Anomaly Detection and Mitigation at Internet Scale: A Survey

Jessica Steinberger, Lisa Schehlmann, Sebastian Abt, and Harald Baier

da/sec - Biometrics and Internet Security Research Group,
Hochschule Darmstadt, Darmstadt, Germany
{Jessica.Steinberger,Lisa.Schehlmann,Sebastian.Abt,Harald.Baier}@h-da.de

Abstract. Network-based attacks pose a strong threat to the Internet landscape. There are different possibilities to encounter these threats. On the one hand attack detection operated at the end-users' side, on the other hand attack detection implemented at network operators' infrastructures. An obvious benefit of the second approach is that it counteracts a network-based attack at its root. It is currently unclear to which extent countermeasures are set up at Internet scale and which anomaly detection and mitigation approaches of the community may be adopted by ISPs. We present results of a survey, which aims at gaining insight in industry processes, structures and capabilities of IT companies and the computer networks they run. One result with respect to attack detection is that flow-based detection mechanisms are valuable, because those mechanisms could easily adapt to existing infrastructures. Due to the lack of standardized exchange formats, mitigation across network borders is currently uncommon.

Keywords: Anomaly Detection, Anomaly Mitigation, Internet Service Provider, Network Security, NetFlow, Correlation.

1 Introduction

Network attacks pose a significant problem to the Internet landscape, which causes substantial financial losses. [1] distinguish methods for attack detection according to their *detection methodology*, their *locality* and the *dataset* they use. The *detection methodology* is classed as either signature-based or anomaly-based [2,3]. Obvious disadvantages of a signature-based approach are the need for up-to-date signatures and the restriction to detect only previously known attacks. The anomaly-based technique, on the other hand, searches for suspicious behavior and so it is also possible to detect new attacks. The *locality* is divided in host-based and network-based approaches [4]. To enforce the host-based method, access to the devices of end-users is needed. This poses some problems, e.g. due to the *bring your own device* concept or due to end-users who do not make use of host-based techniques or do not keep them up-to-date. Even there is an increase of new platforms, such as mobile phones, where possibly no existing host-based approach is available yet. A network-based approach on the contrary provides

G. Doyen et al. (Eds.): AIMS 2013, LNCS 7943, pp. 49–60, 2013.

both a global view and global administration, which makes an event correlation easier. Finally, detection may be performed on different *datasets*. As of today common *datasets* for the network-based methods are raw packet data, NetFlow data or system log files.

A study performed by [5] shows that Internet Service Provider (ISP) networks are considered to be key points for botnet mitigation which is one important aspect of attack detection and mitigation. In order to leverage this key position of ISPs in detection and mitigation of cyber-criminal activities we assume that a network-based anomaly detection system for detecting anomalous events has to be placed at an ISP node. So there is the possibility for correlating events for a better knowledge of isolated anomalous events for detecting distributed attacks, such as shown in [6].

Recently the network security scientific community discusses the advantages of network-based anomaly detection on base of NetFlow data [7]. NetFlow is more feasible at Internet scale as e.g. raw packet data, because it is created by packet forwarding and preserves users' privacy. [8] and [9] propose a NetFlow-based detection mechanism for detecting botnets at large-scale networks.

To sum up an important defense strategy against the underground economy is to implement flow-based anomaly detection algorithms at ISP nodes and to exchange status information with third parties. But will such an approach be adopted by ISPs? Do ISPs share and exchange status information with other providers on base of a standardized format? To get insight in real-world processes, structures and capabilities of IT companies and the computer networks they run, we set up a questionnaire of 56 questions, which was answered by 135 respondents from ISPs and other network operators.

The paper is organized as follows. In Section 2 we describe the setup of our survey. The result set is analyzed and evaluated in Section 3. In Section 4 the paper is concluded and future research problems are discussed.

2 Survey Description

The survey[1] addresses ISPs and network operators. It consists of 56 questions related to 6 categories. These categories adhere a number of questions and are listed in Table 1.

We distributed our survey over several relevant mailing lists. The most important are

- European IP Networks forum RIPE, `http://labs.ripe.net`
- German Network Operators Group DENOG, `http://www.denog.de`
- Association of the German Internet Industry, `http://international.eco.de`
- DE-CIX competence group security, `http://www.de-cix.net`
- Swiss Network Operators Group SwiNOG, `http://www.swinog.ch`
- North American Network Operators Group NANOG, `http://www.nanog.org`
- Competence Center for Applied Security Technology, `http://www.cast-forum.de`

[1] `http://www.dasec.h-da.de/wp-content/uploads/2013/02/SurveyOnNetworkAttackDetectionAndMitigation.pdf`

Table 1. Overview of the survey

Category	# of questions	# of complete category answer set	# of complete answers on average	
			1. Level	*2. Level*
Company and personal info	9	3 out of 9	74	17
Attacks and threats	5	2 out of 5	87	-
Data and tools	17	8 out of 17	47	26
Mitigation and reaction	11	4 out of 11	69	10
Role of ISPs and IXPs	9	2 out of 9	45	23
Contact information	5	0 out of 5	12	-

We provided an online system to collect the answers over a time period of two weeks and got 135 participants. However, 88 of the 135 data sets are somehow incomplete, because the respondents decline to give the requested information or abort the survey before completion. The third column of Table 1 provides an aggregated overview, how many questions in each category are completely answered by all 135 participants. The last column pair of Table 1 displays the number of participants on average, who answered so called level 1 and 2 questions. Level 1 denotes questions that were available for all attendees, whereas level 2 refers to follow-up questions. To handle the incompletely answered questions, we proceed as follows. If the question belongs to the first or last category, no further data preparation is necessary. In case of the remaining 4 categories we simply scale the reference point down from 135 to the number of actual answers, which yields a distinct size of these result sets. As each question could be analyzed isolated with regard to their cross-connections, there is no distortion in our results.

A total of 67 respondents have submitted valuable data concerning their geographic provenance and their business segment. Our respondents origin from Europe, America and Africa. Their market segment may be Carrier/Telco/ISP, Cloud Service Provider, Enterprise, Hosting/Data Center/Colocation Service Provider, Research and Education Network or other.

Figure 1 visualizes the distribution of our anonymous participants by four characteristics in a treemap. A treemap is used to visualize multidimensional, hierarchical data and their relationships. The size of each rectangle is proportional to the number of times a certain combination occurred. Our four characteristics are geographic region, market segment, role of employee (who answered the questionnaire) and finally the monthly average traffic transport (denoted as x). As shown in Figure 1 the majority of the participants are headquartered in Europe and classified their company as Carrier/Telco/ISP. Most of these respondents transport on average more than 100 Gbit per second. As the majority of our participants reside in Europe, our results shall be valid at least for Europe.

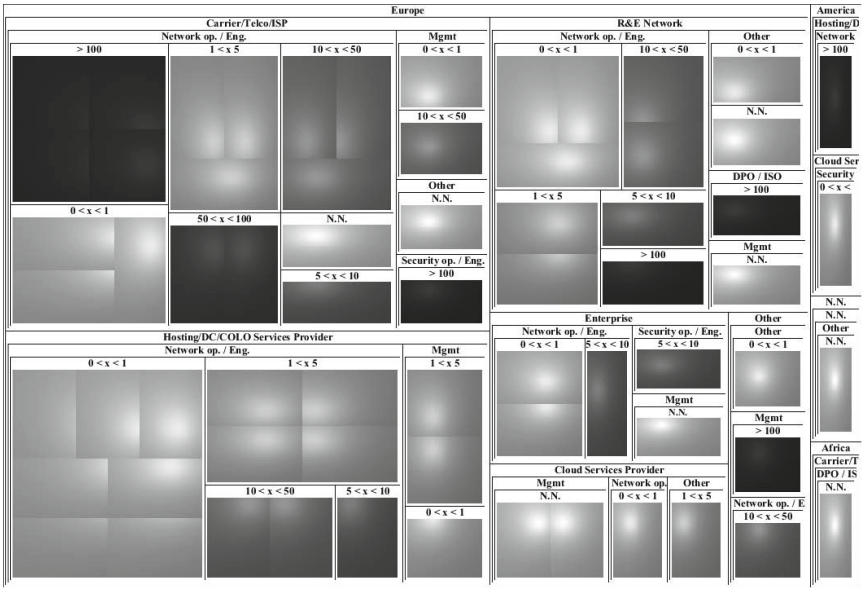


Fig. 1. Geographic and business segment information of our participants

Table 2. Overview of the compliance to common standards

Standard	ITIL	COBIT	ISO 27000	German Grundschutz
Compliance rate	8%	1%	9%	1%

3 Result Set Analysis and Evaluation

In this Section we present the main results of our survey and discuss their respective relevance. Section 3.1 provides information about the compliance of standards and frameworks in the context of ISPs. Section 3.2 shows our results how ISPs rate the risk of common threats and what raises their awareness. In order to assess the feasibility of future detection approaches, Section 3.3 identifies techniques and data that are available for detecting anomalous events. Section 3.4 shows that currently no mitigation with respect to third-parties is implemented and no standardized exchange formats are in use. Finally we discuss in Section 3.5 the self-assessment of providers about their role in network defense.

3.1 Compliance to Security Standards and Frameworks

Being compliant to established standards and frameworks is a common approach to enhance security, however a minority of the responding ISPs actually makes use of them (see Table 2).

In the area of IT Service Management the best practice library called IT infrastructure library (ITIL) is widely known and established [10]. ITIL describes a

process called Information Security Management (ISM), which focuses on alignment of IT security with business security and ensures that information security is effectively managed in all service and management activities (e.g. with respect to the classical security goals information availability, confidentiality, integrity, authenticity). Solely 8% of 135 participants adhere ITIL.

A framework for governance and management of enterprise is called Control Objectives for Information and Related Technology (COBIT) . COBIT provides a document called COBIT Security Baseline, which covers security in addition to all the other risks that can occur with the use of IT. COBIT is only implemented by 1% of our participants.

The standard series ISO/IEC 27000 [11] provides best practice recommendations on ISM, risks and controls within the context of an overall ISM system. Like ITIL only 9% of the respondents adhere to the standards ISO/IEC 27000. Finally, the *IT-Grundschutz* from the German Federal Office for Information Security (BSI) is used by only 1% of the respondents. It uses a holistic approach in various catalogues.

To sum up, providers do not comply to common security standards. We assume that there are two main reasons for the absence of security standard compliance of ISPs. First, as shown in Section 3.5 ISPs do not see a financial incentive to do so. Second the standards only provide a coarse-grained view of IT security related issues, which does not fit well to the segment of ISPs. Especially to the best of our knowledge there is no well-defined process model to detect and mitigate anomalous events in network traffic. Having said that we think if such a process model existed, more ISPs would spend resources to adopt these processes to their business and therefore support the detection and mitigation of network attacks. Hence standardized network defense models at ISP level including detection and mitigation is needed.

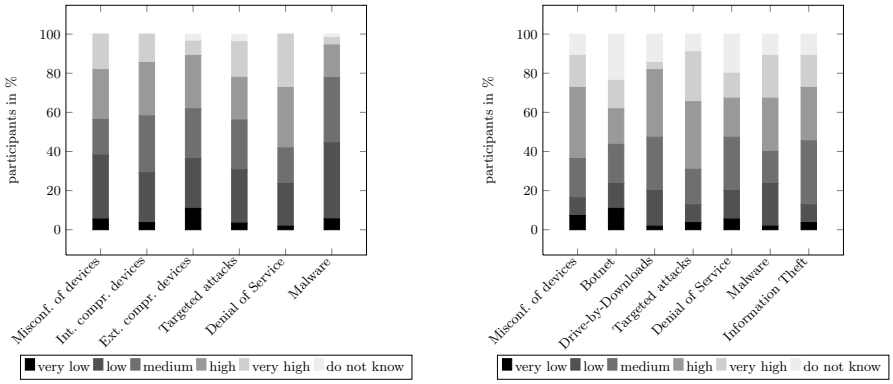
3.2 Attacks and Threats

The results of our survey given in this Section are twofold: First we show which information sources are used by ISPs to keep up-to-date and to raise their security awareness. Second we present results about actual detected attacks and threats against their networks.

With respect to awareness of common threats we are interested to know about reasons that raise the awareness of the 135 participants. As expected the most common source is an attack to the ISP's or its customers' infrastructures, respectively, namely in each case 24%. Presentations and discussions at conferences are close behind with 19%. They are followed from publications in journals, magazines, websites and mailing lists and used by 18%. Legal and regulatory requirements are an insignificant source and only used by 12%. The result shows that besides incidents in their particular networks publications at conferences or in magazines are used and an important way to raise the awareness.

Concerning information sources, the most important way to inform about new attacks and threats are websites, blogs and feeds with 28%, followed from mailing lists with 27%. Security conferences are only marginal relevant and only

used by 12%. The same holds for scientific publications which are relevant for 9%. This result can be attributed to the fact that ISPs need a fast and pragmatic solution to detect and mitigate an attack. Incident information is spread much faster via non-reviewed channels such as web sites or mailing lists. We assume that there is a great challenge and a demand for close collaborations of ISPs and the security research community, so that both parties could benefit from each other's knowledge and experiences to reach expedient results in this area.



(a) pose to company's infrastructure **(b)** pose to company's customer

Fig. 2. Risk assessment of threats

Next we turn to results about actual attacks against the ISPs infrastructures. For 49% of our 54 participants answering this question the number of detected attacks per month is at most 10 and thus rather low. On the other hand 9% detect at least 500 attacks per month to their or their customers' infrastructure. The risk assessment as shown in Figures 2a and 2b) reveals that common threats only pose a *very low* or *low* risk to ISP's or their customers' infrastructure, respectively. Denial of Service attacks pose the most common threat to the ISP's infrastructure (risk is *high* or *very high*), which accords to the result of the Arbor security report 2012 [12]. On the other hand at their customers' infrastructure the most widespread risk is a targeted attack.

To summarize the results of this Section, the awareness of threats should be raised, especially on base of scientific results. This could e.g. be achieved by publications of every kind. Furthermore we see a demand for close collaborations between industry and security research community to ensure a fast exchange of experience and knowledge to gain purposeful results.

3.3 Data and Tools

As stated in the introduction the locality at an ISP node offers great possibilities for real-time detecting and correlating anomalous events. In this Section we

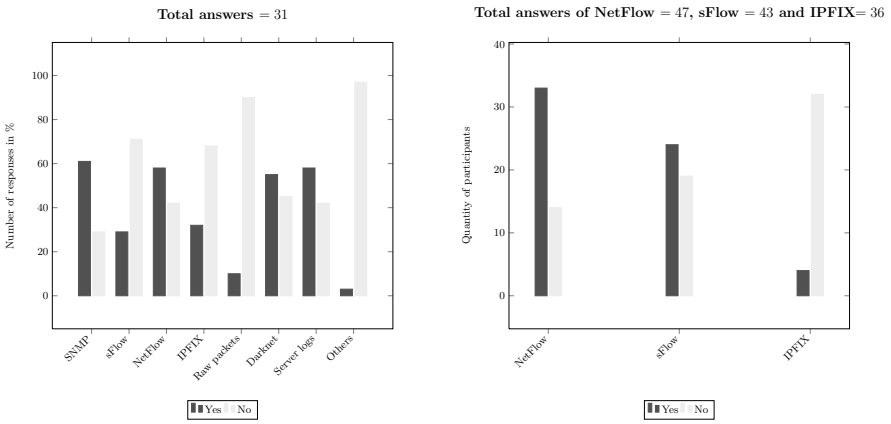
provide the results of our survey with respect to acquired data and tools to detect attacks.

In Section 1 we discussed different kinds of data sources for anomaly detection. We are interested, if flow data is available for anomaly detection. Once again we consider this to be important to assess the feasibility of current scientific anomalous detection approaches, especially the promising algorithms based on network flow data [8,9,1]. Flow data contains statistical network information about a unidirectional data stream between two network devices in a certain time frame (e.g. source/destination IP address, source/destination port etc.). There are different network flow formats. The common ones are NetFlow [7] developed by Cisco, its successor IPFIX (Internet Protocol Flow Information Export, [13]), and sFlow [14].

Figure 3a shows the results to the question, which kind of data the companies currently use for attack detection. The number of responses is 31. The majority of 61% actually use SNMP data, a protocol for exchanging management information between network devices. SNMP is just like NetFlow a passive measurement technology, however, NetFlow provides the advantage of containing more detailed information. So, also shown in Figure 3a, SNMP data is closely followed by NetFlow data and other server logs, namely in each case 58%. Additional flow formats like sFlow and IPFIX are used by 29% and 32% of the attendees. On the other hand only a small minority of 10% make use of raw packet data for the anomaly detection.

The next questions address the technical ability to collect the three common flow data formats. The outcome is illustrated in Figure 3b. Concerning NetFlow (version 5 or version 9) 33 of the 47 participants and hence 70% answering this question provide this possibility. The availability to collect sFlow is given by 24 of 43 responding participants, i.e. 56%. However, only 4 of 36 replying attendees (corresponding to 11%) are able to collect IPFIX data with the current company's infrastructure. But IPFIX is much newer than NetFlow, which perhaps explains this fact.

Finally we aim at comparing flow-based algorithms to the well-known deep packet inspection. We first asked for the technical ability to perform a deep packet inspection, i.e. to collect raw packet data. Although 73% of the 49 responding participants have the ability to do that, only 50% of them think that this is a feasible approach. Their main argument against collecting raw data is the huge amount of network traffic to process. Furthermore 56% of them think, that raw packet data endangers the customers' privacy and requires too much human resources. Further mentioned disadvantages of deep packet inspection are the financial investment (44%) and a prohibition by legal or regulatory requirements (44%). To our mind flow data is privacy friendly. To support this claim we asked the participants if collecting and processing NetFlow data is superior in protecting the customers' privacy to collecting and processing raw packet data. 63% of the 41 respondents agree and 37% disagree to this statement. In summary, flow-based data sources, such as NetFlow, are common, available and privacy-friendly data sources at network nodes. They thus present techniques for



(a) Currently used data **(b)** Technical ability to collect data

Fig. 3. Data used for attack detection

detecting anomalous events in networks. These results support our assumption that there is a demand for network-based anomaly detection systems based on NetFlow data.

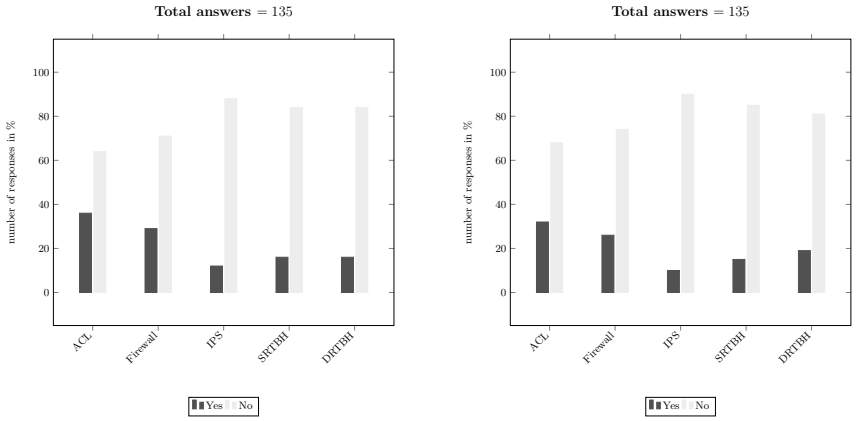
3.4 Mitigation and Reaction

Although incident response requires mitigation and reaction a significantly lower proportion of publications related to this topic is published in the last years compared to detection and correlation approaches. We aim at contributing to the current state of mitigation and reaction processes at network operators.

Our first outcome addresses the time to respond to an attack. Nearly 50% of 43 respondents are able to initially mitigate attacks within 20 minutes. To completely resolve an attack the majority requires up to one day. This is consistent to the results of [12].

Our next question inquires about measures implemented to mitigate an attack. Figure 4a depicts the results if the attack targets the operator's infrastructure itself. Currently 36% of 135 attendees use access-control lists and 29% use a firewall to mitigate the attack. Intrusion prevention systems (IPS), source-based remote-triggered blackhole (SRTBH), and destination-based remote-triggered blackhole (DRTBH) only play a minor role. Figure 4b shows the distribution of the measures used to mitigate network attacks targeting the company's customer. The results are similar to the outcome of the company's infrastructure itself. Again our results are in accordance with the results reported in [12].

Early warning systems require incident information sharing with external third parties (e.g. customers, vendors, competitors, CERTs). However, if an attack is observed the majority of 67% of the 46 respondents do not share attack information. Additionally the 15 attendees sharing information do not use a standardized format for automated data exchange. 13 of them exchange



(a) Targeting company's infrastructure

(b) Targeting company's customers

Fig. 4. Measures used to mitigate network attacks

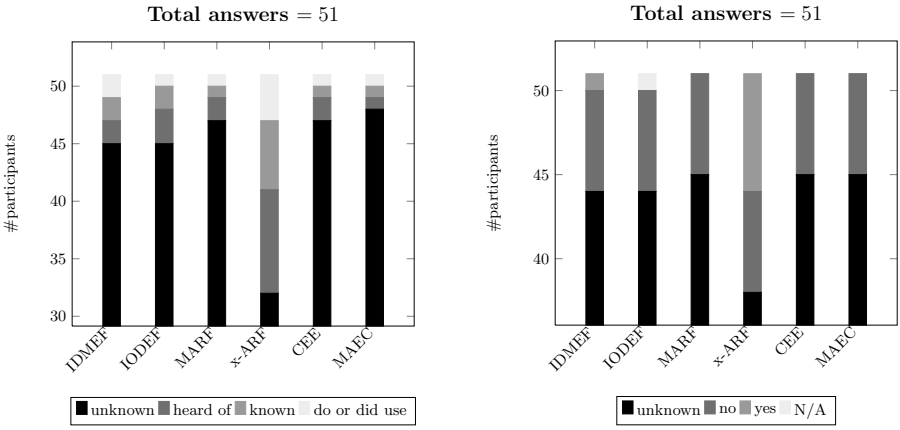
Table 3. Overview of event exchange formats

Name	Abbr.	Responsible
Intrusion Detection Message Exchange Format	IDMEF	IETF
Incident Object Description Exchange Format	IODEF	IETF
Messaging Abuse Reporting Format	MARF	IETF
Extended Abuse Reporting Format	x-ARF	eco
Common Event Expression	CEE	Mitre
Malware Attribute Enumeration and Characterization	MAEC	Mitre

incident information via email, 8 via telephone and 2 automated by a proprietary detection system. Thus no early warning takes place.

A possible explanation is the absence of a well-developed and adopted standardized exchange format for security events/incidents as mentioned in [15]. Several efforts to standardize a number of different exchange formats failed. The reasons therefore can be summarized into three main categories: lack of data of interests, difficulty to handle the information by humans and/or machines, and finally the time of development. Nevertheless effective early warning requires automated and standardized information exchange. The community came up in the past with exchange formats, which are listed in Table 3.

Each exchange format has its own focus and provides special possibilities to exchange attack related information. We asked if the exchange formats are known and to which extent they are used. Figure 5a depicts the distribution of exchange formats, currently in use or known by our participants. On average 86% of 51 respondents do not know about the existence of the 6 exchange formats. Furthermore x-ARF is the most known or used exchange format. Figure 5b shows the future plans of the attendees with respect to incident exchange formats.



(a) Currently used exchange formats **(b)** Used exchange formats in future

Fig. 5. Distribution of used exchange formats

The majority of them do not focus to establish the usage of exchange formats. In fact a few of the participants plan to use some of the mentioned exchange formats. Only the exchange format x-ARF shows an increased usage.

To sum up we suppose that exchanging attack detection data with third parties supports a faster detection and mitigation process. The approach to exchange network attack data with third parties requires a standardized and accepted exchange format, which is able to be used in conjunction with NetFlow data. Although x-ARF is a candidate for that purpose, a lot of convincing has to be done to establish an effective early warning system.

3.5 Role of ISP and IXP

In the last category we ask questions about the subjective view of the role of an ISP and an Internet exchange point (IXP) in network attack detection and mitigation. 94% of 48 respondents answer that an ISP plays an important role in network attack detection and mitigation. Significantly less, namely 69% agree with this opinion related to the role of an IXP.

Even though 62% of 37 attendees think that there is a financial incentive to perform network attack detection and mitigation for ISPs, the remaining respondents are convinced that network security for an ISP is a loss center in a low margin industry. While 54% of 48 of the participants agree that the task of detection and analysis as well as coordination (46%) could be realized at IXP level, only 40% believe that mitigation / response might be a task of an IXP. We assume that the correlation of network security events at network operator level support the detection and mitigation of anomalies, but only 29% of the respondents consider IXP to be responsible for correlation.

Moreover, 42% of 45 respondents are convinced that ISPs shall protect their customers from Internet attacks and 58% agree to protect the Internet from

attacks originating from its costumers. Alongside 27% added a comment that the appropriate way of thinking should include both perspectives. However, it remains the issue of accounting this security add-on. In particular, it was a widespread opinion that removing attack traffic reduces network traffic which reduces how much traffic they can charge their customers for. Hence, in their own perception there is no incentive for ISPs or IXPs to implement security measures. As addressed by [16] we also assume that existing peering agreements might include security as an aspect of their service level agreement. Therefore the network operators should be interested in fulfilling these agreements.

4 Conclusion and Future Work

In this paper we present and discuss the results of a survey, which aims at gaining insights in industry processes, structures and capabilities of IT companies and the computer networks they run. We formulated questions, which cover six categories. Our findings are that most of the participants do not use well-defined processes or standards for security management. Our assumption is, that there is no existing one, which fits to the network operators' business model.

Furthermore we gain knowledge about currently used data sources and detection tools for detection and mitigating anomalies. An important outcome is that NetFlow data is a common approach and an available data source. Sampling raw packets is not considered to be a practical approach. Additionally we reveal that the lack of an accepted exchange format prevents to establish effective early warning systems.

The majority of participants thinks that ISPs play an important role in detecting and mitigating anomalies. This supports our assumption to promote detection algorithms, which fit the requirements of an ISP node.

In the future there is a need for network-based anomaly detection solution based on NetFlow data. Such a solution should be published as open source and well-documented, so that ISPs could easily adapt this appliance to their special needs without investigating too much money and operational time. Appliances used in different ISP networks should be able to collect and correlate security events with each other for better detecting network anomalies.

Acknowledgment. This work was partly supported by the German Federal Ministry of Education and Research under grant number 16BY1201F (iAID) and by CASED.

References

1. Abt, S., Baier, H.: Towards efficient and privacy-preserving network-based botnet detection using netflow data. In: Proceedings of 9th International Network Conference, INC 2012, Port Elizabeth, South Africa (July 2012)
2. Maryam, F., Alireza, S., Sureswaran, R.: A Survey of Botnet and Botnet Detection. In: Proceedings of the 2009 Third International Conference on Emerging Security Information, Systems and Technologies, SECURWARE 2009, Washington DC, USA (2009)

3. Jing, L., Yang, X., Kaveh, G., Hongmei, D., Jingyuan, Z.: Botnet: classification, attacks, detection, tracing, and preventive measures. EURASIP Journal on Wireless Communications and Networking (February 2009)
4. Karen, S., Peter, M.: SP 800-94. Guide to Intrusion Detection and Prevention Systems (IDPS). Technical report, National Institute of Standards & Technology, Gaithersburg, MD, United States (February 2007)
5. van Eeten, M., Bauer, J.M., Asghari, H., Tabatabaie, S., Rand, D.: The Role of Internet Service Providers in Botnet Mitigation: An Empirical Analysis Based on Spam Data. In: The Tenth Workshop on the Economics of Information Security, WEIS 2010 (2010)
6. Prez, M.G., Mrmol, F.G., Prez, G.M., Gmez-Skarmeta, A.F.: RepCIDN: A Reputation-based Collaborative Intrusion Detection Network to Lessen the Impact of Malicious Alarms. Journal of Network and Systems Management 21(1) (March 2013)
7. Cisco Systems, Inc.: Netflow services solutions guide (January 2007), http://www.cisco.com/en/US/docs/ios/solutions_docs/netflow/nfwhite.html
8. François, J., Wang, S., State, R., Engel, T.: BotTrack: tracking botnets using NetFlow and PageRank. In: Domingo-Pascual, J., Manzoni, P., Palazzo, S., Pont, A., Scoglio, C. (eds.) NETWORKING 2011, Part I. LNCS, vol. 6640, pp. 1–14. Springer, Heidelberg (2011)
9. Bilge, L., Balzarotti, D., Robertson, W., Kirda, E., Kruegel, C.: DISCLOSURE: Detecting Botnet Command and Control Servers Through Large-Scale NetFlow Analysis. In: Proceedings of the Annual Computer Security Applications Conference, ACSAC 2012, Orlando, FL USA (December 2012)
10. Bundesamt für Sicherheit in der Informationstechnik: IT Infrastructure Library (ITIL) und Informationssicherheit (2005), https://www.bsi.bund.de/ContentBSI/Publikationen/Studien/ITinf/index_htm.html
11. International Organization for Standardization: Information technology - Security techniques - Information security management systems - Overview and vocabulary (ISO/IEC 27000:2012), 2012 edn. (January 14, 2013)
12. Anstee, D., Bussiere, D., Sockrider, G., Morales, C.: Worldwide Infrastructure Security Report. Technical Report VII, Arbor Networks Inc. (January 2012), http://www.arbornetworks.com/research/infrastructure-security-report
13. Boschi, E., Mark, L., Quittek, J., Stiemerling, M., Aitken, P.: IP Flow Information Export (IPFIX) Implementation Guidelines. RFC 5153 (Informational) (April 2008), http://www.ietf.org/rfc/rfc5153.txt
14. Phaal, P., Lavine, M.: sFlow Version 5 (July 2004), http://www.sflow.org/sflow_version_5.txt
15. ENISA - European Network and Information Security Agency: Cert cooperation and its further facilitation by relevant stakeholders. Technical report, ENISA (December 2006), http://www.enisa.europa.eu/activities/cert/background/coop/files/cert-cooperation-and-its-further-facilitation-by-relevant-stakeholders/at_download/fullReport
16. Molina, M., Paredes-Oliva, I., Routly, W., Barlet-Ros, P.: Operational experiences with anomaly detection in backbone networks. Computers & Security 31(3), 273–285 (2012)