
Pages 1–29
Publisher arXiv.org: Computer Science – Computers and Society
Date July 2015
Accessible Publisher's version (DOI: none) | Author's version (free of charge)

Abstract This articles surveys the existing literature on the methods currently used by web services to track the user online as well as their purposes, implications, and possible user's defenses. A significant majority of reviewed articles and web resources are from years 2012 – 2014. Privacy seems to be the Achilles' heel of today's web. Web services make continuous efforts to obtain as much information as they can about the things we search, the sites we visit, the people with who we contact, and the products we buy. Tracking is usually performed for commercial purposes. We present 5 main groups of methods used for user tracking, which are based on sessions, client storage, client cache, fingerprinting, or yet other approaches. A special focus is placed on mechanisms that use web caches, operational caches, and fingerprinting, as they are usually very rich in terms of using various creative methodologies. We also show how the users can be identified on the web and associated with their real names, e-mail addresses, phone numbers, or even street addresses. We show why tracking is being used and its possible implications for the users. For example, we describe recent cases of price discrimination, assessing financial credibility, determining insurance coverage, government surveillance, and identity theft. For each of the tracking methods, we present possible defenses. Some of them are specific to a particular tracking approach, while others are more universal (block more than one threat) and they are discussed separately. Apart from describing the methods and tools used for keeping the personal data away from being tracked, we also present several tools that were used for research purposes – their main goal is to discover how and by which entity the users are being tracked on their desktop computers or smartphones, provide this information to the users, and visualize it in an accessible and easy to follow way. Finally, we present the currently proposed future approaches to track the user and show that they can potentially pose significant threats to the users' privacy.

3

Authors Tomasz Bujlow and Jens Myrup Pedersen
Title A Practical Method for Multilevel Classification and Accounting of Traffic in Computer Networks
Pages 1–56
Publisher Department of Electronic Systems, Aalborg University
Date February 2014
Accessible Publisher's version (DOI: none) | Author's version (free of charge)
Abstract Existing tools for traffic classification are shown to be incapable of identifying the traffic in a consistent manner. For some flows only the application is identified, for others only the content, for yet others only the service provider. Furthermore, Deep Packet Inspection is characterized by extensive needs for resources and privacy or legal concerns. Techniques based on Machine Learning Algorithms require good quality training data, which are difficult to obtain. They usually cannot properly deal with other types of traffic, than they are trained to work with, and they are unable to detect the content carried by the flow, or the service provider. To overcome the drawbacks of already existing methods, we developed a novel hybrid method to provide accurate identification of computer network traffic on six levels: Ethernet, IP protocol, application, behavior, content, and service provider. Our system built based on the method provides also traffic accounting and it was tested on 2 datasets. We have shown that our system gives a consistent, accurate output on all the levels. We also showed that the results provided by our system on the application level outperformed the results obtained from the most commonly used DPI tools.

2

Authors Tomasz Bujlow, Valentín Carela-Español, and Pere Barlet-Ros
Title Extended Independent Comparison of Popular Deep Packet Inspection (DPI) Tools for Traffic Classification
Pages 1–440
Publisher Department of Computer Architecture (DAC), Universitat Politècnica de Catalunya (UPC)
Date January 2014
Accessible Publisher's version (DOI: none) | Author's version (free of charge)
Abstract Network traffic classification became an essential input for many network-related tasks. However, the continuous evolution of the Internet applications and their techniques to avoid being detected (as dynamic port numbers, encryption, or protocol obfuscation) considerably complicated their classification. We start the report by introducing and shortly describing several well-known DPI tools, which later will be evaluated: PACE, OpenDPI, L7-filter, NDPI, Libprotoident, and NBAR.
This report has several major contributions. At first, by using VBS, we created 3 datasets of 17 application protocols, 19 applications (also various configurations of the same application), and 34 web services, which are available to the research community. The first dataset contains full flows with entire packets, the second dataset contains truncated packets (the Ethernet frames were overwritten by 0s after the 70th byte), and the third dataset contains truncated flows (we took only 10 first packets for each flow). The datasets contain 767 690 flows labeled on a multidimensional level. These datasets are available as a bunch of PCAP files containing full flows including the packet payload, together with corresponding text files, which describe the flows in the order as they were originally captured and stored in the PCAP files.

At second, we developed a method for labeling non-HTTP flows, which belong to web services (as YouTube). Labeling based on the corresponding domain names taken from the HTTP header could allow to identify only the HTTP flows. Other flows (as encrypted SSL / HTTPS flows, RTMP flows) are left unlabeled. Therefore, we implemented a heuristic method for detection of non-HTTP flows, which belong to the specific services. Then, we examined the ability of the DPI tools to accurately label the flows included in our datasets.

1

Authors	Tomasz Bujlow, Valentín Carela-Español, and Pere Barlet-Ros
Title	Comparison of Deep Packet Inspection (DPI) Tools for Traffic Classification
Pages	1–108
Publisher	Department of Computer Architecture (DAC), Universitat Politècnica de Catalunya (UPC)
Date	June 2013
Accessible	Publisher's version (DOI: none) | Author's version (free of charge)
Abstract	Nowadays, there are many tools, which are being able to classify the traffic in computer networks. Each of these tools claims to have certain accuracy, but it is a hard task to asses which tool is better, because they are tested on various datasets. Therefore, we made an approach to create a dataset, which can be used to test all the traffic classifiers. In order to do that, we used our system to collect the complete packets from the network interfaces. The packets are grouped into flows, and each flow is collected together with the process name taken from Windows / Linux sockets, so the researchers do not only have the full payloads, but also they are provided the information which application created the flow. Therefore, the dataset is useful for testing Deep Packet Inspection (DPI) tools, as well as statistical, and port-based classifiers. The dataset was created in a fully manual way, which ensures that all the time parameters inside the dataset are comparable with the parameters of the usual network data of the same type. The system for collecting of the data, as well as the dataset, are made available to the public. Afterwards, we compared the accuracy of classification on our dataset of PACE, OpenDPI, NDPI, Libprotoident, NBAR, four different variants of L7-filter, and a statistic-based tool developed at UPC. We performed a comprehensive evaluation of the classifiers on different levels of granularity: application level, content level, and service provider level. We found out that the best performing classifier on our dataset is PACE. From the non-commercial tools, NDPI and Libprotoident provided the most accurate results, while the worst accuracy we obtained from all 4 versions of L7-filter.

Other scientific contributions

Presentations in seminars

12

Role	Co-author and Participant
Topic	User Tracking Uncovered (Tracking Catalog: Uncovering and analyzing user tracking on the Internet)
Event	Data Transparency Lab (DTL) Launch Workshop
Place	Telefonica, Barcelona, Spain
Date	November 2014
Accessible	http://www.datatransparencylab.org

11

Role	Author and Presenter
Topic	Consistency, Accuracy, and Usefulness of Techniques and Tools for Network Traffic Identification
Event	Seminar organized by the Networks, Systems, Services, and Security (R3S) research team from the Distributed Services, Architectures, Modelling, Validation, and Network Administration (SAMOVAR) research unit
Place	TELECOM Sudparis, Evry, France
Date	May 2014
Accessible	http://samovar.telecom-sudparis.eu/spip.php?article779

10

Role	Author and Presenter
Topic	Obtaining Useful Classification Results by Deep Packet Inspection (DPI)
Event	Complements of Network Management (SGR) course for the 8th semester student group from the specialty of Computer Science
Place	Computer Science Department, University of Pisa, Pisa, Italy

Date April 2014

9

Role Author and Presenter
Topic Usefulness of the Results – a Forgotten Evaluation Metric of Traffic Identification Tools
Event Seminar organized by the Telecommunication Networks Group
Place Department of Electronics and Telecommunications, Polytechnic University of Turin, Turin, Italy
Date April 2014

8

Role Author and Presenter
Topic Advanced Network Traffic Monitoring & Analysis
Event Communication Networks and Ambient Intelligence course for the 7th semester student group from the specialty of Network and Distributed Systems
Place Department of Electronic Systems, Aalborg University, Aalborg, Denmark
Date September 2013

7

Role Author and Presenter
Topic Quality of Service (QoS) Assessment in Computer Networks
Event Second IntelliCIS Training School on Simulation-based design of Complex Infrastructure Systems
Organizer COST Action IC0806: Intelligent Monitoring, Control and Security of Critical Infrastructure Systems (IntelliCIS)
Place RWTH University, Aachen, Germany
Date March 2013
Accessible http://www.intellicis.eu/Pages/Training_Schools.php

6

Role Author and Presenter
Topic Traffic Monitoring and Analysis – Advanced Techniques Based on Machine Learning
Event Seminar on Traffic Monitoring and Analysis
Place Department of Computer Architecture, Universitat Politècnica de Catalunya, Barcelona, Spain
Date November 2012

5

Role Author and Presenter
Topic Classification of Traffic Using Machine Learning Techniques
Event Communication Networks and Ambient Intelligence course for the 7th semester student group from the specialty of Network and Distributed Systems
Place Department of Electronic Systems, Aalborg University, Aalborg, Denmark
Date October 2012

4

Role Author and Presenter
Topic Advanced Network Traffic Analysis
Event Life Long Learning course for external participants
Place Aalborg University, Aalborg, Denmark
Date August 2012

3

Role Author and Presenter
Topic Advanced End-User Traffic Monitoring
Event Internet Quality – More Than Bandwidth, an international industrial conference
Organizer Collaborating Living Labs (COLL) project: Compare Testlab – Karlstad University, NettOp – University of Stavanger, and CNP – Aalborg University
Place IDA House of Engineers, Copenhagen, Denmark
Date June 2012
Accessible https://mit.ida.dk/IDAforum/u0631a/Documents/Internet%20kvalitet%20-%2018-06-2012/Tomasz%20Bujlow.pdf

2

Role Author and Presenter

Topic Volunteer-based System for Classification of Traffic in Computer Networks
Event First IntelliCIS Training School on Intelligent Monitoring of Critical Infrastructures
Organizer COST Action IC0806: Intelligent Monitoring, Control and Security of Critical Infrastructure Systems (IntelliCIS)
Place Albena Resort, Bulgaria
Date October 2011
Accessible http://www.intellicis.eu/Pages/Training_Schools.php

1
Role Author and Presenter
Topic Classification of Traffic in Integrated Computer Networks
Event Life Long Learning course for external participants
Place Aalborg University, Aalborg, Denmark
Date August 2011

Reviews of journal articles and conference papers

15
Publication Journal of Cyber Security Technology
Publisher Taylor & Francis Group
Type Article in a journal
Date February 2018

14
Publication Wireless Communications and Mobile Computing
Publisher Hindawi
Type Article in a journal
Date February 2018

13
Publication IEEE Communications Letters (IEEE COMML)
Publisher IEEE
Type Article in a journal
Date February 2017

12
Publication IEEE Communications Letters (IEEE COMML)
Publisher IEEE
Type Article in a journal
Date July 2016

11
Publication SoftwareX
Publisher Elsevier
Type Article in a journal
Date May 2016

10
Publication Computer Communications (COMCOM)
Publisher Elsevier
Type Article in a journal
Date March 2016

9
Publication IEEE Transactions on Network and Service Management (TNSM)
Publisher IEEE
Type Article in a journal
Date May 2015

8
Publication IEEE Transactions on Network and Service Management (TNSM)

Publisher IEEE
Type Article in a journal
Date October 2014

7 **Publication** IEEE Transactions on Network and Service Management (TNSM)
Publisher IEEE
Type Article in a journal
Date April 2014

6 **Publication** Scientia Iranica
Publisher Sharif University of Technology
Type Article in a journal
Date December 2013

5 **Publication** Proceedings of the 21th Telecommunications Forum (TELFOR 2013)
Organization Telecommunications Society, Belgrade; School of Electrical Engineering, University of Belgrade; IEEE Serbia; Montenegro COM Chapter
Type Conference paper
Date October 2013

4 **Publication** Proceedings of the 20th Telecommunications Forum (TELFOR 2012)
Organization Telecommunications Society, Belgrade; School of Electrical Engineering, University of Belgrade; IEEE Serbia; Montenegro COM Chapter
Type Conference paper
Date October 2012

3 **Publication** Proceedings of the 2012 International Conference on Computing, Networking and Communications (ICNC'12)
Type Conference paper
Date September 2011

2 **Publication** Zeszyty Naukowe. Telekomunikacja i Elektronika
Publisher University of Technology and Life Sciences in Bydgoszcz
Type Article in a journal
Date May 2011

1 **Publication** Computer Standards & Interfaces
Publisher Elsevier B.V.
Type Article in a journal
Date March 2011

I declare that I agree to have my personal data, if it necessary, processed for the recruitment process