Stanisław Kozielski
Dariusz Mrozek
Paweł Kasprowski
Bożena Małysiak-Mrozek
Daniel Kostrzewa (Eds.)

Communications in Computer and Information Science 613

Beyond Databases, Architectures and Structures

Advanced Technologies for Data Mining and Knowledge Discovery

12th International Conference, BDAS 2016
Ustroń, Poland, May 31 – June 3, 2016
Proceedings

BDAS

Springer

Communications in Computer and Information Science 613

Commenced Publication in 2007
Founding and Former Series Editors:
Alfredo Cuzzocrea, Dominik Ślęzak, and Xiaokang Yang

Stanisław Kozielski · Dariusz Mrozek
Paweł Kasprowski · Bożena Małysiak-Mrozek
Daniel Kostrzewa (Eds.)

Beyond Databases, Architectures and Structures

Advanced Technologies for Data Mining and Knowledge Discovery

12th International Conference, BDAS 2016
Ustroń, Poland, May 31 – June 3, 2016
Proceedings

Springer

Editors
Stanisław Kozielski
Institute of Informatics
Silesian University of Technology
Gliwice
Poland

Dariusz Mrozek
Institute of Informatics
Silesian University of Technology
Gliwice
Poland

Paweł Kasprowski
Institute of Informatics
Silesian University of Technology
Gliwice
Poland

Bożena Małysiak-Mrozek
Institute of Informatics
Silesian University of Technology
Gliwice
Poland

Daniel Kostrzewa
Institute of Informatics
Silesian University of Technology
Gliwice
Poland

ISSN 1865-0929 ISSN 1865-0937 (electronic)
Communications in Computer and Information Science
ISBN 978-3-319-34098-2 ISBN 978-3-319-34099-9 (eBook)
DOI 10.1007/978-3-319-34099-9

Library of Congress Control Number: 2016938385

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Preface

Collecting, processing, and analyzing data have become important branches of computer science. Many areas of our existence generate a wealth of information that must be stored in a structured manner and processed appropriately in order to gain the knowledge from the inside. Databases have become a ubiquitous way of collecting and storing data. They are used to hold data describing many areas of human life and activity, and as a consequence, they are also present in almost every IT system. Today's databases have to face the problem of data proliferation and growing variety. More efficient methods for data processing are needed more than ever. New areas of interests that deliver data require innovative algorithms for data analysis.

Beyond Databases, Architectures and Structures (BDAS) is a series of conferences located in Central Europe and very important for this geographic region. The conference intends to give the state of the art of the research that satisfies the needs of modern, widely understood database systems, architectures, models, structures, and algorithms focused on processing various types of data. The aim of the conference is to reflect the most recent developments of databases and allied techniques used for solving problems in a variety of areas related to database systems, or even go one step forward — beyond the horizon of existing databases, architectures, and data structures.

The 12th International BDAS Scientific Conference (BDAS 2016), held in Ustroń, Poland, during May 31–June 3, 2016, was a continuation of the highly successful BDAS conference series started in 2005. For many years BDAS has been attracting hundreds or even thousands of researchers and professionals working in the field of databases. Among attendees of our conference were scientists and representatives of IT companies. Several editions of BDAS were supported by our commercial, world-renowned partners, developing solutions for the database domain, such as IBM, Microsoft, Oracle, Sybase, and others. BDAS annual meetings have become an arena for exchanging information on the widely understood database systems and data-processing algorithms.

BDAS 2016 was the 12th edition of the conference, organized under the technical co-sponsorship of the IEEE Poland Section. We also continued our successful cooperation with Springer, which resulted in the publication of this book. The conference attracted more than 100 participants from 15 countries, who made this conference a successful and memorable event. There were five keynote talks given by leading scientists: Prof. Jarek Gryz from the Department of Computer Science and Engineering, York University, Toronto, Canada, had a keynote talk on "Interactive Visualization of Big Data." Prof. Abdelkader Hameurlain from Pyramid Team, Institut de Recherche en Informatique de Toulouse IRIT, Paul Sabatier University, Toulouse Cedex, France, gave an interesting lecture entitled "Big Data Management in the Cloud: Evolution or Crossroad?" Prof. Dirk Labudde from Bioinformatics group Mittweida (bigM) and Forensic Science Investigation Lab (FoSIL), University of Applied Sciences, Mittweida, Germany, gave an excellent talk entitled "From Data to Evidence: Digital

Forensic Analyses of Communication Networks." Prof. Jean-Charles Lamirel from SYNALP team, LORIA, Vandoeuvre-lès-Nancy, France, had a very enlightening speech on "Performing and Visualizing Temporal Analysis of Large Text Data Issued for Open Sources: Past and Future Methods." Prof. Zbigniew W. Raś from the Department of Computer Science, University of North Carolina, Charlotte, USA, honored us with a presentation on "Reduction of Readmissions to Hospitals Based on Actionable Knowledge Mining and Personalization." The keynote speeches and plenary sessions allowed participants to gain insight into new areas of data analysis and data processing.

BDAS is focused on all aspects of databases. It is intended to have a broad scope, including different kinds of data acquisition, processing, and storing, and this book reflects fairly well the large span of research presented at BDAS 2016. This volume consists of 57 carefully selected papers. The first four papers accompany the stunning keynote talks. The remainder of the papers are assigned to seven thematic groups:

- Artificial intelligence, data mining, and knowledge discovery
- Architectures, structures, and algorithms for efficient data processing
- Data warehousing and OLAP
- Natural language processing, ontologies, and Semantic Web
- Bioinformatics and biomedical data analysis
- Data processing tools
- Novel applications of database systems

The first group, containing eight papers, is related to various methods used in data mining and knowledge discovery. Papers assembled in this group show a wide spectrum of applications of various exploration methods, like decision rules, knowledge-based and neuro-fuzzy systems, clustering and memetic algorithms, rough sets, to solve many real problems. The second group contains 11 papers devoted to database architectures, structures, and algorithms used for efficient data processing. Papers in this group discuss hot topics of effectiveness of query execution, Big Data, testing performance of various database systems, NoSQL, scalability, task scheduling in processing data in OLTP systems, and in-memory, cloud, and probabilistic databases. The next group of papers concerns issues related to data integration, data warehousing, and OLAP. The group consists of three papers presenting research devoted to the scalability of extraction, transformation and load processes, novel data integration architectures, and spatiotemporal OLAP queries. The fourth group consists of five papers devoted to natural language processing, text mining, ontologies, and the Semantic Web. These papers discuss problems of extraction of concepts from text, mapping semantic features to words, text classification, building ontology for underutilized crops, and querying large RDF data with GPUs. The research devoted to bioinformatics and biomedical data analysis is presented in six papers gathered in the fifth group.

The next group of papers is focused on different data processing tools. It presents tools for content modeling, multitenant applications, frameworks for Big Data and biometric identification, as well as benchmarks and simulators constructed by authors.

The last group consisting of ten papers introduces novel applications for which database systems proved to be useful. Some examples include: water demand forecasting, combat identification, drug abuse extraction, hand pose recognition, or methane concentration value prediction.

We hope that the broad scope of topics related to databases covered in this proceedings volume will help the reader to understand that databases have become an important element of nearly every branch of computer science.

We would like to thank all Program Committee members and additional reviewers for their effort in reviewing the papers. Special thanks to Piotr Kuźniacki — builder and for 12 years administrator of our website: www.bdas.pl. The conference organization would not have been possible without the technical staff: Dorota Huget and Jacek Pietraszuk.

April 2016

Stanisław Kozielski
Dariusz Mrozek
Pawel Kasprowski
Bożena Małysiak-Mrozek
Daniel Kostrzewa

Contents

Invited Papers

Artifical Intelligence, Data Mining and Knowlege Discovery

Architectures, Structures and Algorithms for Efficient Data Processing

Data Warehousing and OLAP

Natural Language Processing, Ontologies and Semantic Web

Bioinformatics and Biomedical Data Analysis

Data Processing Tools

Novel Applications of Database Systems

AI Implementation in Military Combat Identification – A Practical Solution

Łukasz Apiecionek[1], Wojciech Makowski[2](✉), and Marcin Woźniak[2]

[1] Institute of Technology, Casimir the Great University in Bydgoszcz,
ul. Chodkiewicza 30, 85-064 Bydgoszcz, Poland
lapiecionek@ukw.edu.pl

[2] TELDAT Sp. z o. o.sp.k., Bydgoszcz, Poland
{wmakowski,mwozniak}@teldat.com.pl

Abstract. This paper presents the architecture of a communication system which was implemented in MiG-29 airplanes. This system provides a continuous on-line access to the situational awareness information which is necessary for the pilot. The interoperability of this system with other NATO systems allows to collect and transfer data between them. Artificial Intelligence methods are used to implement and improve this system. This modification enables the system to work faster and increases the situational awareness of the pilot on the battlefield.

Keywords: CID · Security · Network · Artificial intelligence

1 Introduction

The Polish Air Force has been undergoing intense changes since several years. During this time, the Air Force was modernized through investments in modern technologies, among others by acquiring brand new F-16 aircrafts and introducing numerous other changes which generate many powerful capabilities. Besides the investments in modernity, withdrawing older generation aircraft decreased the number of combat units. New machines, equipped with the systems compatible with NATO standards, allow the connection of the pilot and the aircraft to a network-centric system, present on todays battlefield. The use of these applications allows among others to exchange data with the tactical situation Link-16, which provides constant on-line access to information and increases the pilots situational awareness. This impacts directly the increase of the efficiency of the pilot, helps to finish the task and allows the pilot to return safely to the home airport.

The Polish Air Force still possesses aircrafts such as MiG-29, which are not interoperable with the aforementioned protocols due totheir level of technology and equipment. In order to restore the interoperability of this kind of aircraft, there is a possibility to perform complicated system changes.

Piloting of the combat airplane can be very stressful. Decisions need to be made quickly by the pilot during the flight. Possessing all the information which can help to complete his tasks is essential. The solution to this problem is to implement a Command Support System (CSS) to these aircrafts. This article features the elaborated system.

S. Kozielski et al. (Eds.): BDAS 2016, CCIS 613, pp. 659–667, 2016.
DOI: 10.1007/978-3-319-34099-9_51

2 CID Architecture

2.1 Template Selection

The aforementioned airplanes use older technology for military aviation that are used nowadays by the Polish Air Force. MiG-29 in the number of more than 30 units, can successfully fulfill provided tasks for a long time. These aircraft have a great combat value, and with the F-16 they form the core of the Polish combat aviation. For many years these aircraft were modernized and upgraded, but only half of them have recently undergone a comprehensive exchange of avionics, which moved them closer to the Western standards and allowed the cooperation in NATO structures [1]. The second half of the aircraft is still waiting for modernization. The planned upgrade includes increasing the pilots situational awareness on the modern battlefield. One of the improvements is the installation of liquid crystal displays, which, among other features also have the ability of viewing digital maps with the tactical situation. It is applied on the basis of the information provided by the new mission computer. However, at present, the data are entered into the computer before take-off and are not updated during the mission in the air. It is not possible for the pilot to build the current situational awareness.

2.2 System Requirements

Modern battlefield situation changes are extremely dynamical. These changes apply to the position of air and ground targets, as well as own troops. Transfer of such a large amount of dynamically changing data via the audio channel, as it is currently the case for the aircraft without implemented Combat Support System solutions, is archaic and does not meet the basic requirements of the modern battlefield. The pilot must receive updated information about the tactical situation regularly, in a transparent manner and in a way which does not distract him. This is very important for the effectiveness of eliminating the enemy targets, but also for preventing the accidents of own troops.

2.3 Information Source

The methods for obtaining the information about the situation from different sources on the battlefield make it necessary to implement many protocols used nowadays for data exchange standards. However, it is possible to use the COTS type finished product (commercial off-the-shelf - ready, working product off the shelf). One of these products that offers a very wide range of supported functionality is the Network Centric Data Communication Platform JASMINE, developed by TELDAT, a Polish company. This kind of system provides many unique internationally interoperable capabilities. It is a network-centric platform which enables the users to access a range of data, collected using different kinds of protocols. The JASMINE platform:

- supports (including automatization) processes of the command and management of troops at all levels, including a single soldier;
- enables achieving the information superiority and thereby creates the situational awareness of troops;
- provides possibility for building Common Operational Picture;
- substantially increases the security of military components and their elements, including soldiers and vehicles;
- creates modern, efficient, scalable, mobile and cheap multi-service ICT infrastructure, which enables construction of many independent networks for command center at operational and tactical level.

The system architecture proposed in this article is based largely on the solutions presented by the NATO STANAG ADatP-37 document (NATO Standard For Services To Forward Friendly Force Information To Weapon Delivery Assets). This standard sets guidelines for CID (Combat Identification) system class. High-level architecture of the system is presented in the Fig. 1.

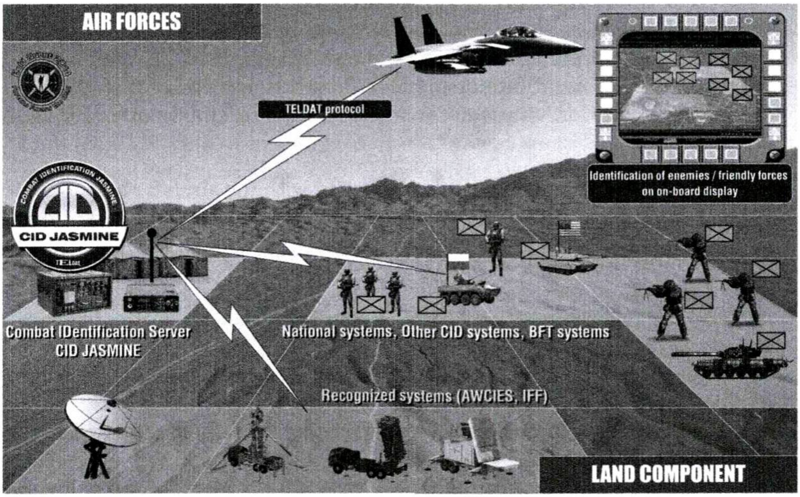
AIR FORCES
TELDAT protocol
Identification of enemies / friendly forces on on-board display
CID JASMINE
Combat IDentification Server CID JASMINE
National systems, Other CID systems, BFT systems
Recognized systems (AWCIES, IFF)
LAND COMPONENT

Fig. 1. System architecture [5]

The solution proposed by NATO was suitably modified considering the characteristics of the aircraft without implemented on-line protocols, among other things, the current restrictions on the transmission and presentation of data. The most important difference, compared to the original concept of NATO, is the introduction of a dedicated protocol and the data exchange medium between the aircraft and the CID JASMINE system [11]. Thanks to this protocol and medium, that aircrafts will receive the same information as the other objects in the air, supported by Link-16 data exchange protocol [12]. In addition, it will be possible to transfer the results from the analysis of information from reconnaissance systems. This functionality will be offered by JASMINE CID System improved under this project. The primary task of the system is to provide information about the current position of enemy and friendly forces, which is implemented as a series of the following stages:

- stage I - gathering information;
- stage II - information analysis;
- stage III - information passing to aircraft.

On stage I information is gathered in the system. The system uses the available protocols and data exchange means to acquire information about the current position of friendly forces and the location of the targets. Reports can be obtained from many sources, mainly from:

- sensors placed directly in the system, such as system IFF (Identification Friendly Foe) components;
- BFT (Blue Force Tracking) class systems, identifying friendly ground forces;
- systems of situational awareness and recognition of AWCIES class;
- command support systems of allied and own troops;
- joint MIP (Multilateral Interoperability Programme) database;
- JC3IEDM (Joint Consultation, Command and Control Information Exchange Data Model).

On stage II, the collected data must be processed with the aircraft characteristics and the earlier assumptions. The system is directed using the following rules:

- only the most important targets for the current run of the mission are transmitted to each aircraft;
- on each airplane the position of the allied troops that are in the immediate vicinity is presented in the first place;
- as the aircraft approaches it receives the first information about the position of allied forces in its vicinity;
- location reports are periodically supplemented with textual information provided to the pilot.

On stage III, the properly prepared data is transferred to the radio stations and received by the on-board radio telemetry channel.

Thus, after appropriate processing, the signals are transmitted to the mission computer, which after decoding it, sends the information to a display in the cockpit where it is presented on the background of a digital map display. Periodically sent text data is transmitted from the computer to the missions speech synthesizer and played there. The effect of the system work is the presentation of the current situation image on the pilots screen in real time.

2.4 System Security

One of the most important aspects of the proper operation of the presented system is the adequate protection of the whole process of data acquisition and transmission. The system can also operate on the interconnection of different classification levels. In that case it will need to use separating safety gates [4,6,7]. The process of the data transfer to the aircraft must also be protected. It is assumed to use the following mechanism [3]:

- authorization information based on a unique data type known only to the sender and the recipient;
- obfuscation of location information through the use of predetermined reference points;
- encryption of the transmitted information.

3 Data Acquiring Platform

The elements of the JASMINE System serve as a platform for data collection. This solution provides the standards and protocols, such as NFFI, FFI-XML-MTF, Link-16, VMF, including data acquired from IFF combat identification sensors, which, despite they seem to be useful both on land and in the air, do not enable the use of information from all these systems at the same time. NFFI, Link-16, VMF have been approved by NATO and are used, among others, during joint operations in Afghanistan. These protocols are very useful, but can be applied only in a few areas and scenarios. In order to improve, especially when the air-land operations are conducted, NATO's Combat Identification Server (CID Server) has been introduced, which collects information from various sources, such as:

- CID sensors such e.g. own enemy;
- BFT (Blue Force Tracking) position monitoring systems;
- situational awareness systems (SA).

CID is the process of obtaining a reliable, accurate view and the attributes of entities in the operations area in order to ensure the real use of tactical possibilities and weapons resources. JASMINE CID Server is an implementation of the NATOs conceptNATO in this area.

In order to provide a precise support, CID System has to be supplied with information from multiple of reliable sources. In this case, the data are delivered with the JASMINE system C3 components and database. This solution uses, among others:

- MIP database
- JC3IEDM model

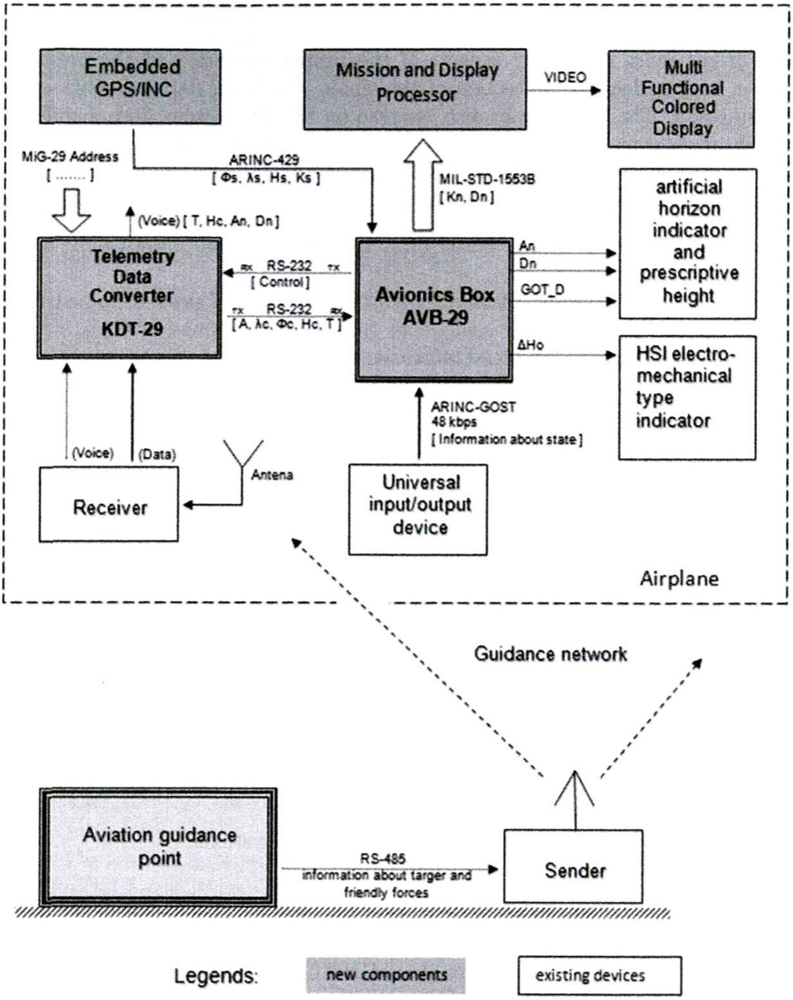

Fig. 2. System block diagram

4 The Use of AI Methods

In order to implement the Artificial Intelligence methods, the data flow of information which was presented has to be modified. The architecture solution diagram with AI method implemented is shown in the Fig. 3.

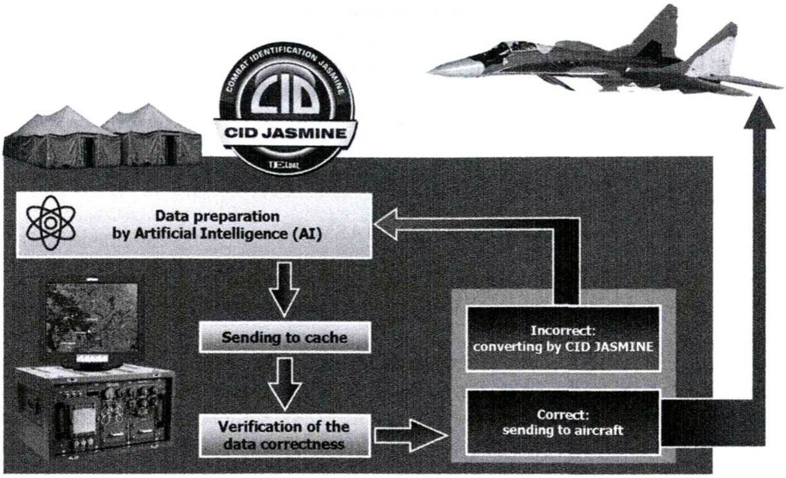

Fig. 3. Flow of information

According to this architecture, Artificial Intelligence is used for preparing the data which is to be placed in the cache. This data is selected based on the actual situation on the battlefield. The system prepares the package for the cache based on AI methods. This package contains a prepared prediction of the plane trajectory, using the collected data of the maneuvers which are performed most frequently.

The most effective AI method for this type of solution is one of the machine learning implementation [2,9,10]. Such a mechanism will improve significantly the dynamics of decision-making and eliminate the least delays [13]. This type of system would learn by examining its behavior patterns based on the pilots operations. After a sufficient number of inputs, AI will be able to predict how the pilot will behave. By this method the machine thinks with the pilot, which is a much faster solution.

For this purpose, a mechanism of Artificial Intelligence uses the following inputs:

- the position of own and foreign units;
- the type of foreign entities;
- the type of aircraft.

The output is a predicted concept of the future trajectory of the aircraft on the battlefield.

As a result, the prepared data is sent to cache. If the pilot decides to fly as provided by the AI method, the data is already given, otherwise it will be converted by CID JASMINE.

One of the best machine learning mechanisms which could be used in that kind of system is the reinforcement learning [8]. This method was chosen for the first test because of the possibility to provide the system description as an input model.

Described solution actively participated in the international exercises Bold Quest 2014 and Bold Quest 2015. These exercises were performed in the USA, and JASMINE CID was the only product from Poland in the history of this international exercise.

During Bold Quest exercises the CID JASMINE:

- has been accredited to connect to the training network;
- made positive tests with NATO military systems, including USA, Denmark and Italy;
- was an agent for the exchange of data, e.g. with airplanes A10, B1, F15, F16, F18, F18D and F18E-Super Hornet in the terms of: management and visualization of data availability from JASMINE CID server in plain text on remote terminal, and tracking the listed aircraft;
- passed load tests with positive results;
- collaborated with Patriot air defense system on the exchange of information on air situation

These exercises were performed with the MIP database, which contained the information about the aircrafts: A10, B1, F15, F16, F18, F18D and F18E-Super Hornet. By using the JC3IEDM, JASMINE CID had on-line access to this information.

5 Conclusion

The practical tests have already been performed. The implemented design concept is presented in this publication. CID JASMINE has already been tested on BQ 2014 and BQ 2015 exercises.

The first stage of the tests, the communication of the headquarters with the plane on the ground, has already been performed. The next step, air testing, will be performed in the near future. Implementation of the AI machine learning has been launched. It should be mentioned that the JASMINE System will be the first to use AI to increase the system performance for military use.

References

1. Biuletyn konstrukcyjny p/o/r/u/5034/k/08
2. Angryk, R.A., Czerniak, J.: Heuristic algorithm for interpretation of multi-valued attributes in similarity-based fuzzy relational databases. Int. J. Approximate Reasoning **51**(8), 895–911 (2010)
3. Apiecionek, Ł., Romantowski, M.: Secure IP network model. Comput. Method Sci. Technol. **4**, 209–213 (2013)

4. Apiecionek, Ł., Romantowski, M., Śliwa, J., Jasiul, B., Goniacz, R.: Safe exchange of information for civil-military operations. In: Military Communications and Information Technology: A Comprehensive Approach Enabler, pp. 39–50 (2011)
5. Apiecionek, Ł., Biernat, D., Makowski, W., Lukasik, M.: Practical implementation of AI for military airplane battlefield support system. In: 2015 8th International Conference on Human System Interactions (HSI), pp. 249–253. IEEE (2015)
6. Apiecionek, Ł., Czerniak, J.M., Zarzycki, H.: Protection tool for distributed denial of services attack. In: Kozielski, S., Mrozek, D., Kasprowski, P., Małysiak-Mrozek, B. (eds.) BDAS 2014. CCIS, vol. 424, pp. 405–414. Springer, Heidelberg (2014)
7. Apiecionek, L., Romantowski, M.: Security solution for cloud computing (2014)
8. Bradtke, S.J., Barto, A.G.: Learning to predict by the method of temporal differences. Mach. Learn. **22**, 33–57 (1996). (Springer)
9. Kosinski, W., Prokopowicz, P., Slezak, D.: On algebraic operations on fuzzy reals. In: Rutkowski, L., Kacprzyk, J. (eds.) Neural Networks and Soft Computing. Advances in Soft Computing, vol. 19, pp. 54–61. Springer, Heidelberg (2003)
10. Kozielski, M., Skowron, A., Wróbel, Ł., Sikora, M.: Regression rulelearning for methane forecasting in coal mines. In: Kozielski, S., Mrozek, D., Kasprowski, P., Małysiak-Mrozek, B., Kostrzewa, D. (eds.) BDAS 2015. CCIS, vol. 521, pp. 495–504. Springer, Heidelberg (2015)
11. Kruszynski, H., Kosowski, T., Apiecionek, L.: CID server JASMINE. In: V Communications Conference in Sieradz (2014)
12. Lojka, T., Zolota, M., Zolotová, I., et al.: Communication engine in human-machine alarm interface system. In: Sincak, P., Hartono, P., Vircikova, M., Vascak, J., Jaksa, R. (eds.) Emergent Trends in Robotics and Intelligent Systems. Advances in Intelligent Systems and Computing, pp. 129–136. Springer, Heidelberg (2015)
13. Vidhate, D., Kulkarni, P.: Cooperative machine learning with information fusion for dynamic decision making in diagnostic applications. In: 2012 International Conference on Advances in Mobile Network, Communication and its Applications (MNCAPPS), pp. 70–74. IEEE (2012)

Intelligent FTBint Method for Server Resources Protection

Łukasz Apiecionek[1(✉)] and Wojciech Makowski[2]

[1] Institute of Technology, Casimir the Great University in Bydgoszcz,
ul. Chodkiewicza 30, 85-064 Bydgoszcz, Poland
lapiecionek@ukw.edu.pl

[2] TELDAT Sp. z o. o.sp.k., Bydgoszcz, Poland
wmakowski@teldat.com.pl

Abstract. The subject of this article is the issue of security of network resources in computer networks. One of the main problems of computer networks are Distributed Denial of Service attacks, which can take all server resources and block them. The FTBint intelligent method can manage the amount of network traffic passed to a server and help the server to work during the attack. After the attack is recognized the number of connections provided to the server can be changed in time in an intelligent way. Such solution gives time to the server to dispose of the resources which were allocated incorrectly by the attacker. This new concept is different from the one used in the currently existing methods, as it enables the user to finish his work which had been started before the attack occured. Such user does not suffer from DDoS attacks when the FTBint method is used. The proposed method has already been tested.

Keywords: DDoS · Security · Network

1 Introduction

There are lots of useful servers operating in the Internet network nowadays. Users need a fast access to information provided by servers from every part of the network. One of the obstacles they can face are Denial of Service attacks, or rather Distributed Denial of Service nowadays. They cause network unavailability by blocking services via seizing system resources in computers in the network until they stop working. A user who has already started working in the system loses the connection and cannot even log out of the system, which has to do it for him after the connection timeout is reached or when a broken connection is detected. DDoS attacks are presently a serious obstacle for IT systems' efficient functioning and some new idea of dealing with them is necessary. There are only few common methods of fighting the DDoS attack problems [2–5, 11]. The main idea behind them is to use the Intrusion Detection System and Intrusion Prevention System (IDS/IPS in short) solutions. Such systems are efficient provided that they have a description of well-known attacks or some kind of Artificial Intelligence solution which could learn the actions in some specific scenarios of attack.

S. Kozielski et al. (Eds.): BDAS 2016, CCIS 613, pp. 683–691, 2016.
DOI: 10.1007/978-3-319-34099-9_53

Other solutions suggest using a firewall mounted on the edge of the network. However, the systems based on this concept will only block the incoming traffic on specific ports or IP address ranges, which is not sufficient. During the attack the server will stop responding and become unavailable. This paper presents a new intelligent method called FTBint. This method is able to limit the traffic during the attack in an intelligent way and after the attack is over the network can smoothly return to its previous state. This solution was implemented and tested in a real environment. The structure of this paper is as follows. Section 2 shortly describes the issue of the DDoS attacks and introduces the proposed method for fighting them. Section 3 presents the results of the implementation of the described method. Section 4 provides a conclusion and discussion over the developed method.

2 FTB Intelligent Method Description

The DDoS attacks are widely described in the literature [5,11]. These attacks can be performed on various system resources: TCP/IP sockets [11,12] or DNS servers. Regardless of the method, the main principle is to simulate so many correct user connections that their number exceeds the actual system performance and drives it to abnormal operation. The transmission of the attackers' packets is done through the provider's network and if it cannot be blocked, it leads to data link saturation. Such saturation results in a lack of connection to the server. The proposed solutions designed to prevent such situations are not specific and their implementation is associated with many problems. The most common concern is the limited performance of the network devices. However, it is possible to limit the incoming traffic on a firewall and allow the servers to deal with the already established connection. This will let the users finish their work and the new users will be able to connect to the server. This is achieved via implementing intelligent network protection using the FTBint method.

The role of the input firewall is to control the incoming traffic on the edge of the network. When the network is about to give access to the server to the external users, a specific type of traffic has to be allowed by the incoming rules. For instance, in the case of an http server, usually the TCP port 80 has to be opened for the incoming connections. When an attack on the server occurs, this port is still open.

This situation leads in turn to the server overload. Thus, a special firewall FTBint module was developed, the role of which is to filter the traffic on the server's open port and to limit it according to the determined policy in an intelligent way.

During the server's regular work, all packets are passed through and the network is not under any attack. Recognizing the attack by the FTBint method is based on counting the opening network connections to the server during time slots *t1*. The attack is recognized when the opened connections counted in time slots exceed *packet_limit*. As a result, the intelligent FTBint filtration process is launched. At the beginning of the filtration a list of the IP addresses which communicate with the server correctly, i.e. which are not a part of the DDoS attack - *listIP* - is taken from the server. During the filtration each packet is checked whether it is on the list of the valid IP addresses *listIP*; if so, the packet is sent to the network, if not, a counter of the passed packets *packet_counter* is checked whether its value is greater than the allowable packet limit *packet_limit* in a time slot *t1*. If the limit of the packets is exceeded, the packet is dropped - *DROP*. In the following time slot the number of current packets *packet_counter* is zeroed and the above mentioned filtration process is restarted. Afterwards the FTBint method has to regulate the opening connection limit in an intelligent way. It is made in the following way:

- when the limit of the packets is exceeded in a given number of the subsequent time slots *subsequent_time_limit* the FTBint method recognizes that the network is facing a large attack on the server and in order to give the server some time to regain efficiency, the limit of packets *packet_limit* allowed to pass in time slots is decreased,
- if in the following time slots the packet limit is not exceeded, the FTBint method recognizes that the attack on the server started to lose its intensity and the *packet_limit* can be increased.

Changing the packet limit allows the server to handle the incoming connections which may be potentially correct or to release the resources used incorrectly by the attacker. Despite the attack, the server is still accessible to the users who were working on it when the attack was detected.

The process of decreasing the packet limit can depend on the server type, its needs and kind of work. Moreover, the limit values may require experimental determination or setting them basing on the server's resources, its operating system, the amount of memory and processor type.

A pseudocode of the main part of the algorithm responsible for passing the packets as well as narrowing the limits is shown below:

```
Algorithm pseudocode

packet_counter:=packet_counter + 1

if packet_counter < packet_limit then
    packet pass
else
    begin
    if IP address in listIP then
       packet pass;
    else
       packet drop;
    end;
if times_slots ends then
   begin
```

```
if packet_counter>packet_limit then
    overdrop_times=overdrop_times + 1;
    packet_counter=0;
if overdrop_times> subsequent_time_limit then
    packet_limit=packet_limit/2;
    overdrop_times=0;
else
    packet_limit=packet_limit*2;
    overdrop_times=0;
end;
```

The *packet_counter* variable contains the number of packets which are passed through in a certain time slot. When its value does not exceed the permissible limit *packet_limit*, the packet is sent, but when the limit is exceeded, further tests are performed. If the packet is present in the database of the known IP addresses *listIP*, the packet is passed through, otherwise it is dropped. When the *time_slots* timeout expires, a verification is performed whether the limit of packets was exceeded in this time slot. In this case, in the implemented method the counter of the limit was increased in the subsequent time slots. When the limit was exceeded in the following two time slots, the allowable limit of packets was decreased (*packet_limit/2*). If in the next time slots the limit was not exceeded, the limit of packets was increased (*packet_limit*2*).

3 Implementation Results

In order to verify if the server works continuously indeed, the method was implemented and tested. The implementation consisted of a module for the firewall's IPTables module on a Debian Linux system with 2.6.32 kernel. The tests simulating the most common types of attacks on the servers were performed on a simple network, which was built for this purpose according to well known structures and routing protocol requirements 789. A graphical interface to Asterix FreePBX distribution, working under CentOS with kernel 2.6.32 was used as an http server and an Apache 2.2.15 server, equipped with 1 GB of RAM, was used as a receiver. In order to perform the attack, a Sender machine was used - based on a Debian Linux with 2.6.32 kernel (equipped with 512 MB of RAM) and DDOSIM software (Layer 7 DDoS Simulator v0.2). The server was connected to a real network through the firewall with the FTBint method implemented, which was running on a Linux Debian with 2.6.32 kernel with 512 MB of RAM.

During the test, the memory usage and the http server response times were observed. The attack consisted of sending a large amount of HTTP Get requests to the server. Without the FTBint method, the server's memory usage rose up to 100 % and the server stopped responding. After launching the FTBint method, RAM memory usage was observed.

The FTBint method was configured with the following limits:

- time slot for analyzing the amount of transmitted data *time_slots* = 1 s;
- allowable limit of packets *packet_limit* = 30;
- minimal limit of packets = 10;

Five tests of the attack on the server were performed, each lasting one hour and under the following conditions:

- 100 HTTP GET requests sent every 30 s;
- 1000 HTTP GET requests sent every 30 s;
- 2000 HTTP GET requests sent every 10 s;
- 10000 HTTP GET requests sent every 10 s;
- 50000 HTTP GET requests sent every 30 s.

None of the cases resulted in a server overload. Prior to the attack, a connection with the server through the firewall was established and it remained active because it had been started before the attack and the computer was recognized as allowed to communicate with the server (Fig. 1).

During the test the amount of free RAM memory was observed on the http server and on the firewall. On the http server the memory remained mostly on a constant level (Fig. 2).

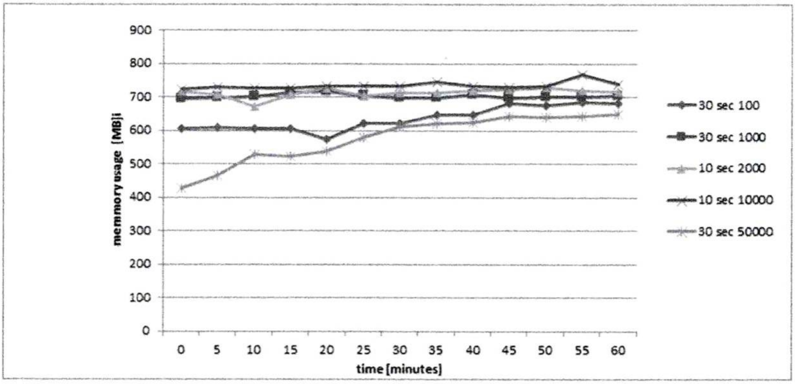

Fig. 1. RAM usage on http server

On the firewall the RAM usage remained on a constant level. During the test which consisted of sending 1000 packets every 10 s, the RAM memory usage was higher, which was probably a result of other operations of the device.

The RAM memory usage test was also performed in a network built without using the proposed FTBint method. The results are shown in Fig. 3. Without

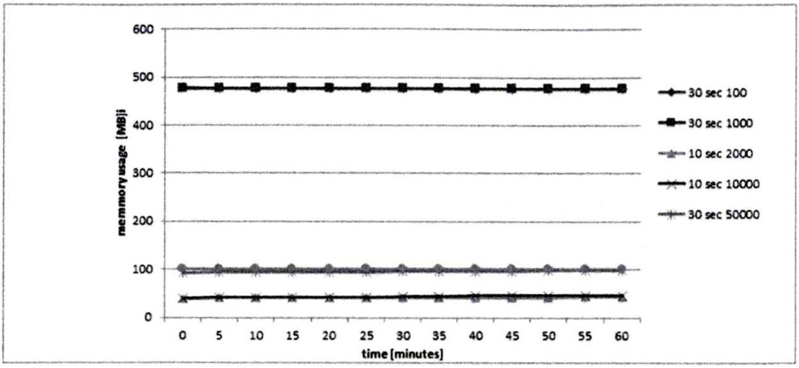

Fig. 2. RAM usage on firewall

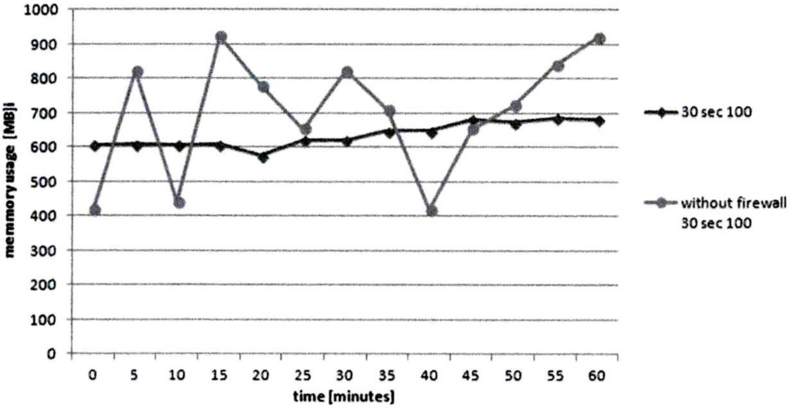

Fig. 3. RAM usage on http server with and without FTB method on firewall

the active method, the memory usage rose up to 100 % several times (the server was equipped with 1 GB of RAM).

In the description of the method a mechanism for intelligent narrowing the amount of packets in a single time slot was mentioned. Figure 4 shows a fragment of the operation of the implemented FTBint mechanism. It regulates the number of packets which can be transmitted in a time slot dynamically, depending on the recognized server load.

It is worth emphasizing that after the attack, the network regains its ordinary state. It means that using the FTBint method allows the network to work without requiring the administrator's action. The authors are planning to implement an improved solution for recognizing DDoS attacks, which will not be limited to counting the connections during the time slots. Some literature concerning using fuzzy sets for this purpose has already been published [1,6,8–10]. The authors are conducting research on this topic as well as in order to improve this element [7].

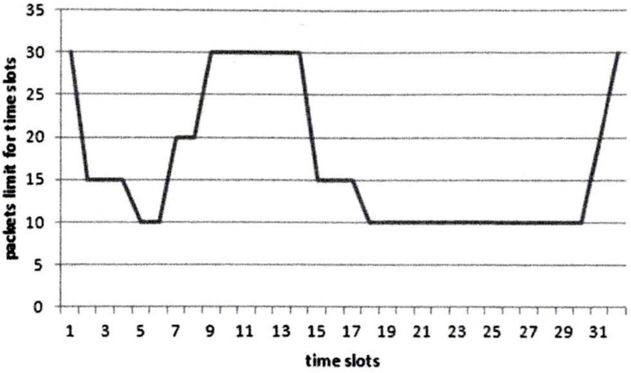

Fig. 4. Intelligent packet limit on time slots by FTBint method

Table 1. Method comparison

Criteria	Existing methods	FTBint method
Server memory usage	Up to 100 %	Constant defined amount
Server availability during attack	No response	Response
Possibility to connect new user to the server during attack	Impossible	Possible, Network browsers tries to connect multiple time, so the user get the chance to connect
Possibility to finish the task with the server during attack	impossible	Possible
Connection limit during attack	Zero	Limited to defined amount by administrator according to proposed algorithm
Connection limit after the attack	Needs time and administrator's work to get back	No work required to back to previous state
Administrator work required after the attack	Required	Not required

The Table 1 compares the FTBint method with other existing methods according to the server memory usage, possibility of initiating a connection during the attack, the connection limit after the attack stops and the required administrator's actions after the attack.

In this article the FTBint concept of eliminating DDoS attacks was introduced. While the methods suggested in the literature can block the access to the resources when the attack occurs, by using a firewall along with IDS/IPS mechanisms, during the time of the blockage no user from an external network can connect to the desired resources. Moreover, such solution does not allow to complete the work started by the users who were already connected. The users who worked with the server lose their connection. The FTBint method described in this article allows the users to continue their work even if the attack occurs. It is possible for the users who were connected to the server prior to the attack and the server informed the method about this fact. The FTBint method limits the connections to the server in an intelligent way, which lets a new user connect to the server. The method was implemented and tested in practice. It does not cause an increase of the firewall load but it prevents the server from overload by keeping the its load at a stable level during the attack. The proposed method may be successfully implemented on any firewall-type device.

The author is ready to provide the sources of the described method for further analysis.

References

1. Angryk, R.A., Czerniak, J.: Heuristic algorithm for interpretation of multi-valued attributes in similarity-based fuzzy relational databases. Int. J. Approximate Reasoning **51**(8), 895–911 (2010)
2. Apiecionek, Ł., Czerniak, J.M., Zarzycki, H.: Protection tool for distributed denial of services attack. In: Kozielski, S., Mrozek, D., Kasprowski, P., Małysiak-Mrozek, B., Kostrzewa, D. (eds.) BDAS 2014. CCIS, vol. 424, pp. 405–414. Springer, Heidelberg (2014)
3. CA-1996-01, C.A.: UDP port denial-of-service attack. http://www.cert.org/advisories/CA-1996-01.html
4. CA-1996-21, C.A.: TCP syn flooding and ip spoofing attacks. http://www.cert.org/advisories/CA-1996-21.html
5. Chang, R.K.: Defending against flooding-based distributed denial-of-service attacks: a tutorial. IEEE Commun. Mag. **40**(10), 42–51 (2002)
6. Czerniak, J.: Evolutionary approach to data discretization for rough sets theory. Fund. Inform. **92**(1–2), 43–61 (2009)
7. Czerniak, J.M., Apiecionek, Ł., Zarzycki, H.: Application of ordered fuzzy numbers in a new OFNAnt algorithm based on ant colony optimization. In: Kozielski, S., Mrozek, D., Kasprowski, P., Małysiak-Mrozek, B., Kostrzewa, D. (eds.) BDAS 2014. CCIS, vol. 424, pp. 259–270. Springer, Heidelberg (2014)
8. Dobrosielski, W.T., Szczepański, J., Zarzycki, H.: A proposal for a method of defuzzification based on the golden ratio-GR. In: Atanassov, K.T., et al. (eds.) Novel Developments in Uncertainty Representation and Processing. AISC, vol. 401, pp. 75–84. Springer, Switzerland (2016). http://dx.doi.org/10.1007/978-3-319-26211-6_7

9. Ewald, D., Czerniak, J.M., Zarzycki, H.: Approach to solve a criteria problem of the ABC algorithm used to the WBDP multicriteria optimization. In: Angelov, P., et al. (eds.) IS 2014. AISC, vol. 322, pp. 129–137. Springer, Heidelberg (2015)
10. Kosiński, W., Prokopowicz, P., Ślęzak, D.: On algebraic operations on fuzzy reals. In: Rutkowski, L., Kacprzyk, J. (eds.) Neural Networks and Soft Computing. ASC, vol. 19, pp. 54–61. Springer, Heidelberg (2003)
11. Moore, D., Shannon, C., Brown, D.J., Voelker, G.M., Savage, S.: Inferring internet denial-of-service activity. ACM Trans. Comput. Syst. (TOCS) **24**(2), 115–139 (2006)
12. Schuba, C.L., Krsul, I.V., Kuhn, M.G., Spafford, E.H., Sundaram, A., Zamboni, D.: Analysis of a denial of service attack on TCP. In: 1997 IEEE Symposium on Security and Privacy, pp. 208–223. IEEE (1997)

Communications in Computer and Information Science

The CCIS series is devoted to the publication of peer-reviewed proceedings of conferences and workshops. Its aim is to efficiently disseminate original research results in computer science. All CCIS proceedings are available in electronic form from the SpringerLink digital library, and as printed books, and reach libraries and readers worldwide via Springer's distribution network.

Besides globally relevant meetings with internationally representative program committees guaranteeing a strict peer-reviewing and paper-selection process, conferences run by societies or of high regional or national relevance are also considered for publication. Application-oriented and interdisciplinary conferences are also welcome.

The topical scope of CCIS spans the entire spectrum of computer science ranging from foundational topics in the theory of computing to information and communications science and technology and a broad variety of interdisciplinary application fields.

CCIS proceedings can be published in time for distribution at conferences or as revised proceedings after the event. The publication is free of charge and an Open Access option is available at a fee. The language of publication is exclusively English.

CCIS is abstracted/indexed in DBLP, Google Scholar, EI-Compendex, Mathematical Reviews, SCImago, and Scopus. CCIS volumes are also submitted for inclusion in ISI Proceedings.

To start the evaluation of your proposal for inclusion in the CCIS series, please send an e-mail to ccis@springer.com.

ISSN 1865-0929

ISBN 978-3-319-34098-2

9 783319 340982

CCIS

❯ springer.com