

CarmentiS

A German Early Warning Information System - Challenges and Approaches -

Klaus-Peter Kossakowski¹, Jürgen Sander¹, Bernd Grobauer² and Jens Ingo Mehlau²

¹ PRESECURE Consulting GmbH, Beelertstiege 2, D-48143 Münster
kpk | js@pre-secure.de

² Siemens AG, CT IC CERT, Otto-Hahn-Ring 6, D-81730 München
bernd grobauer | jens.mehlau@siemens.com

Abstract: CarmentiS, a joint effort of the early warning working group within the German CERT association, provides an infrastructure and organizational framework for sharing, correlating and cooperatively analyzing sensor data. This article gives an overview of the CarmentiS infrastructure and organizational framework, and describes the current status of the project. It also addresses open questions that can only be solved by experimenting with co-operative analysis and gives an outlook of possible further developments of the CarmentiS approach towards improved situation awareness and early warning.

1 Scope

The early warning working group within the German CERT association [1] has started to implement an early warning information system (EWIS) called CarmentiS. Like in any known EWIS, one building block of CarmentiS is a network of decentralized sensors, which constitutes the basis of the system. Most of the technical challenges involved in setting up this basis are presented at the FIRST 2005 conference in Singapore [2].

This paper focuses on the second building block of CarmentiS – co-operative human analysis and the aggregation and correlation of different kinds of sensor data that are achieved by CarmentiS. Technological challenges for co-operative human analysis lie in the need to support the analysts such that they can concentrate on the essentials and efficiently pool their know-how, resources and non-sensor-based information sources – also across the boundaries of analyst teams. Also the correlation of different kinds of sensor data poses technological challenges. (It should be pointed out, that existing approaches that pool sensor data from various organizations only operate with one kind of data: DShield [3] and MyNetWatchman [4] operate with firewall logs, which in most cases means that connection attempts to blocked ports are being logged; eCSIRT.net [5] collects and correlates IDS alerts; the IMS project [6] analyzes darknet traffic.) A significant non-technical challenge lies in the legal and organisational as well as human issues in building and using an EWIS based on data sharing and co-operation.

Section 2 of this paper provides a closer overview of the CarmentiS approach to situation awareness and early warning. To promote information sharing, a new method for the sanitization of sensor data was defined, which is described in Section 3. Section 4 informs about the current development status. Section 5 provides an outlook on the next steps in the CarmentiS project.

CarmentiS is supported by the German Federal Office of Information Security (BSI [7]).

2 CarmentiS Approach

Pursuing a cooperative approach for building an early warning system, one has to bring together different teams of different organizations. Challenges to do so lie both in supplying an appropriate technical infrastructure that supports cooperation as well as defining an adequate organizational framework. The following section describes the main participants of CarmentiS that have been identified so far, lists common requirements that have to be fulfilled by CarmentiS, and the CarmentiS architecture to bring these participants together.

2.1 Participants

In a first step, three types of stakeholders are identified and supported by CarmentiS:

- **Partners:** Partners represent organizations, which deliver data of interest towards the CarmentiS central. Rules and regulations regarding the use of the data and analysis results have to be established between the partners and the host of CarmentiS. Each partner has to accept these rules for the data delivered to the CarmentiS central. In other words, it is each partner's responsibility to assure that the delivered data may indeed be exported to CarmentiS.
- **CERTs:** Analysis results and early warning information created by co-operative analysis within CarmentiS are not only of interest to the CarmentiS Partners' CERTs, but also to other CERTs: in most cases, an organization's CERT is the ideal contact for delivering information and warnings relevant for that organization's IT security. Therefore, CarmentiS envisions CERTs that for some reason cannot act as CarmentiS partners as ideal recipients for information and warnings concerning their constituency.
- **Governance / CIIP:** Critical Information Infrastructure Protection (CIIP) is a main task for national governance systems. Protecting critical infrastructures, such as communications, transportation, and energy, against disruption of any kind is increasingly crucial in maintaining both domestic stability and national security.

2.2 Architecture

The cooperative approach of CarmentiS is based on the following simple idea: organizations have situation awareness of their own networks, but knowledge of what is going on behind their perimeters is often missing. In order to broaden the range of vision, participants deliver different types of data of interest to an independent third party. This intermediary, named CarmentiS central, provides the main functionalities for receiving data from partners, conducting analyzes of this data, and presenting appropriate user functions for analysts as well such as CERTs and CIIP-related users. It consists of four main components: the *Import Interface and Storage component*, the *Main Analyze Component*, the *Analysts Workbench*, and the *User Workbench* (see Figure 1). The following sections describe these parts of the CarmentiS architecture including the dynamic behavior of the data export process.

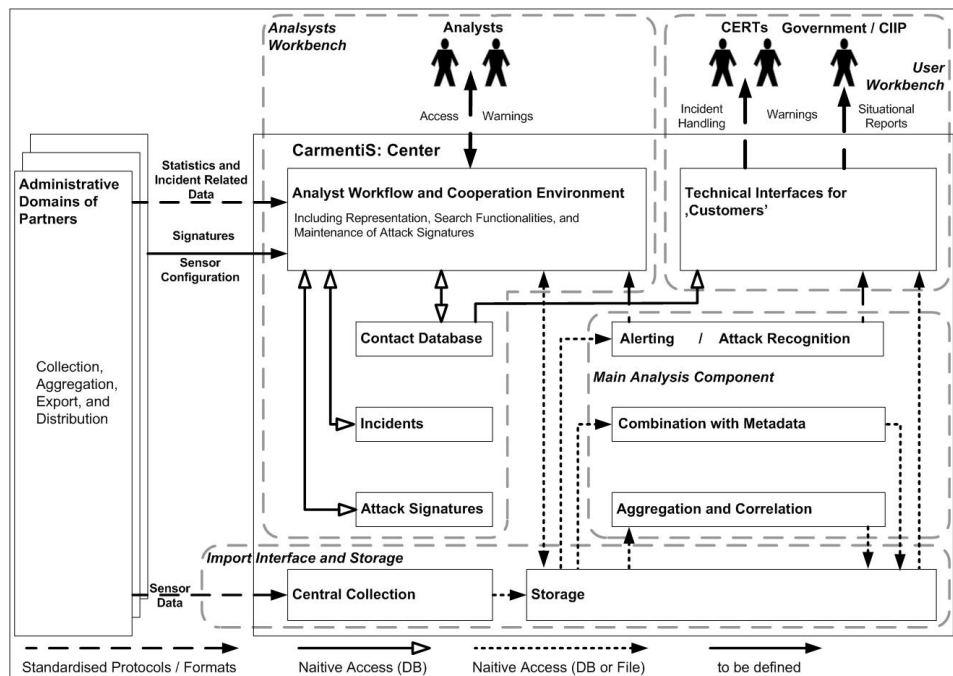


Figure 1: Architecture

Import Interface and Storage

The three main issues regarding data import are various data sources, data volume and privacy concerns. Depending on characteristics and placement of the deployed sensors, very large data sets may be generated. Much of this data is likely to be of a sensitive nature, either because of data-protection laws or because information could be gathered from the data that is considered confidential by the institution whose traffic is being monitored. It is to be expected that CarmentiS partners differ in their assessment of what should be exported and what should not.

Therefore, Information about the export policy that was applied to the data is one important aspect of meta information that must accompany all submitted data; other examples of meta information are sensor configuration, sensor location, etc. Without such meta information, sensible interpretation and correlation of data is impossible. The Carmentis data exchange format therefore must support the communication of meta data.

Because of the sensitive nature of the transmitted data, the transmission channels to the Carmentis central have to be secured using state-of-the-art authentication and encryption mechanisms. Within Carmentis central, the component *Central Collection* receives the data sent by the partner and removes the envelopes of the encapsulated data-files. After that, the extracted data are stored in the main *Storage* for further analysis. Because of the nature of the delivered data, mechanisms have to be found that can deal with very large data sets.

Data Analysis

The main task of the *Data Analysis* component is to aggregate and correlate the data delivered from the partners, to conduct analyzes and to give alerts.

- **Correlation of different data** - Because various kinds of data from different organizations are collected, the data has to be aggregated in an appropriate way. Correlation measures are needed for data resulting from NIDS and Netflows; the integration of additional kinds of data will require additional correlation mechanisms.
- **Profile-based analysis** - Carmentis has chosen profile-based analysis as an appropriate mechanism for cooperative analysis. Profiles are developed by analysts and dynamically updated by the system. The analyzes may be based on the overall data of Carmentis, data of a single partner, or on the aggregated data of specific groups of interest. The latter approach could provide a possibility to examine specific sectors of critical infrastructures by grouping partners into their respective sectors. By conducting the same analysis on different input data, one can gather additional information by comparing the findings.
- **Automated analysis** - Complementing profile-based analysis carried out by analysts, proven automated analysis methods will be necessary to support the analysts, e.g., by creating notifications about events of interest that warrant closer analysis.
- **Alarm notification** - Automatically generated warnings should be distributed using a push model to ensure timely response. In order to further improve response times, the such messages must be based on communication standards such as IODEF [8] thus facilitating an import to standard incident response tools like SIRIOS [9].

Analysts Workbench

The left part of the CarmentiS central in Figure 1 depicts the components necessary for the analysts workbench. In order to provide a cooperative analyze, it is necessary to build virtual teams of analysts, which are employees of the participating partners. The cooperation of the analysts is coordinated and supported by the analysts workbench as follows:

- Presentation of information describing the actual overall security status, analyzes, and technical as well as non-technical indicators for possible malicious activities.
- Presentation of the findings of the analyzes conducted by other analysts regarding the danger of attacks.
- Presentation of reports describing well-known as well as upcoming attack techniques. This provides valuable background information for finding new attacks, accurately adjusting the CarmentiS sensors, and designing appropriate countermeasures.
- Providing an interface, which enables the analyst to develop methods or countermeasures for identifying and combating new attacks.
- Providing capabilities for distributing warnings and advisories via email or SMS.
- Providing an interface for adjusting CarmentiS sensors. This includes capabilities for directly update sensors placed at the partners as well as providing new signatures for download.

In detail, the analysts workbench administrate the signatures and require access to a contact database, an incident database, and an knowledge database.

User Workbench

The findings of the analyzes are presented to CarmentiS users via a web-portal. The user interface should support different views specialized for each participating group of stakeholders. Warnings are sent using a push model (e.g., email) but may very well be duplicated within the the user workbench to provide a comprehensive overview.

3 Information Sharing

At present a proposal on the “retention of data processed in connection with the provision of public electronic communication services” [10] of the Commission of the European Union is discussed controversially in Europe. This proposal demands the implementation of rules which guarantee the retention of at least traffic data for anti-terrorism investigations.

Depending on the characteristic of the used sensors (e.g. NIDS) within an EWIS, traffic data could easily be elicited. Whether such regulation will finally become effective is still very much open: it seems that in its present form, the Commission's proposal is not compatible with national laws. For example, in Germany strong regulations with respect to the protection of privacy exist: each individual has the right to decide about what happens with his or her personal data. This includes the secrecy of communications, which forbids unauthorized eavesdropping as well as storing connection information except accounting purposes, i.e. not only the payload of a data packet is protected, also the header information, like dynamic or static IP addresses.

Obviously, CarmentiS must adhere to rules and regulations set on the national and European level. Privacy concerns could be met by restricting sensor data to purely statistical data. This, however, clearly is not enough for an EWIS.

In order to fulfill these divergent requirements, we divide the data captured by sensors into two groups: connections, which conform to a given security policy specified by the partner that is contributing the sensor data in question and those, which are rated as an attack. In the context of CarmentiS we utilize only the second category of data. In conjunction with proper sanitation of the delivered data, the data can be processed and stored in compliance with national laws. Thus the exchange of aggregated data between participants of CarmentiS will not represent a legal problem, and organizations delivering data can make sure that their interests are not negatively affected.

Sanitization of sensor data must be performed such that as much information as possible is retained: simply replacing IP-addresses and computer names with randomized values, as in the case of simple methods for anonymization, makes correlation of different events impossible. A better approach for sanitization are pseudonymization techniques, which are for instance implemented in CryptoPan [11]. This technique is cryptography based and preserves the prefix relationship among IP addresses. The property is that two pseudonymized IP addresses match on a prefix of n bits if the unpseudonymized addresses match also on this n bits and only then. This approach works fine within a single administrative domain, but in the context of an EWIS where information is processed centrally, this leads to similar problems as in the case of simple anonymization: different IP-addresses from the source set are mapped to equal IP addresses in the target set.

In order to avoid this undesirable effect, we choose a two-stage procedure for the sanitization of IP addresses. According to the given netmask each address will be split into the network and the host part. In doing so each part will be treated differently and finally joined again. During the registration process, a unique CarmentiS Network-Id is assigned to each observed network of the future partner. This Network-Id is produced by the registration authority of the CarmentiS central or any other accepted Trusted Third Party under application of CryptoPan with a secret key (an example is shown in Figure 2).

Step 1: Registration process	
Network-ID of a CarmentiS partner 192.168.64.0/18	assigned CarmentiS Network-ID 214.7.192.0/18
(CryptoPAN) → with privat key of the registration authority	
Step 2: Data export	
IP-Address of a CarmentiS partner 192.168.73.114	calculation
(CryptoPAN) → with privat key of the partner	80.197.133.73 anonymized IP-Address
	and not
	255.255.192.0 Network mask
	=
	0.0.5.73 host part of IP-Addr
	xor
	214.7.192.0 CarmentiS Network-ID
	=
	214.7.197.73 Result

Figure 2: CarmentiS Pseudonymisation

During the export process the second part of each IP-address is processed in the responsibility of the data supplier. This can be done again under application of CryptoPan. But also other methods can be used as long as it is made sure that a clear mapping is realized. The advantage of this approach is, that any pseudonymized IP address has a unique mapping within the whole IP address space, and a reconstruction of the original IP address can only be done by the administrative domain itself.

4 Development and Status

This Section gives an overview of the development status of CarmentiS components for importing, storing and analyzing sensor data. Of the components described in the previous section, the export component (placed at each partner's site) and the data storage unit have been implemented completely. With the extension of an existing open source tool for the analysis of Netflows to cover also NIDS data and meta data, a powerful tool for data analysis has been integrated into the existing infrastructure. The user interface of this analysis tool forms the first component of the analysts' workbench.

4.1 Data Import

Figure 3 shows the data export work flow, as implemented by most of the CarmentiS partners. An export tool was developed within the CarmentiS framework with the following features:

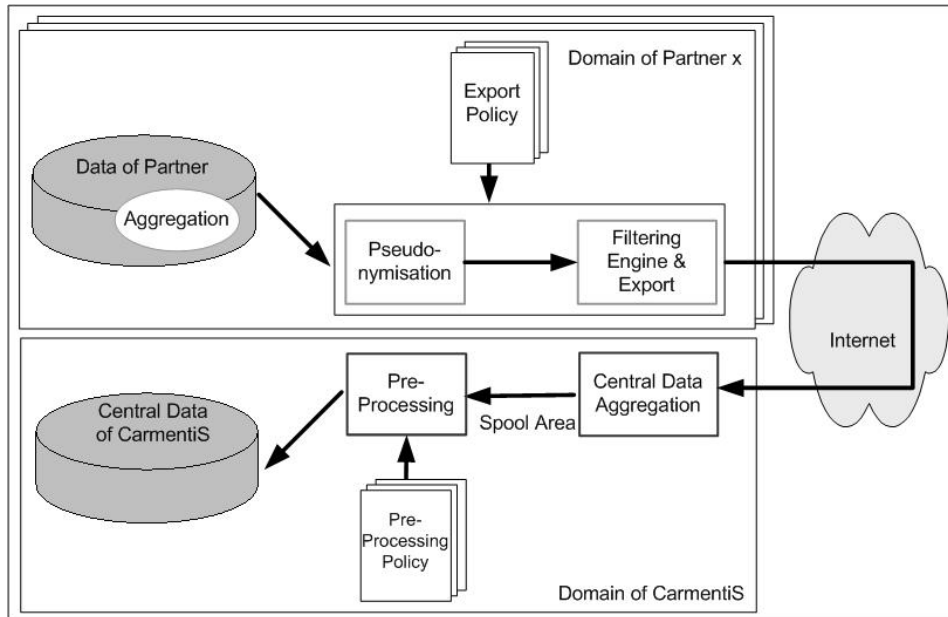


Figure 3: Data Export

- **Support of various data formats** – The export tools supports up to now three main input formats, namely CISCO Netflows [12], Argus [13] and IDMEF [14]. The design of the tool is modular, therefore additional formats can be simply integrated. All data will be delivered in a common format specified within the framework. The common format is not only a container for existing data formats, but provides meta information needed for proper analysis, e.g. class of information, kind of sensor, time frame of the data, location of the sensor, etc.
- **Filtering-Engine** - This module can be used to filter the captured data according to a given policy. In order to fulfill the requirements of an administrative domain, it is possible to drop any connection or event based on network, IP address, port or protocol.
- **Pseudonymization** (see Section 3)

Several transport mechanisms for sending data from a partner to the CarmentiS central can be used. An obvious is the Prelude framework [15] which offers strong authentication with X.509 certificates, integrity and confidentiality via SSL encryption, and avoids data loss in case of network problems by output-buffering and auto-reconnect. Originally, however, Prelude only treats IDMEF data; plugins for the other data kinds used by CarmentiS have been developed.

4.2 Data storage and analysis

Due to the huge amount of data to be processed continuously and the requirements regarding processing time, the exclusive use of a relational database had to be ruled out from the beginning. Instead, a file-based approach that has proved feasible in productive use by SWITCH-CERT has been chosen. Their Netflow toolset worked in practice for the monitoring of the SWITCH internet gateways [16]. In this approach, which is geared towards Netflow data, sensor data is maintained in files representing five minute slices. After new sensor data are available, these are normalized and stored in an internal data format. In the following step the data is aggregated and supplied to the database. Active analysis methods are triggered and start processing the data. In particular, all active profiles defined by the CarmentiS analysts are executed and the analysis results are stored. This carries a twofold benefit

- Analysts checking the result of standard profiles in regular intervals can access analysis results without time delay, because analysis is triggered automatically for new data. Thus, the analyst only has to wait for the processing of modified or newly created profiles.
- After original sensor data has to be discarded to create space for new data, analysis results for profiles executed while the data was available can still be accessed. Analysis results can be stored for a long time because they are very compact in comparison to original sensor data.

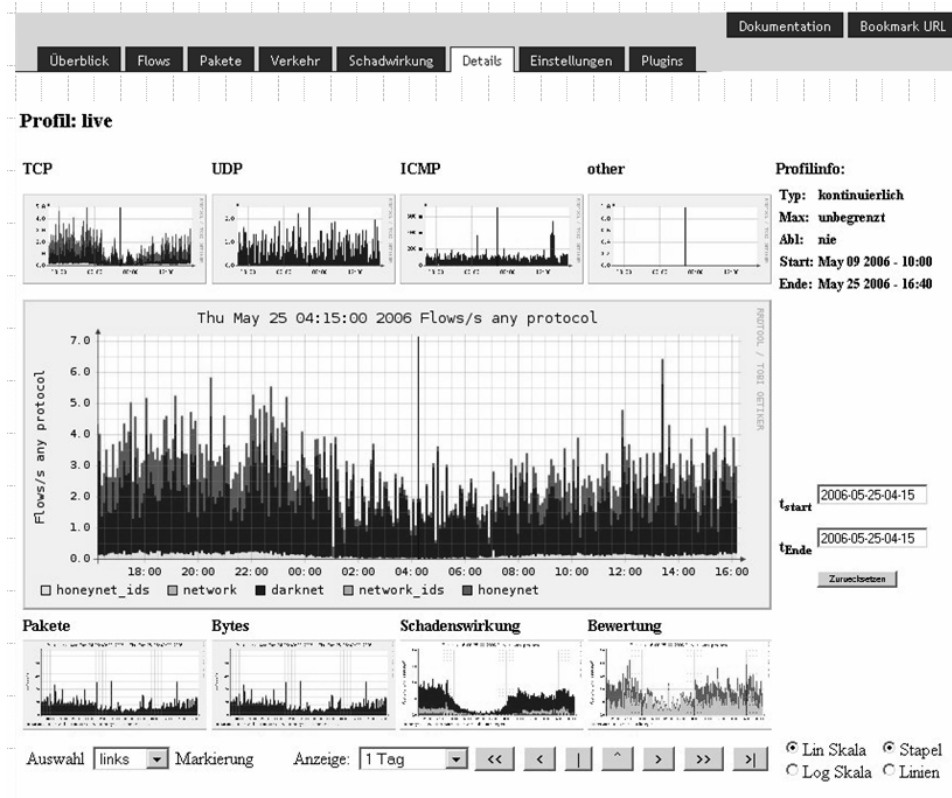


Figure 4: Screenshot from Carmentis Analysis Details Page

All processing tools have only command lines interfaces and are programmed in ,C', in order to be able to fulfill the high requirements on performance. Because the backend part of the netflow toolset (NFDUMP [17]) was developed exclusively for the processing and presentation of Netflow data, extensive modifications had to be made to be able to process IDS data and meta information.

4.3 Analysts Workbench

The analysts workbench was realized as a web front-end, programmed in 'Perl' and 'PHP' has been adopted from NFSEN [18]. The workbench offers the analysts different views and includes a graphical user interface for the tools for data processing. Figure 4 shows the details tab for a detailed analysis of sensor data. The page is divided into two parts: The upper part gives a detailed view and allows to navigate through the sensor data. The content of the main graph can be selected by clicking into any of the smaller graphs which are arranged above and below the main graph (for instance in Figure 4 the selection is: *flows / any protocol*). The time window can be selected by a marking of the area of interest in the main graph with a pointing device. The pages are automatically refreshed every 5 minutes to update the graphs. Figure 5 shows the lower part of the page, which contains all the necessary controls to process the sensor data within the given time window. For instance the top 10 IDS signatures during the given time window.

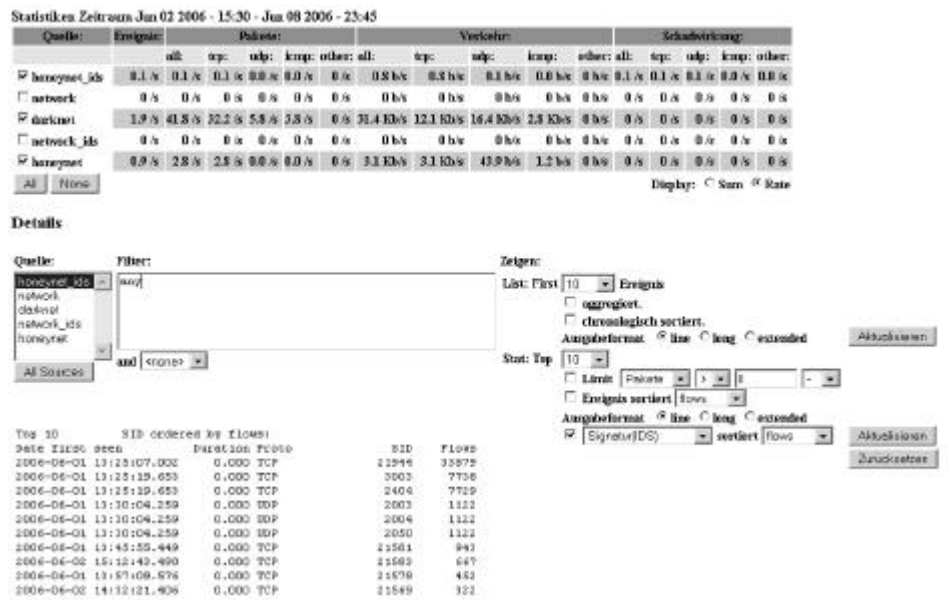


Figure 5: Screenshot from Carmentis Analysis Details Page (2)

5 Outlook

With the basic CarmentiS infrastructure nearing completion, focus is shifting from planning and developing to using the infrastructure. Many aspects of correlating data of different kinds and/or from different sources and co-operatively analyzing data can only be researched with actual trials. Hence, the next step of the project will consist of conducting experiments to research these aspects, which are described in Section 5.1. Further work is also required to go from situation awareness achieved by CERTs taking part in the co-operative analysis towards achieving situation awareness for possible users of CarmentiS as identified in Section 2.1 and, finally, effective early warning; first steps in that direction are described in Section 5.2.

5.1 Experimenting with correlation and co-operative analysis

The following aspects of correlation and co-operative analysis are to be researched by conducting experiments with the basic CarmentiS infrastructure described in the last section.

- **The practicalities of co-operative analysis** - To the best of our knowledge, CarmentiS is the first project to attempt co-operative analysis of sensor data between several CERT teams using a web-accessible analysts' workbench. So far, the part of the analysts' workbench that supports analysis of the collected data using an extended version of the NFDUMP / NFSEN tools has been implemented.

Only experiments in co-operative analysis will be able to tell how such analysis can be organized and which additional features of the analysts' workbench must be added to support such co-operative analysis. This collaborative environment will be built on top of SIRIOS, a modular application framework designed for CERTs with main focus on incident management and vulnerability handling developed under the auspices of CERT-Bund, the German governmental CERT. The core system is based on OTRS, an open source trouble ticket system, licensed under the GPL. To be able to use this system for our purposes, we will adapt available modules and integrate additional databases. In first place stands a database for the maintenance of attack signatures.

- **Leveraging the combination of flow data and IDS data** - As described in Section 4, the Netflow tools used within the CarmentiS analysts' workbench have been extended to handle IDS data, allowing the user to browse through IDS alerts and Netflows via an integrated console. However, ways to leverage the *combination* of IDS data and Netflows must be researched through experimental analysis. It seems likely that IDS data could be used to automatically create profiles for slicing flows according to known attacks detected by IDS systems.

5.2 From co-operative analysis to early warning

CarmentiS partners taking part in the co-operative analysis will be the first to benefit from CarmentiS: at the very least, they will have an additional tool useful for incident handling: events observed within the own network can be compared with events seen on other networks. The next step must be to support situation awareness and early warning also for possible users of the CarmentiS system not involved in cooperative analysis.

- **Creating user-specific situation awareness** - In order to support the CarmentiS stakeholders with information about the current IT-security situation, a web-portal will be implemented. At the moment, we envision to offer different views for each group of stakeholders, containing statistical information, reports and warnings prepared by the analysts, information provided by partners and the CERT community as well as publicly available information sources. How different users can be served best, however, is still subject to research.
- **Improving early warning with automated analysis techniques** - As described in Section 4.2, automated analysis techniques are to be used to notify analysts of events of interest that warrant further examination. Keeping in line with the modular approach that characterizes the CarmentiS approach so far, a well-defined interface for integrating automated analysis techniques into the CarmentiS framework will be implemented.

Preliminary results have shown, that simple threshold schemes are quiet efficient compared to scientific more elaborate methods i.e. neural networks or statistical analysis [18], but that the later can identify more complex behaviours simple methods cannot. Therefore we plan to start with the implementation of two algorithms:

- 1) A threshold scheme based on volume classes high, medium, low. Instead of concentrating on the raw data records we will focus on the number of raw data record sets that conform to pattern created by association rule mining [19] in all records. To achieve usability, the past will be used to determine the appropriate volume class for such pattern. The creation of new pattern will, by itself, trigger manual analysis to assess the relevance of that pattern.
- 2) The second approach will probably concentrate on hidden markov models [20] based on further evaluation of available scientific results.

6 Conclusion

CarmentiS provides a co-operative approach towards situation awareness and early warning in the Internet. At its core is an infrastructure and organizational framework for sharing, correlating and cooperatively analyzing sensor data. Development and deployment of the basic infrastructure components have progressed such that experiments in collecting sensor data from several institutions, correlating and co-operatively analyzing this data can commence. Experiences collected with these experiments are crucial for the definition of a viable organizational framework and the further development of the existing infrastructure.

In order to recognize future network based attacks in time, nationwide IT security management requires an early warning information system with the broadest possible basis. We are strongly convinced, that a regulatory approach cannot achieve a broad enough basis: a suitable cooperation of industry, research organizations as well as the government is indispensable.

7 References

- [1] German CERT Association (Deutscher Cert-Verbund): Website, (<http://www.cert-verbund.de/>)
- [2] Klaus-Peter Kossakowski, EWIS in a BOX -or- How to build a national early warning information system in 80 days! - 17th Annual FIRST Conference
- [3] DSield: DSield Website, (<http://www.dshield.org/>)
- [4] MyNetWatchman: MyNetWatchman Website., (<http://mynetwatchman.com/>)
- [5] eCSIRT.net; eCSIRT.net Website, (<http://www.ecsirt.net/>)
- [6] IMS: Internet Motion Sensor (IMS) - Website, (<http://ims.secs.umich.edu/>)
- [7] BSI: German Federal Office of Information Security (Website), (<http://www.bsi.bund.de>)
- [8] Danyliw, Roman and Meijer, Jan and Demchenko, Juri, The Incident Object Description Exchange Format Data Model and XML Implementation, IETF Incident Handling WG, 2006, <http://www.cert.org/ietf/inch/docs/draft-ietf-inch-iodef-06.txt>
- [9] Sirios, System for Incident Response in Operational Security - Website, (<http://www.sirios.org>)
- [10] Commission of the European Community, The Retention of Data Processed in Connection with the Provision of Public Electronic Communication Services –

September 2005

- [11] Jinliang Fan, Jun Xu, Mostafa Ammar and Sue Moon - Prefix-Preserving IP Address Anonymization , Computer Networks, Volume 46, Issue 2 , 7 October 2004
- [12] Netflows: NetFlow Services Solution Guide,
<http://www.cisco.com/univercd/cc/td/doc/cisintwk/intsolns/netflsol/nfwhite.htm>
- [13] Argus: Audit Record Generation and Utilization System – Website,
(<http://qosient.com/argus/>)
- [14] Debar, H. and Curry, D. and Feinstein, B., The Intrusion Detection Message Exchange Format, IETF Intrusion Detection Exchange Format Working Group, 2006, <http://www.ietf.org/internet-drafts/draft-ietf-idwg-idmef-xml-16.txt>
- [15] Prelude: Prelude IDS Website, (<https://www.prelude-ids.org/>)
- [16] Haag, Peter, NfSen and NFDUMP, 16. TF-CSIRT Meeting, 2005,
<http://www.terena.nl/activities/tf-csirt/meeting16/nfsen-haag.pdf>
- [17] NFDUMP: nfdump tools Website, (<http://nfdump.sourceforge.net/>)
- [18] NfSen: Netflow Sensor Website, (<http://nfsen.sourceforge.net>)
- [19] Chyssler, T. and Nadjm-Tehrani, S. and Burschka, S. and Burbeck, K., Reduction and Correlation in Defence of IP Networks, Proceedings of the 13th IEEE International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), 2004
- [20] Agrawal, R. and Imielinski, T. and Swami, A., Mining Association Rules Between Sets of Items in Large Databases, Proceedings of the ACM SIGMOD Conference on Management of Data, 1993
- [21] Jha, S. and Tan, K. and Maxion, R.A., Markov Chains, Classifiers, and Intrusion Detection, 14th IEEE Computer Security Foundations Workshop (CSFW), 2001