Flow-Based Intrusion Detection

Anna Sperotto and Aiko Pras
Centre for Telematics and Information Technology
University of Twente
The Netherlands
Email: {a.sperotto, a.pras}@utwente.nl

Abstract—**The spread of 1-10 Gbps technology has in recent years paved the way to a flourishing landscape of new, high-bandwidth Internet services. At the same time, we have also observed increasingly frequent and widely diversified attacks. To this threat, the research community has answered with a growing interest in intrusion detection, aiming to timely detect intruders and prevent damage. We believe that the detection problem is a key component in the field of intrusion detection. Our studies, however, made us realize that additional research is needed, in particular focusing on validation and automatic tuning of Intrusion Detection Systems (IDSs).**

The contribution of this thesis is that it develops a structured approach to intrusion detection that focuses on (i) system validation and (ii) automatic system tuning. We developed our approach by focusing on network flows, which offer an aggregated view of network traffic and help to cope with scalability issues. An interesting approach to validation is the creation of appropriate testbeds, or ground-truth data sets, for which it is known when an attack has taken place. First, we obtained ground-truth information for flow-based intrusion detection by *manually* creating it. The outcome of our research is the first publicly released flow-based labeled data set. Second, we generated ground truth information in an *automatic* manner, by means of probabilistic traffic models based on Hidden Markov Models (HMMs).

Finally, we approached the problem of automatic tuning of IDSs. The performance of an IDS is governed by the trade-off between detecting all anomalies (at the expense of raising alarms too often), and missing anomalies (but not issuing many false alarms). We developed an optimization procedure that aims to mathematically treat such trade-off in a systematic manner, by automatically tuning the system parameters.

I. Motivations

As users, we are day by day increasingly dependent on the Internet. In Fig. 1, we show the world average number of Internet users per 100 inhabitants in 2009, as reported by the International Telecommunication Union (ITU) [11]. Among the developed countries, Internet usage is pervasive. For example, the Netherlands has on average more than 89 Internet users per 100 inhabitants. If we consider that 90% of the Dutch population is between 5 and 80 years old[1], we may say that virtually everybody in a suitable age is nowadays using the Internet. Internauts all over the world are connected at work and at home, and e-mails have almost completely substituted envelopes and stamps. We browse the latest news on our phone and we smile when our favorite café offers free WiFi. Moreover, and for some a bit worryingly, the Internet is the virtual space in which we are nowadays used to managing our money and our personal data. We depend on the Internet not only at the personal level. Internet is by now such a ubiquitous technology that almost any company, university, governmental organization or, even, critical infrastructure such as power plants or water treatment plants, are globally connected.

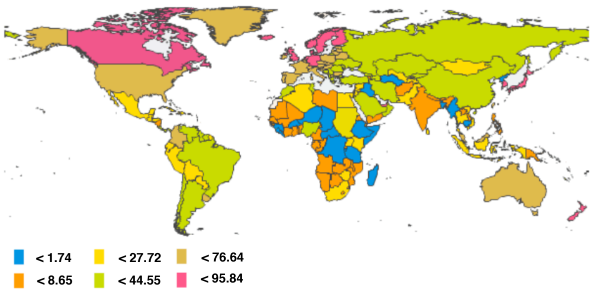

Fig. 1. Percentage of Internet users in 2009 (source: ITU ICT Eye) [11].

At the basis of the situation that we have described, there is certainly the technological push at infrastructure level which we witnessed in the last decade. Nowadays an access speed of 1-10 Gbps is not unusual, and since bandwidth for wired connections is definitely not a problem anymore, more and more high-bandwidth services are being offered to the users.

We should realize, however, that, behind the scenes, this technological improvement has paved the way to new challenges. First, the amount of Internet traffic, as well as the line speed, continues to grow. A university network, for example, reaches traffic averages in the order of hundreds of Mbps, with high activity peaks in the order of Gbps. On backbone networks, as for example Internet2 [12], the USA research and education backbone, the throughput is even higher (see Fig. 2). Such amounts of data need to be managed and monitored, and new strategies to cope with an average load of multiple Gbps have to be developed. Second, the number of attacks does also continue to grow. The reason behind this is in itself very simple: attacks are getting economically more and more profitable. SPAM is an example of this phenomenon. SPAM represents a form of pervasive advertisement, but it is, increasingly often, also a mean to gather personal information. Experts estimate that 90% of the worldwide sent mail messages are SPAM [1] [13], and the phenomenon does not seem to decline.

The combination of increasing network load and attack frequency is challenging if we are aiming to effectively detect

[1] Source: Centraal Bureau voor de Statistiek (Statistic Netherlands), 2010

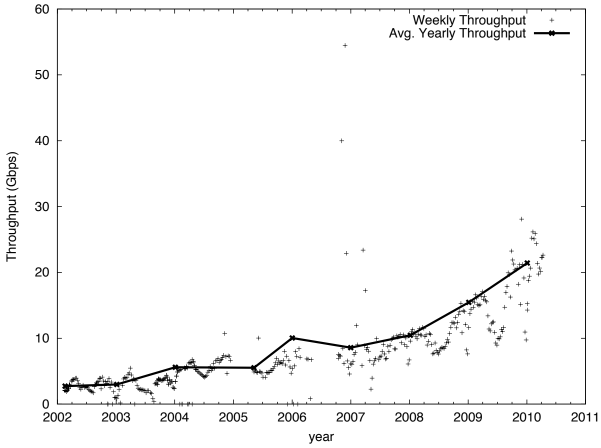

Fig. 2. Network throughput (Gbps) for the network Abilene [12].

intruders. In this thesis, we reacted to the ever growing amount of data by focusing on network flows, rather than individual network packets [14]. A flow is defined as a set of packets that have common properties, as, for example, having the same source and the same destination. Measuring flows offers an aggregated view of traffic information and drastically reduces the amount of data to be analyzed. Flow monitoring is therefore an important approach to cope with scalability issues in IP monitoring. However, from a security perspective, we do not yet see a definite answer to the problem of intrusion detection in situations, as high-speed networks, in which the traditional packet-based solutions may no longer be feasible. Flows therefore appear as a promising approach that may lead to improved results in the field of intrusion detection in high-speed networks.

In the following, we first introduce background on the topics of intrusion detection and flow-based intrusion detection (Sec. II), and then point out what we consider the main open issues in the field (Sec. III). The analysis of such open issues leads us to the presentation of the research questions addressed in this thesis and our approach (Sec. IV). We then present a summary of the contribution of our work (Sec. V).

II. Intrusion Detection

According to Krügel *et al.* [15], "intrusion detection is the process of identifying and responding to malicious activities targeted at computing and network resources". An intrusion attempt, also named *attack*, refers to a sequence of actions by means of which an intruder attempts to gain control of a system.

Since research on intrusion detection started in the 1980s, many flavors of IDSs have been proposed. Traditional IDS taxonomies [16], [17] classify IDS according to several characteristics, e.g., the data the IDS analyzes (log, application or network data), the type of analysis (real-time or offline) or the type of data processing (centralized or distributed). However, the most widely known classification feature regards how an IDS identifies intrusions, i.e., if it is a *misuse-based* or an *anomaly-based* IDS.

A misuse-based IDS, also known as signature-based or knowledge-based IDS, performs detection by comparing new data with a knowledge base of known attacks. An alarm is generated if a previously specified pattern is recognized. The strength of a misuse-based IDS lies in being highly accurate, but its effectiveness depends on the completeness of the signatures. Therefore, a misuse-based system cannot recognize new attacks.

An anomaly-based IDS, also known as behavior-based IDS, compares input data with a *model of normality*, which describes the expected behavior of the monitored system. A significant deviation from the model is marked as an anomaly. The main advantage of an anomaly-based IDS is that it can potentially detect also attacks that have never been seen before.

An IDS aims to discriminate between intrusion attempts and normal activities. In doing so, however, an IDS can introduce classification mistakes, usually known as *false positives* and *false negatives*. There is a natural trade-off between detecting all malicious events (at the expense of raising alarms too often, i.e., having high false positives), and missing anomalies (i.e., having high false negatives, but not issuing many false alarms). Which component of the trade-off is more important is a case-specific decision, and ideally, we would want to optimize both components. We might want to identify all malicious attempts, because this would make our network safer. However, this would be of no use if the number of alerts would overload the IT specialist responsible for handling them.

A. Flow-based Intrusion Detection

An IDS would need to be able to handle the growing number of attacks, the rise in the amount of traffic as well as the increase in line speed [2]. However, researchers assess the payload-based IDSs processing capability to lie between 100 Mbps and 200 Mbps when commodity hardware is used [18], and close to 1 Gbps when dedicated hardware is employed [19]. Well-known systems like, e.g., Snort [20], exhibit high resource consumption when confronted with the overwhelming amount of data found in today high-speed networks [21].

Given these problems, flow-based approaches seem to be promising candidates for intrusion detection research. Flows are monitored by specialized accounting modules usually placed in network routers. These modules are responsible for exporting reports of flow activity to external collectors. Flow-based IDSs will analyze these flows to detect attacks. Compared to traditional IDSs, flow-based IDSs have to handle a considerably lower amount of data. For example, in the case of the University of Twente (UT) network, we calculated that the ratio between packets exported by NetFlow (containing the flow records) and the packets on the network is on average equal to 0.1%. Moreover, considering the network load measured in bytes, the overhead due to Netflow is on average 0.2%. Flow-based intrusion detection is therefore a logical choice for high-speed networks.

The question remains whether flows do carry enough information, compared to packet payload, to be useful for intrusion detection. Flow measurements are by nature aggregated information. They, therefore, do not provide the detection precision of payload-based inspection. However, information on the safety status of a monitored network can be obtained from flows, for example by studying evolutions of network interactions. Flow measurements provide an aggregated view of the data transferred over the network and between hosts, in terms of number of packets, bytes and measured flows themselves. In this context *time series* are a powerful tool to describe network evolution patterns. In network monitoring, time series are usually accepted as the natural way to look at network traffic, i.e., in a streaming manner. The popularity of this approach is reflected by the widespread use of tools like, among others, the Netflow Sensor (NfSen) suite [22]. In this thesis, we combine time series analysis with flow-based intrusion detection. We aim to investigate how it is possible to describe anomalous events by mathematically considering the evolution over time of flows, packets and bytes.

In any case, it is important to underline that flow-based intrusion detection is not supposed to substitute the packet-based one, but rather to complement packets-based methods by allowing early detection in environments in which payload-based inspection is not scalable.

III. Open issues in Intrusion Detection

Flow-based intrusion detection is a relatively new research field, the early contributions of which date back to the first decade of this century [2]. Flow-based intrusion detection builds upon previous experiences in intrusion detection, but faces different challenges, for example the absence of payload. Like payload-based intrusion detection, however, it has a distinct problem-oriented attitude: with the number of attacks almost exponentially increasing [2], and the attackers' motivations moving from ideological to economical, the researchers' attention is focused on developing new techniques to timely detect intruders and prevent damage. Our studies in the field of flow-based intrusion detection, however, made us realize that other research directions are possible, especially if we consider that flow-based intrusion detection is a discipline placed at the intersection between network monitoring and security. Here we point out two problems that, in our opinion, need attention.

First, in the network monitoring community, the importance of sharing network traces for development, validation and comparison purposes is well understood. Examples are repositories like Caida DATA [23], Crawdad [24] and Simpleweb [25]. In intrusion detection, data sets also play a central role, with the difference being that researchers usually evaluate new approaches by testing them on data sets for which the malicious or benign nature of the data is known. We refer to these traces as *ground-truth* data sets. Considering the current situation, research on IDS generally suffers from a lack of shared ground-truth data sets. High-quality ground-truth data sets are time consuming to create and often rely on privacy-sensitive data. Therefore, most publications use non-public traffic traces for evaluation purposes, and we have no knowledge of any publicly available ground-truth flow-based data set. This is generally an obstacle to the comparison of different IDS approaches.

Second, we may certainly say that each new approach to intrusion detection brings us a step forward towards having a safer network. However, we often forget that, for an intrusion detection system to be deployed, we keep "the man in the loop". With this, we mean that, once a new approach has been developed, it is generally assumed that expert IT personnel will take care of the operational aspects in a specific network. This assumption is motivated by the fact that any IDS, to be effective, needs to be tuned according to the specific characteristics of the monitored network. However, this also means that only the expertise of the security operator ensures us that an IDS is tuned to be used in the best way possible. We wonder therefore whether the problem of parameter tuning for IDS could be treated in a more systematic manner, leaving the security expert free to focus on high-level policies (such as maximum allowed false positive rate, or relative importance assigned to false positive and false negative rates) instead of understanding how the IDS works. This reasoning relates to the wider concept of *autonomic management*. Citing Horn [26], who proposed in 2001 the idea of the Autonomic Computing Initiative (ACI), an autonomic system allows "users to concentrate on what they want to accomplish rather than figuring how to rig the computing systems to get them there." Applied to an IDS, this means that we want to achieve the desired (optimal) performance without knowing the details of the underlying system. The task of the security expert would therefore be to specify security policies, and the system should implement these policies to provide detection according to a provided optimality criterion.

IV. Goal and Approach

In light of the reasoning so far, the goal of this thesis is to develop a structured approach to intrusion detection that focuses on (i) system validation, by means of shared ground-truth data sets and (ii) automatic system tuning, by tuning the system parameters. We develop our approach in the context of detecting anomalies using flow data and time series. To achieve this goal, we answer the following research questions:

- **Research Question 1:** What is the state of the art in the field of flow-based intrusion detection?
- **Research Question 2:** Are time series a good approach for intrusion detection at flow level? If yes, how can they be exploited best?
- **Research Question 3:** How can we determine ground-truth information for flow-based intrusion detection?
- **Research Question 4:** How can we tune the parameters of a flow-based IDS based on high-level policies?

Flow-based intrusion detection is a relatively new research field that has only recently attracted the researchers' attention. The objective of Research Question 1 is therefore to identify the main contributions and research trends in flow-based intrusion detection so far. To do this, we perform a literature study that presents a structured overview of the research field.

The starting point for the research in this thesis are real network measurements. Therefore, Research Question 2 aims at gaining domain knowledge about how anomalies look like “in the wild”, in cases where only flow data are available. To answer this question, we perform an extensive data analysis on flow data from the University of Twente and SURFnet, the Dutch national research and education network. We focus, in particular, on how anomalies affect metrics as the number of flows, packets and bytes, in the form of time series.

The goal of Research Question 3 is to shed light on a basic problem that all IDSs as well as other classification systems have in common: the need for labeled data, or *ground truth*. We are not able to evaluate a system if we do not know the nature of the data it has processed. Ground-truth data sets are fundamental in the development phase, for validation purposes and, if publicly available, for comparison between different IDSs. We identify two viable solutions to answer Research Question 3. First, we obtain ground-truth information for flow-based intrusion detection by *manually* creating it. In order to do so, after identifying the requirements a ground-truth data set should meet, we inspect possible network infrastructures suitable for the task. Following the outcome of this phase, we proceed by creating a flow-based labeled data set by tracking malicious activities on a monitoring point that is optimized for collecting security information. Finally, we release the flow-based data set in anonymized form for public use. Our second approach to answer Research Question 3 is to create ground truth in an *automatic* manner. To answer this part of Research Question 3, we model malicious and benign flow information at time series level. We rely upon the well-known framework offered by Hidden Markov Models (HMMs) [27], since they allow a probabilistic, compact representation of (flow-based) time series and they can be used for generation purposes. We validate our approach by verifying that the generated time series statistically approximated the original ones.

Finally, Research Question 4 deals with the topic of tuning the parameters of an IDS according to high-level policies. The general expectation for an IDS is to have a high true positive rate while maintaining a low false positive rate. However, the tuning of the system parameters is typically left to the experience and the skill of IT personnel and little work has been done in proposing structured approaches to address this problem. To answer Research Question 4, we propose an optimization procedure for tuning the parameters of a flow-based, time-series based intrusion detection system. We approach the parameter tuning problem in a probabilistic manner by addressing the trade-off between correct detection and detection errors. We explicitly take into account the fact that the optimal solution may depend on the situation, meaning that it can change according to specific network and user requirements. Finally, we support our approach by extensively validating it on synthetic and original data sets.

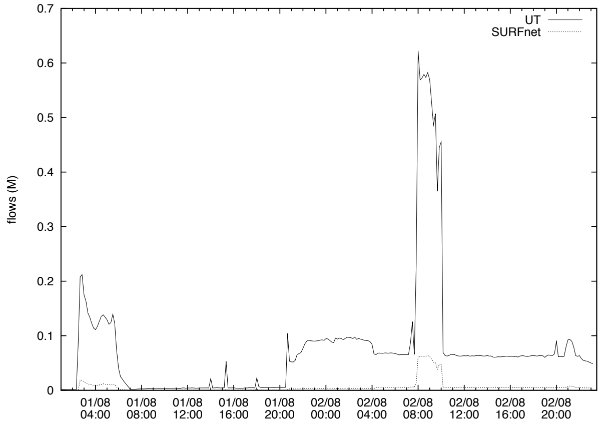

Fig. 3. Flow time series, for UT and SURFnet SSH traffic: the plateau and the peaks indicate a prolonged attack activity.

V. Contribution and impact

We believe that the problem of anomaly detection in flow data should be approached by bringing into focus topics as *validation* and *tuning* of IDSs. Our general conclusion is therefore that the research attention, focused mainly on solving the detection problem by developing new IDSs, should be enlarged to include issues that can be considered as a basis of intrusion detection: publicly available ground-truth data sets and optimal parameter tuning. Considering the trends we observed regarding bandwidth provisioning and traffic load, but also number and intensity of attacks, it is likely that intrusion detection is going to remain a core research area for future networks. We expect that this thesis can bring new awareness in perceiving intrusion detection as a structured group of research areas: not only development of detection systems, but also system validation and tuning (see Fig. 4).

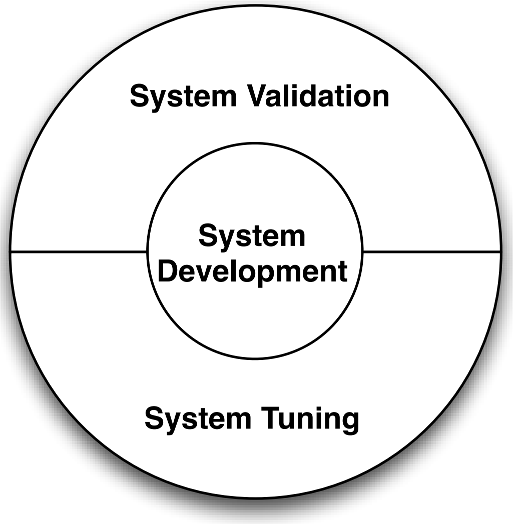

Fig. 4. Intrusion detection as consisting of system development, system validation and system tuning.

Our research provided the following answers to the Research Questions presented in Sec. IV.

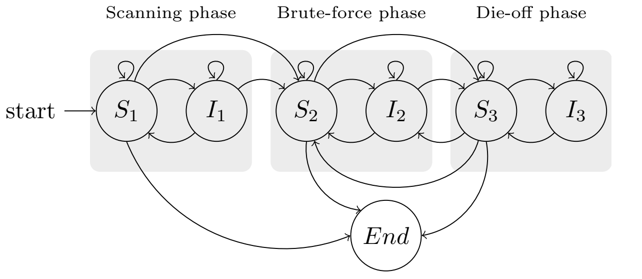

Fig. 5. Model for SSH dictionary attack traffic. The attack evolves through a *scan* phase, followed by a *brute-force* dictionary phase and it ends with a *die-off* phase.

A. Research Question 1: What is the state of the art in the field of flow-based intrusion detection?

In our work, we presented a survey of the state of the art in flow-based intrusion detection [2]. Flow-based intrusion detection is a relatively recent field of research, whose first contributions date back to 2002. Since then, several approaches to the problem of detection have been proposed. By analyzing and categorizing them, we identified the major trends in the field. We concluded that the research efforts are at the moment focused on passive and centralized solutions with, primarily, centralized data collection. Moreover, we also noticed an evenly shared interest between anomaly-based and misuse-based systems and a clear attention for real-time systems.

B. Research Question 2: How can traffic anomalies be characterized in time series derived from flow data?

In our research, we focused on anomaly characterization in time series. In [3], we presented an extensive data analysis on flow-data from the University of Twente and SURFnet, the Dutch national research and education network . Our analysis led us to the following conclusions. First, *time series* of flows, packets and bytes can be a suitable approach to flow-based intrusion detection, since they allow data analysis by keeping into account temporal relations between events, in this case the amount of traffic. Second, to more clearly identify and characterize anomalies, we suggest performing an application-based *traffic breakdown*. With this we mean that anomalies that are not noticeable by considering the whole traffic can be identified by looking at, for example, only SSH or DNS traffic (see Fig. 3). The traffic breakdown can therefore empower flow-based intrusion detection by reducing the amount of data to be analyzed and by facilitating anomalies exposure. This observation suggests that a feasible approach to flow-based intrusion detection should encompass the design of modular intrusion detection systems targeting specific applications. Moreover, the combined analysis of flow, packet and byte time series can strengthen the certainty of the presence of an attack. However, for certain classes of attacks, the choice to monitor only some of the aforementioned metrics can be sufficient.

C. Research Question 3: How can we determine ground-truth information for flow-based intrusion detection?

Our results cover the fields of *manual* and *automatic* ground truth generation.

First, we investigated the creation of a flow-based data set of security-relevant events [4], where each of the attacks has been manually labeled. Building this type of data set can be challenging. Our research covered several aspects of the problem, namely which requirements should such a data set meet, which infrastructure is suitable for data collection and, finally, how the collected data can be labeled. Our findings showed that the most promising measurement setup among the analyzed ones is monitoring a single host with enhanced logging capabilities. The collected information permitted us to create a database of both flows and security events (derived by the logs). However, we are aware of the limitations that our approach entails, in particular, the fact that the collected trace mainly consists of malicious traffic. As results of our research, we built and publicly released a flow-based labeled data set. At the best of our knowledge, our effort constitutes the first publicly available labeled flow-based traffic trace. The data set is available in anonymized form at the address: *http://traces.simpleweb.org*. Despite this favorable outcome, the lesson learned is that, although we limited our experiments to a single host, labeling remains a complex task that requires human intervention.

In our work, we also investigated the possibility of generating ground-truth information in an automated manner [5]. We proposed a modeling approach for flow-based traffic time series based on HMMs. We showed that the models that we developed provided a compact representation of the traffic, where only few states are needed to fully describe the traffic evolution. Moreover, HMMs can be used for generative purposes, allowing us to create time series for which the ground truth is known. The models we proposed are inferred by studying real SSH traffic time series for both attack and normal traffic, captured at the University of Twente. Fig. 5 shows the model describing an SSH dictionary attack. The research in the thesis showed that: (i) our HMM-based approach can capture the main statistical characteristics of the original time series; (ii) by qualitative investigation, our approach can reproduce time series that resemble the real traffic; and (iii) as far as ground truth is concerned, we are now able to create labeled data sets in an automated manner.

D. Research Question 4: How can we tune the parameters of a flow-based IDS based on high-level policies?

The performance of an IDS is governed by the trade-off between false positives and false negatives. We proposed a mathematical framework, in terms of an optimization procedure, to treat such trade-off in a systematic manner. We tested our procedure on a *history-based*, *probabilistic* and *anomaly-based* detection system. Such a simple model became our demo system. We then proposed a probabilistic optimization procedure that, by solving a non-linear optimization problem, tunes the parameters of the detection system such as the history

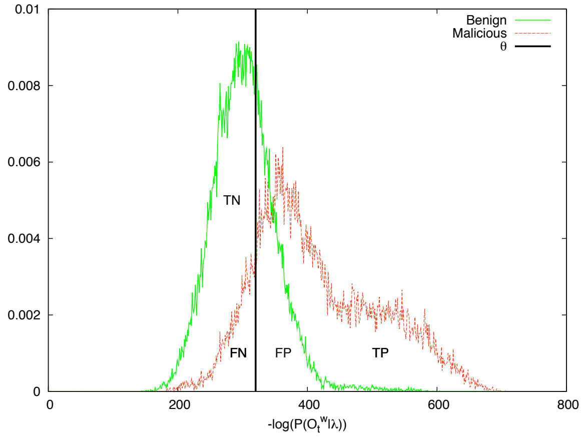

Fig. 6. Automatic system tuning: calculate the optimal threshold θ that maximizes TN and minimizes FP, given a probabilistic representation of benign and malicious traffic (window size 100).

length and the alert threshold. The goal of the procedure was to maximize the correct detection (*true negative*) and minimize the errors (*false positive*) (see Fig. 6). Since the proposed procedure explicitly regarded optimality according to the high-level policies, we studied the impact of such policies on the overall performance of the system.

VI. Thesis Material

The thesis can be downloaded from *http://wwwhome.cs.utwente.nl/~sperottoa/research.html*. The page also contains the list of publications concerning the context of this thesis.

Acknowledgment

This research has been supported by the EC IST-EMANICS Network of Excellence (#26854) and the EU FP7-257513 UniverSelf Collaborative Project.

Author's Publication List

[1] A. Sperotto, G. Vliek, R. Sadre, and A. Pras, "Detecting Spam at the Network Level ," in *Proc. of the 15th Open European Summer School and IFIP TC6.6 Workshop (EUNICE '09)*, 2009, pp. 208–216.

[2] A. Sperotto, G. Schaffrath, R. Sadre, C. Morariu, A. Pras, and B. Stiller, "An Overview of IP Flow-Based Intrusion Detection," *IEEE Communications Surveys & Tutorials*, vol. 12, no. 3, pp. 343–356, 2010.

[3] A. Sperotto, R. Sadre, and A. Pras, "Anomaly Characterization in Flow-Based Traffic Time Series ," in *Proc. of the 8th IEEE International Workshop on IP Operations and Management (IPOM '08)*, 2008, pp. 15–27.

[4] A. Sperotto, R. Sadre, D. F. van Vliet, and A. Pras, "A Labeled Data Set For Flow-based Intrusion Detection," in *Proc. of the 9th IEEE International Workshop on IP Operations and Management (IPOM '09)*, 2009, pp. 39–50.

[5] A. Sperotto, R. Sadre, P. T. de Boer, and A. Pras, "Hidden Markov Model modeling of SSH brute-force attacks," in *Proc. of the 20th IFIP/IEEE International Workshop on Distributed Systems: Operations and Management (DSOM '09) (**Best Paper Award**)*, 2009, pp. 164–176.

[6] R. J. Hofstede, A. Sperotto, T. Fioreze, and A. Pras, "The Network Data Handling War: MySQL vs. NfDump," in *Proc. of the 16th EUNICE/IFIP WG 6.6 Workshop on Networked Services and Applications*, 2010, pp. 167–176.

[7] A. Pras, R. Sadre, A. Sperotto, T. Fioreze, D. Hausheer, and J. Schoenwaelder, "Using NetFlow/IPFIX for Network Management," *Journal of Network and Systems Management*, vol. 17, no. 4, 2009.

[8] T. Fioreze, L. Granville, A. Pras, A. Sperotto, and R. Sadre, "Self-Management of Hybrid Networks: Can We Trust NetFlow Data?" in *Proc. of the 11th IFIP/IEEE International Symposium on Integrated Network Management (IM '09)*, 2009, pp. 577–584.

[9] A. Sperotto and R. van de Meent, "A Survey of the High-Speed Self-Learning Intrusion Detection Research Area," in *Proc. of the First International Conference on Autonomous Infrastructure, Management and Security (AIMS '07)*, 2007, pp. 196–199.

[10] A. Sperotto and M. Pelillo, "Szemeredi's Regularity Lemma and Its Applications to Pairwise Clustering and Segmentation," in *Proc. of the 6th International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR '07)*, 2007, pp. 13–27.

References

[11] International Telecommunication Union, "Ict statistics," http://www.itu.int/ITU-D/icteye/, Jan. 2010.

[12] Internet 2, "Internet 2 research network," http://www.internet2.edu/, Jan. 2011.

[13] W. van Wanrooij and A. Pras, "Filtering Spam from Bad Neighborhoods," *International Journal of Network Management (accepted for publication)*, 2010.

[14] B. Claise, "Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of IP Traffic Flow Information," RFC 5101 (Proposed Standard), 2008.

[15] C. Kruegel, F. Valeur, and G. Vigna, *Intrusion Detection and Correlation: Challenges and Solutions*. Springer-Verlag Telos, 2004.

[16] H. Debar, M. Dacier, and A. Wespi, "A revised taxonomy for intrusion detection systems," *Annales des Telecommunications*, vol. 55, no. 7–8, pp. 361–378, 2000.

[17] S. Axelsson, "Intrusion detection systems: A survey and taxonomy," Chalmers Univ., Tech. Rep. 99-15, 2000.

[18] M. Gao, K. Zhang, and J. Lu, "Efficient packet matching for gigabit network intrusion detection using TCAMs," in *Proc. of 20th Int. Conf. on Advanced Information Networking and Applications (AINA'06)*, 2006, pp. 249–254.

[19] G. Vasiliadis, S. Antonatos, M. Polychronakis, E. P. Markatos, and S. Ioannidis, "Gnort: High Performance Network Intrusion Detection Using Graphics Processors," in *Proc. of the 11th Int. Symp. on Recent Advances in Intrusion Detection (RAID '08)*, 2008, pp. 116–134.

[20] M. Roesch, "Snort, intrusion detection system," http://www.snort.org, Jan. 2011.

[21] H. Dreger, A. Feldmann, V. Paxson, and R. Sommer, "Operational experiences with high-volume network intrusion detection," in *Proc. of the 11th ACM Conf. on Computer and Communications Security (CCS '04)*, 2004, pp. 2–11.

[22] P. Haag, "Nfsen: Netflow sensor," http://nfsen.sourceforge.net, Jan. 2011.

[23] The Cooperative Association for Internet Data Analysis, "CAIDA DATA," http://www.caida.org/data, Jan. 2011.

[24] CRAWDAD, "Community Resource for Archiving Wireless Data At Dartmouth," http://crawdad.cs.dartmouth.edu/, Jan. 2011.

[25] Simpleweb, " Trace repository," http://traces.simpleweb.org, Jan. 2011.

[26] P. Horn, "Autonomic computing: IBM's Perspective on the State of Information Technology," http://www.research.ibm.com, 2001.

[27] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.