Teaching Network Security With IP Darkspace Data

Tanja Zseby, *Member, IEEE*, Félix Iglesias Vázquez, *Member, IEEE*, Alistair King, and K. C. Claffy

***Abstract*—This paper presents a network security laboratory project for teaching network traffic anomaly detection methods to electrical engineering students. The project design follows a research-oriented teaching principle, enabling students to make their own discoveries in real network traffic, using data captured from a large IP darkspace monitor operated at the University of California, San Diego (UCSD). Although darkspace traffic does not include bidirectional conversations (only attempts to initiate them), it contains traffic related to or actually perpetrating a variety of network attacks originating from millions of Internet addresses around the world. This breadth of coverage makes this darkspace data an excellent choice for a hands-on study of Internet attack detection techniques. In addition, darkspace data is less privacy-critical than other network traces, because it contains only unwanted network traffic and no legitimate communication. In the lab exercises presented, students learn about network security challenges, search for suspicious anomalies in network traffic, and gain experience in presenting and interpreting their own findings. They acquire not only security-specific technical skills but also general knowledge in statistical data analysis and data mining techniques. They are also encouraged to discover new phenomena in the data, which helps to ignite their general interest in science and engineering research. The Vienna University of Technology, Austria, first implemented this laboratory during the summer semester 2014, with a class of 41 students. With the help of the Center for Applied Internet Data Analysis (CAIDA) at UCSD, all exercises and IP darkspace data are publicly available.**

***Index Terms*—Communication system security, data analysis, engineering education, security.**

I. Introduction

NETWORK SECURITY is one of the most challenging fields in communication networks research. A well-focused and forward-looking education is required to generate scientists able to cope with the ever-changing challenges and threats in communication networks. Besides the need in academia, industry and governments also demand well-trained security experts who can cope with new developments in attack and defense strategies. Even engineers who are not directly involved in developing security solutions benefit from a security education when designing or deploying new systems. For a balanced security education, students should not only learn principles underlying preventative approaches to prevent or minimize the likelihood of attacks (e.g., encryption or access control) but also how to detect attacks in progress or once they have occurred. This latter objective requires intelligent observation and interpretation of network traffic in order to find anomalies and discover new attacks. Working with real traffic data reveals data analysis challenges that are difficult to convey otherwise, and this project design reflects the essential benefit students get from hands-on experience with traffic data.

This paper presents a network security lab experiment for teaching an important class of network security methods—traffic anomaly detection—to electrical engineering students (upper undergraduate or Master's level). The lab uses data from a large IP darkspace monitor at the University of California San Diego (UCSD), which contains traffic from a variety of different network attacks originating from millions of Internet addresses around the world. This breadth of coverage makes this darkspace data an excellent choice for a hands-on study of Internet attack detection techniques. Additionally, darkspace data is less privacy-critical than other network traces, because it contains only unwanted network traffic and no legitimate communication.

The Vienna University of Technology (TU Vienna), Austria, first implemented this laboratory in a Network Security course (NetSec-I) during the summer semester of 2014, with a class of 41 students. With the help of the Center for Applied Internet Data Analysis (CAIDA) at UCSD, all exercises and IP darkspace data are publicly available in order to encourage adoption of this class by other instructors of network security in electrical engineering and computer science [1].

II. Related Work

Increasing demand for security experts has motivated many universities to include classes in security engineering in computer science curricula, ideally supplemented by lab exercises to help students to digest conceptual knowledge. Early approaches, [2], [3] provided basic recommendations for designing cyber warfare labs for teaching security auditing methods. For example, in [2], the instructors split undergraduates into three groups (attack, defense and forensics) to carry out experiments in an isolated lab with 50 different attack/defense tools. Students signed an agreement committing to appropriate use of the tools and documented their lab work. The authors recommend a more heterogeneous lab to provide more exploration opportunities for students. Lee *et al.* also presented a competition-based lab experiment [4], where students set up and defend their own machines and attack machines set up by other students. Georgia Tech instructors established an isolated laboratory network to support attack and defense exercises [5], through its use they learned that for such a lab

Manuscript received October 09, 2014; revised March 04, 2015; accepted March 17, 2015. This work was supported by the USA National Science Foundation (NSF) under grants CNS-1059439 and CNS-1228994 and by the DHS S&T Cyber Security Division PREDICT project via Cooperative Agreement FA8750-12-2-0326.

T. Zseby and F. Iglesias Vázquez are with the Institute of Telecommunications, Vienna University of Technology, 1040 Vienna, Austria (e-mail: fiv@gmx.at).

A. King and K. C. Claffy are with the Center for Applied Internet Data Analysis (CAIDA), University of California San Diego, La Jolla, CA 92093-0505 USA.

Digital Object Identifier 10.1109/TE.2015.2417512

to be successful, it must be *"as realistic, large scale, and interactive as possible"* [5].

Others have developed security labs using virtual machines [6], [7], recognizing that students are more motivated to take a security class that includes a lab than they are with a pure conceptual class. Othmane *et al.* [7] found that sometimes the answers students provided were unexpected but not wrong; in these cases they let students demonstrate their work to determine its correctness. Marsa-Maestre *et al.* [8] presented a tool that generates different security scenarios for Internet security education and found that student grades improved with the addition of a laboratory component.

The TU Vienna NetSec-I lab teaches network anomaly detection methods using real IP darkspace traffic from millions of actual Internet sources. It provides a real, large-scale and heterogeneous setting as recommended in [2] and [5]. Following further suggestions from [2] and [7], students are required to sign a lab agreement, continuously document their findings and demonstrate new skills in a review session. They also individually explore assigned parts of the data.

III. Approach and Intended Outcome

The TU Vienna network security class combines lectures about security concepts with the NetSec-I lab that complements that conceptual knowledge with hands-on experiments. The target group consists of Master's students in electrical engineering, who have already passed a communication networks class.

The NetSec-I lab concentrates on traffic anomaly detection techniques for three reasons. First, designing and implementing anomaly detection techniques requires data analysis experience, which is best acquired through practical training. Second, covering a few methods in depth is more rewarding for students than shallow experience with several different concepts. Third, expertise with data analysis techniques is a valuable skill for any future scientific career.

The educational objectives of the lab are to meet the following.

1) *Familiarize students with network data analysis methods:* After completing the lab, students should be able to understand network data traces and apply statistical analysis techniques to network trace data.
2) *Deepen students' network security knowledge:* Students should gain knowledge about a variety of attacks and attack preparation techniques and insight into network protocols (TCP, UDP and ICMP). The lab should also improve students' general security awareness.
3) *Enable students' general scientific work skills:* Students should learn to solve scientific problems autonomously and in small (potentially heterogeneous) teams. They should gain skills in presenting and interpreting results and responsibly handling critical data.
4) *Awaken the scientist in each student:* The lab should teach students the fun in discovering new phenomena and should generate interest in scientific work. It should motivate students to take further classes or a Master's thesis in the area of network security, or to apply the acquired general data analysis skills in other fields.

In addition to student learning objectives, the design incorporates five teaching objectives, as follows:

- implement research-oriented teaching;
- ensure equal treatment of students;
- prevent cheating;
- use objective criteria for evaluating students;
- limit the effort required for evaluation.

Research-oriented teaching is the concept of bringing teaching and research closer together [9]. TU Vienna subscribes to this concept and the laboratory exercises are designed to implement research-oriented teaching.

While many existing security labs run isolated testbeds or work with artificial traffic, the TU Vienna NetSec-I Lab uses real network traffic from a large IP darkspace monitor at UCSD [10], operated by CAIDA [11]. The darkspace monitor uses an IP network address range that is announced to the Internet but has nearly no actual hosts attached. The resulting *darkspace traffic data* is heterogeneous, since it originates from millions of different sources, with different operating systems and access network technologies located all over the world. So it offers many exploration opportunities, which helps to hold students' interested, as described in [2].

The darkspace traffic is collected at UCSD using an entire /8 network with 2^{24} darkspace addresses, which corresponds to 1/256 part of the whole IPv4 Internet. Access to such a large IP darkspace is rare, because IPv4 addresses are a scarce resource nowadays.

IV. Key Design Decisions

The NetSec-I lab teaches hands-on skills on network anomaly detection methods to complement conceptual knowledge acquired in lectures.

A. Prerequisites for Students

Basic knowledge about IP networks, as well as experience with MATLAB and shell scripting, are expected of students enrolled in the NetSec-I lab. However, TU Vienna initially did not mandate formal prerequisites, allowing participation of students from other master programs, e.g., computer technology or technical computer science.

The first six lectures (90 min each) provide conceptual background. In these lectures students learn security basics and network traffic analysis techniques. The lectures cover different analysis techniques, such as analyzing time series and feature distributions, finding periodicities in the frequency spectrum and using entropy as a condensed metric to assess feature dispersion or concentration. The lectures also cover specific characteristics of IP darkspace traffic. A written test allows evaluation of student understanding of the conceptual background.

Students are not required to have passed the test before participating in the lab, although it is recommended. Before lab appointments, students take a lab introduction session to review concepts learned in the first part of the course, finalize group assignments, learn lab rules and receive exercise sheets.

B. Dataset

To design a lab that is *"realistic, large scale, and interactive"* [5], the TU Vienna NetSec-I lab uses real network traffic

captured from a large IP darkspace monitor. The traffic mainly consists of real network attacks, attack preparation activities or victims responding to attacks and is used to study new attack patterns [12], [13] and other global Internet phenomena [14]. Raw IP packets are captured in the commonly used pcap format [15] and then processed to extract the most important packet header fields for analysis (FlowTuple format [16]).

C. Data Protection

Although darkspace traffic is less privacy-critical than standard network traffic, some filtering helps to protect potentially sensitive information.

1) The first byte of the destination IP addresses denotes the darkspace network address; this is always set to zero to prevent potential attackers from knowing the exact address range of the darkspace.
2) Source IP addresses are anonymized with Crypto-PAn [17] to protect victims against further attacks.
3) IP packet payload, not needed for the exercises, is removed to reduce file size.
4) Students must sign an agreement (similar to that in [2]) not to copy any data or attempt to de-anonymize IP addresses. The CAIDA Acceptable Use Agreement [18] serves as the basis for this, and it is extended with the general IT rules of the Vienna University of Technology and some specific lab rules.

D. Lab Sessions Setup

Students perform the exercises in three separate three-hour sessions, but they may request extra time. Students work in pairs, which enables them to exchange ideas and learn teamwork, but still allows the instructor to verify that both students contributed to the exercise. Students are encouraged to team with someone with complementary skills, e.g., a student with MATLAB experience might pair with one with shell scripting skills.

All students receive the same exercises to promote equal treatment and to reduce evaluation effort (see the teaching objectives in Section III), yet each team is assigned a different part of the data set to reduce the possibility of cheating. This also increases student interest, since teams can then discover different phenomena from each other.

E. Required Software and Tools

The exercises require the following software tools.

1) *tcpdump* [15], a standard tool for capturing and filtering pcap packet traces.
2) *corsaro* [16], a specialized tool to convert and aggregate IP darkspace data (provided by CAIDA, UCSD).
3) *MATLAB* [19], a tool for numerical computation and mathematical analysis. Exercises are also fully compliant with the open source alternative *Octave* [20].
4) *RapidMiner* [21], for machine learning and data mining.

Both MATLAB and RapidMiner can show students the strengths of different tools and the advantages of their combined deployment. Additional files and scripts required for the exercises are available at [1].

F. Exercise Description

Table I summarizes the exercises; the full exercise sheet is available at [1].

1) Students first get familiar with *pcap* and *FlowTuple* data formats. They learn to use *corsaro* to transform *pcap* data into *FlowTuple* files, which summarize the most relevant traffic features. Students are required to:
   a) list and explore *pcap* files;
   b) transform *pcap* files into *FlowTuple* files;
   c) list and explore *FlowTuple* files.
2) Since *Flowtuple* traces contain some categorical information not amenable to statistical analysis, students aggregate information to obtain numerical time series. They are required to:
   a) obtain aggregated values of packets and unique IP sources for different time rates (hour, minute);
   b) extract descriptive statistics of aggregated data and compare outcomes;
   c) generate time series in Comma Separated Values (CSV) files.
3) Using univariate analysis of aggregated traffic, students investigate the evolution of different time series, e.g., amount of packets or unique IP sources, Fig. 1, and analyses network protocol and destination port distributions, which often reveal abnormal phenomena. Students are required to:
   a) plot time series of aggregated data (packets, unique IP sources; per minute, per hour);
   b) infer traffic phenomena from the obtained plots;
   c) aggregate data for analyzing protocol distributions and TCP destination port distributions;
   d) extract descriptive statistics and histograms;
   e) reason about the significance of the outcomes.
4) Students study the frequency spectrum of the aggregated signals using fast Fourier transformation (FFT). Here they can apply signal processing knowledge acquired in other classes to find temporal patterns in network data. Students are required to:
   a) apply FFT on time series (packets, unique IP sources);
   b) plot FFT results and detect periodicities;
   c) obtain average-day plots with error bars of both signals and compare with the results of descriptive statistics;
   d) study the correlation between packets and unique IP sources time series.
5) Students use RapidMiner to perform additional univariate analysis, and create histograms and metadata of *Flowtuple* traces. RapidMiner handles datasets that mix categorical and numerical data, and allows filtering to isolate unusual phenomena, e.g., a source attempting to connect many times to a specific destination port. Students are required to:
   a) import data for analysis considering scale-types;
   b) analyze traffic feature by feature (*FlowTuple* format) by using meta-data statistics and histograms; detect anomalies and explain observed phenomena;
   c) filter and analyze the specific case of the most commonly recurring source IP.

TABLE I
NETWORK SECURITY LAB EXERCISES

Exercise set	Tools	Result	Teaching Objectives
Data Pre-processing	tcpdump, corsaro	Data in FlowTuple format	Students become familiar with data formats (pcap, FlowTuple), traffic characteristics, and analysis tools: tcpdump, corsaro.
Data Aggregation	corsaro	Time series of number of packets, number of active sources.	Students familiarize themselves with corsaro tool aggregation capabilities and learn how to treat numeric and categorical data.
Univariate analysis, time series plotting	MATLAB	Basic statistics, plots of packets and active sources time series; protocol and destination port histograms	Students learn to abstract knowledge from aggregated signals, basic MATLAB commands, preparation of graphics to present results. Students discover Patch Tuesday effects and Conficker worm tracks.
Analysis of temporal patterns	MATLAB	Frequency spectrum of time series, daily average curves	Students learn how to apply signal processing methods (e.g., FFT) to discover periodic behaviors (e.g. diurnal patterns). They explore the evolution and correlation of traffic aggregated signals.
Univariate analysis of FlowTuples	RapidMiner	Metadata and histograms of FlowTuple features	Students get familiar with RapidMiner, reasoning based on feature value distribution, application of feature filters, and detection of anomalous and recurrent traffic phenomena.
Bivariate analysis	RapidMiner	Metadata and scatter plots of FlowTuple features	Students discover suspicious phenomena and try to characterize the involved hosts, boosting their curiosity with the prospect of discovery.

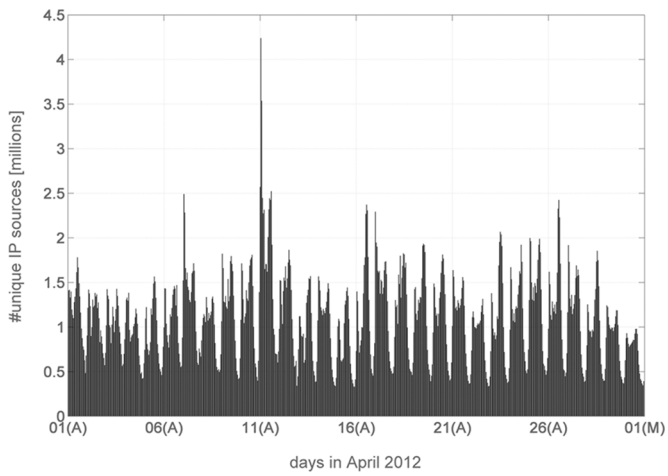

Fig. 1. IP sources per hour observed in the darkspace for April 2012. The peak on April 11 shows the effect of Microsoft's Patch Tuesday release [13].

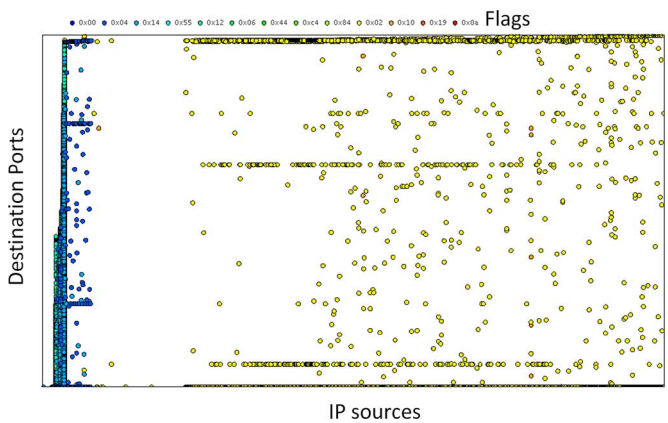

Fig. 2. Vertical and horizontal lines of dots on this RapidMiner capture scatter plot (showing IP sources, destination ports and flags) suggest suspicious traffic behaviors, such as ports attacked by many sources or individual sources scanning multiple ports.

6) Finally, Rapidminer can perform bivariate analysis to detect anomalies by plotting different traffic features against each other, see Fig. 2. Students are encouraged to find traffic sources that show outlying behaviors and try to interpret which kind of attack or traffic phenomenon they detected. Since each group works on a different dataset their discoveries vary. Students are required to:
   a) perform bivariate analysis on different features (*FlowTuple* format) by using scatter plots; detect anomalies and explain observed phenomena;
   b) filter and analyze the specific case of the source IP sending packets to the maximum number of different destination ports;
   c) filter and analyze the specific case of the destination port getting more packets from different IP sources.

G. Initial Trial Session

A trial run of the laboratory experiment with two students several weeks before the semester started provided an opportunity to test lab operations, validate assumptions about the required time per exercise, and detect any technical issues. This trial session was extremely valuable and led to changes in the final design of the project, including improvements in the description of the exercises, adjustment of timing, and resolution of some system administration issues. The trial session also verified that students were captivated by the data and wanted to learn about findings from other students. The final version of the laboratory course thus included a general discussion session after the lab to allow students to exchange findings and lessons learned.

H. Correction and Assessment of Student Results

Each team prepares a written lab report with their results, which are evaluated with prefixed correction criteria and solver scripts, both available at [1]. Since unexpected outcomes are possible by design, the evaluation process includes a short lab review, similar to [7]. Reviews took approximately 30 minutes per team and consisted of asking students three to five questions, at least one of which was addressed to each student alone to evaluate the individual performance of both team members. Students can obtain up to 30 points for the report, plus up to ten points based on their individual performance during the review. Passing the lab requires 21 points total. The final grade is combined from the earlier exam testing conceptual knowledge and the lab exercise.

V. EVALUATION OF THE LAB

During the registration period for the first offering of the new class, in summer semester 2014, 62 students registered in order

TABLE II
RELATIONSHIP BETWEEN EDUCATIONAL OBJECTIVES AND TESTS

Objectives	*Familiarize students with network data analysis methods*	*Deepen students' network security knowledge*	*Enable students' general scientific work skills*	*Awaken the scientist in each student*
Report	Students learn how to use tools for the preprocessing and aggregation of traffic data (A2), the extraction of descriptive traffic values (A3), the calculation of trends and periodicities (A4), and the detection and filtering of traffic anomalies (A5 and A6).	Students acquire a deeper understanding of IP feature scale-types (A2), common traffic-types and expected feature entropies (A3), temporal patterns in darkspace traffic (A4), expected IP feature values (A5) and identification of traffic-types (A6).	Students develop skills on: correct plot design; presentation of results; suitable scientific writing style (A1); data parsing and transformation (A2); descriptive statistics (A3); basic time series analysis (A4); univariate and bivariate analysis (A5 and A6).	Students are encouraged to suggest hypothesis based on statistics (A4), and to carry out anomaly searching, reasoning on meta-data and forensic exploration (A5 and A6).
Oral exam	Students are required to use lab tools to solve analogous problems (O4) and to propose how to solve specific traffic analysis situations (O2).	Students' consolidation of fundamental concepts and relationships is tested (O1).	Students are required to show their expertise with data mining tools (O4) and to propose how to solve problems by means of data analysis (O2).	Students are provided with clues (obtained from analysis) to interpret what is happening in the network (O3).

to get access to slides and material.[1] Of these, 30–40 students regularly attended the lectures and 34 students took the theoretical exam before the lab started. A total of 41 students attended the lab. The lab reports generated by the students and a final oral exam per group (review session) were used to evaluate if the learning objectives were achieved. In addition, students could provide feedback about the class.

A. Methods for Evaluation of Acquired Knowledge

In order to assess the acquired network data analysis skills, the reports generated by students were evaluated according to the accomplishment of six different skill Areas (A):

A1: plotting and presentation of results;
A2: data preprocessing and aggregation;
A3: univariate analysis of aggregated data (basic);
A4: analysis of temporal patterns;
A5: univariate analysis of flow features (advanced);
A6: bivariate analysis of flow features.

The oral reviews completed the evaluation of students' security and data analysis knowledge in addition to the reports. Questions were selected from the fields at the report evaluation indicated presented the most difficulties to the specific group of students under review. Questions were also intended to assess individual consolidation of the knowledge presented. Every team was asked at least three questions (and each individual student was asked at least two). Oral review questions (O) covered three of the following four fields:

O1: consolidation of network security knowledge, e.g., *Explain why there are so many TCP SYN-ACK packets*;
O2: understanding of network security analysis methods, e.g., *Give an example of a network anomaly that is best detected using bivariate analysis of network flow features*;
O3: interpretation of statistical analysis, e.g., *Why does the histogram of the TTL show three isolated regions?*;
O4: application of lab tools, e.g., *Given a pcap file, write a command that gets a list of used protocols every ten minutes with the number of distinct IPs they are addressed to*.

The methodology presented (report and oral review) was devised not only to check if the education objectives introduced in Section III were reached, but also to make them more achievable. Table II relates the educational objectives to the skill areas and oral questions introduced previously.

[1]Registration is required to get access to material but has no other obligations.

TABLE III
DISTRIBUTION OF LEVEL OF DIFFICULTY ENCOUNTERED BY STUDENTS: 0-NO DIFFICULTY, 1-MINOR DIFFICULTIES, 2-CONSIDERABLE DIFFICULTIES, 3-MAJOR DIFFICULTIES (KEY: S.DEV.: STANDARD DEVIATION, C.INT.: CONFIDENCE INTERVAL 95%, STS.: STUDENTS)

	mean	s.dev.	C.int	0	1	2	3	sts.
A1	0.6	0.8	0.2	55%	30%	18%	0%	41
A2	0.5	0.5	0.2	58%	43%	3%	0%	41
A3	1.3	1	0.3	25%	40%	23%	15%	41
A4	1.3	1.3	0.3	28%	25%	38%	13%	41
A5	1	1.1	0.3	45%	20%	25%	13%	41
A6	1	1.1	0.3	45%	25%	20%	13%	41
O1	0.9	0.9	0.3	47%	30%	17%	7%	30
O2	0.8	0.8	0.3	47%	38%	15%	3%	34
O3	0.9	0.9	0.3	35%	41%	18%	9%	34
O4	1	1	0.5	41%	36%	14%	14%	22

B. Evaluation Results

Table III gives a statistical analysis of the level of difficulty (on a scale of 0 to 3) experienced by students in answering the report and the oral evaluation questions. Since oral questions were tailored for every specific team, students were not required to answer a question from each of the expertise fields (which explains that the total number of students—in the far right-hand column—is less than 41). The left-hand column gives the mean of the *level of difficulty* encountered by students in demonstrating they had acquired that particular knowledge.

0: no problems or negligible issues that were not penalized;
1: minor difficulties or a few oversights that do not imply an inadequate understanding of the subject matter;
2: considerable problems that indicate that students did not understand part of the issues under study;
3: major problems indicating that students did not understand the basis of the exercise or did not do the exercise (or part of it).

Table III shows, based on the sample of 41 students, that an average grade of 0.6 ± 0.2 would be expected from a group of students with similar technical background when facing A1 requirements. In general, for all report areas (A) and oral questions

(O), average performance levels were around the acceptable descriptor of *minor difficulties*.

ANOVA tests revealed that there are statistical differences between the various parts of report (A1 to A6),[2] but not between the results for the questions (O1 to O4).[3] Students have more problems in the exercises on the analysis of temporal patterns (A4) and the univariate analysis of the aggregated data (A3). The exercises for A3 and A4 demand abstract reasoning and establishing analogies not previously introduced in class, a more demanding task for students without experience in statistics. Exercises in data preprocessing and aggregation (A2) show a minor rate of difficulties. This may be due to the repetitive nature of this section, where students can check the outcomes, and results are either right or wrong, and not too open to interpretation.

Finally, the high values of standard deviations in Table III (similar to or higher than the mean) reveal that there is a strong variation performance between groups, which coincides with the observations of the instructors. Motivated students who attended the lectures and familiarized themselves with the exercise sheet prior to the first lab class had no problem in passing the lab.

In terms of the objective of acquiring general scientific skills, the last part of the lab, where students searched for suspicious sources themselves, was helpful. Although not all of their attempts at interpretation in their reports were correct, most demonstrated not only data analysis skills but also a reasonable understanding of the use and misuse of network protocols. Students searched additional sources of information to help interpret their findings and were eager to discuss their results, indicating an increased interest in scientific work in the field.

Most students showed encouraging development and were successful in the course. Seven students, however, did not pass the lab due to incomplete or poor quality reports, or poor performance in the lab review. Three of these did not even attempt to take the written test of understanding of the theory; three who did barely passed the test. This suggests that passing the written test should become a prerequisite for attending the lab.

C. Student Feedback

At the end of the semester (in a discussion session) students filled out the anonymous standard TU Vienna evaluation sheet for lectures with labs [22]. For this laboratory class, 14 students returned an evaluation form, so the results as follows only reflect the opinions of these 14 and are not necessarily representative of the whole group of 41 students.

Students were asked to respond to 18 questions, on a scale of 1 (strongly agree) to 5 (strongly disagree). For all 18 statements the average ratings were between 1.07–1.54, indicating that students were quite satisfied with the class.

[2]ANOVA test (single factor, fixed effects) of A1...A6 (using mean, s.dev. and sts. from Table III): $p-\text{value} = 6.3 \times 10^{-5} < 0.05 \rightarrow$ null hypothesis discarded ($\alpha = 0.05$).

[3]ANOVA test (single factor, fixed effects) of O1...O4 (using mean, s.dev. and sts. from Table III): $p-\text{value} = 0.61 > 0.05$ null hypothesis not discarded ($\alpha = 0.05$).

In addition to questions about instructor behavior and class coordination, the evaluation sheet included four questions about the usefulness of the class. Responses included:

1) *"The course raised interest in exploring the topic further"* (average: 1.36);
2) *"Information was provided during the course about how I will be able to use the contents in the future."* (1.54);
3) *"The course increased my knowledge"* (1.36);
4) *"I am capable of using the knowledge I gained from the course."* (1.50).

In addition, students *"enjoyed attending the course"* (1.29) and were *"overall satisfied with the course"*(1.43). Students also stated what they *"particularly enjoyed"*. They explicitly mentioned the lab part (*"practical part", "lab quality", "lab topic"*, etc.) and enjoyed *"working with real data"*. Asked about possible improvements, students suggested more lab time, a permanently open lab, less data preparation, and more emphasis on the free exploration exercises.

VI. Lessons Learned

This section describes lessons learned and planned future improvements.

1) Hold a Trial Session: The lack of major technical problems during the lab can be clearly attributed to the initial trial run, which allowed refinements to be made to the lab settings before actual instruction began.

2) Enforce Prerequisites: In future it will be mandatory to have taken the introductory communication networks course and achieved a passing grade on the network security theory test before registering for the lab. In addition, the lab will include a short TCP exercise to familiarize students with TCP usage, using *Wireshark* [23] to analyze public pcap files [24].

3) Encourage Free Exploration of Data: Student feedback suggests that students respond well to the research-oriented teaching approach. Future versions of the course will expand the opportunities for free exploration tasks.

4) Profit From Student Diversity: Student diversity turned out to be an opportunity rather than a challenge. Computer science students with less experience in signal processing paired with electrical engineering students who had less experience in programming. Such complementary pairs performed quite well in the lab.

5) Offer Flexible Lab Times: Several students asked for additional lab time to finalize reports or further explore data. Future offerings of this course will include more free lab time, remote access to the laboratory, and a booking system to allow students to check on-line if computers are available.

VII. Conclusion

The NetSec-I lab teaches network traffic anomaly detection security methods to electrical engineering students. The lab follows a research-oriented teaching approach and uses real network traffic from a large IP darkspace monitor to evoke student interest in data exploration techniques. The first implementation of the lab was successful in enabling students to gain both the technical and general problem-solving skills required. Students particularly enjoyed the free data exploration exercises, which demonstrated the power of research-oriented teaching concepts.

All exercises, data and supplementary material (scripts, etc.) are available at [1], for use by other instructors.

ACKNOWLEDGMENT

This material represents the position of the authors and not of NSF or DHS. The authors would like to thank V. Bernhardt and D. Frkat, who performed the initial lab trial and provided valuable suggestions. CAIDA provided the darkspace data [25].

REFERENCES

[1] Communication Networks Group of TU Vienna, TU Vienna NetSec Lab. 2014 [Online]. Available: http://www.tc.tuwien.ac.at/netsec-lab
[2] M. Micco and H. Rossman, "Building a cyberwar lab: Lessons learned: Teaching cybersecurity principles to undergraduates," in *Proc. 33rd SIGCSE Tech. Symp. Computer Science Education*, New York, NY, USA, 2002, pp. 23–27, ACM.
[3] P. J. Wagner and J. M. Wudi, "Designing and implementing a cyberwar laboratory exercise for a computer security course," in *Proc. 35th SIGCSE Tech. Symp. Computer Science Education*, New York, NY, USA, 2004, pp. 402–406, ACM.
[4] C. Lee, A. Uluagac, K. Fairbanks, and J. Copeland, "The design of netSecLab: A small competition-based network security lab," *IEEE Trans. Edu.*, vol. 54, no. 1, pp. 149–155, Feb. 2011.
[5] R. Abler, D. Contis, J. Grizzard, and H. L. Owen, "Georgia Tech information security center hands-on network security laboratory," *IEEE Trans. Edu.*, vol. 49, no. 1, pp. 82–87, Feb. 2006.
[6] M. Wannous and H. Nakano, "NVLab, a networking virtual web-based laboratory that implements virtualization and virtual network computing technologies," *IEEE Trans. Learn. Tech.*, vol. 3, no. 2, pp. 129–138, Apr. 2010.
[7] L. Ben Othmane, V. Bhuse, and L. Lilien, "Incorporating lab experience into computer security courses," in *Proc. World Congr. Computer and Information Technology (WCCIT)*, June 2013, pp. 1–4.
[8] I. Marsa-Maestre, E. de la Hoz, J. Gimenez-Guzman, and M. Lopez-Carmona, "Using a scenario-generation framework for education on system and internet security," in *Proc. IEEE Global Engineering Education Conf. EDUCON*, Marrakesh, Morocco, Apr. 2012, pp. 1–7.
[9] A. Brew, *Research and Teaching: Beyond the Divide*. New York, NY, USA: Palgrave Macmillan, 2006.
[10] "The CAIDA UCSD network telescope," [Online]. Available: http://www.caida.org/projects/network_telescope/
[11] "Center for applied internet data analysis," [Online]. Available: http://www.caida.org
[12] E. Aben, Conficker/Conflicker/Downadup as Seen from the UCSD Network Telescope 2009, Tech. Rep. [Online]. Available: http://www.caida.org/research/security/ms08-067/conficker.xml
[13] T. Zseby, A. King, N. Brownlee, and K. Claffy, "The day after patch Tuesday: Effects observable in IP darkspace traffic," in *Proc. Passive and Active Network Measurement Workshop (PAM)*, Hong Kong, China, Mar. 2013, vol. 7799, pp. 273–275.
[14] A. Dainotti, C. Squarcella, E. Aben, K. C. Claffy, M. Chiesa, M. Russo, and A. Pescapé, "Analysis of country-wide internet outages caused by censorship," in *Proc. ACM SIGCOMM Conf. Internet Measurement*, New York, NY, USA, 2011, pp. 1–18, ACM, ser. IMC'11.
[15] "tcpdump and libpcap," [Online]. Available: http://www.tcpdump.org/
[16] A. King, "corsaro v2.1," Oct. 2012 [Online]. Available: http://www.caida.org/tools/measurement/corsaro/
[17] J. Xu, J. Fan, M. Ammar, and S. Moon, "Prefix-preserving IP address anonymization: Measurement-based security evaluation and a new cryptography-based scheme," in *Proc. 10th IEEE Int. Conf. Network Protocols*, Paris, France, Nov. 2002, pp. 280–289.
[18] "The CAIDA acceptable use agreement (AUA) for publicy accessible datasets," [Online]. Available: http://www.caida.org/home/legal/aua/public aua.xml
[19] "MATLAB," [Online]. Available: www.mathworks.de/products/matlab/
[20] "GNU octave," [Online]. Available: www.gnu.org/software/octave/
[21] "Community edition, RapidMiner version 5.3," [Online]. Available: http://rapidminer.com/
[22] "TU vienna evaluation sheet for lectures with labs," TU Vienna [Online]. Available: http://www.tuwien.ac.at/fileadmin/t/tuwien/fotos/lehre/LVABewertung/VUde.pdf
[23] "Wireshark," [Online]. Available: http://www.wireshark.org/
[24] "Public pcap files," [Online]. Available: http://nostarch.com/packet2.htm
[25] "The CAIDA UCSD network telescope educational dataset," CAIDA [Online]. Available: http://www.caida.org/data/passive/telescope-educational dataset.xml

Tanja Zseby (M'07) received the Dipl.-Ing. degree in electrical engineering and the Ph.D. (Dr.-Ing.) degree from Technical University Berlin, Germany.

She is a Professor of communication networks in the Faculty of Electrical Engineering and Information Technology, Vienna University of Technology, Vienna, Austria. Before joining Vienna University of Technology she led the Competence Center for Network Research at the Fraunhofer Institute for Open Communication Systems (FOKUS), Berlin, and worked as a Visiting Scientist at the University of California, San Diego.

Félix Iglesias Vázquez (M'13) was born in Madrid, Spain, in 1980. He received the Ph.D. degree in technical sciences in 2012 from the Vienna University of Technology, Vienna, Austria.

He currently holds a University Assistant position at the Vienna University of Technology. He has worked on fundamental research and project development for diverse Spanish and Austrian firms and lectured in the fields of electronics, physics and automation. His research interests include machine learning, data analysis and network security.

Alistair King received the M.Sc. degree from The University of Waikato, New Zealand, in 2010.

He is a Research Programmer at the Center for Applied Internet Data Analysis (CAIDA), University of California San Diego, La Jolla, CA, USA. His current interests are centered around software and infrastructure development for efficient, real-time analysis of large-scale Internet measurement datasets.

K. C. Claffy received the Ph.D. degree from the Department of Computer Science and Engineering, University of California at San Diego, La Jolla, CA, USA.

She is now an Adjunct Professor at the University of California at San Diego and is Director of the distributed Center for Applied Internet Data Analysis (CAIDA) and resident Research Scientist based at the University of California's San Diego Supercomputer Center. Her research focuses on the collection, analysis, and visualization of workload, routing, topology, performance, and economic data on the Internet.