

Selbstorganisierende Netzwerksensoren und automatisierte Ereigniskorrelation

Falko Dressler¹, Reinhard German¹, Peter Holleczeck²

¹Rechnernetze und Kommunikationssysteme, Institut für Informatik,
Universität Erlangen-Nürnberg

{dressler,german}@informatik.uni-erlangen.de

²Regionales Rechenzentrum der Universität Erlangen-Nürnberg (RRZE)
holleczeck@rrze.uni-erlangen.de

Kurzfassung – In dieser Arbeit werden Forschungsaktivitäten des Lehrstuhls für Rechnernetze und Kommunikationssysteme sowie des Regionalen Rechenzentrums der Universität Erlangen-Nürnberg im Kontext eines großflächigen Internet-Frühwarnsystems beleuchtet. Schwerpunkte der Darstellung und Diskussion sind die praktische Anwendbarkeit bisheriger Erkenntnisse sowie die weiteren Forschungsaktivitäten zu verschiedenen Aspekten einer massiv-verteilten Analyse- und Erkennungsinfrastruktur. Zentrale Eigenschaften im Fokus der Untersuchung sind das effiziente Monitoring und Messen, die Selbstorganisation und Adaption von Netzwerksensoren an veränderliche Umgebungszustände, sowie die automatisierte Ereigniskorrelation und Anomalieerkennung.

1 Einleitung

Der Aufbau eines verteilten IT- oder Internet-Frühwarnsystems ist ein wesentlicher Aspekt bei der Bekämpfung von Angriffen aus dem Internet. Wir sehen die Entwicklung eines solchen Systems als Herausforderung in Bezug auf die Integration und Weiterentwicklung verschiedenster Teilsysteme. Die Architektur eines IT-Frühwarnsystems ist vermutlich durch Dynamik und Heterogenität geprägt. Verteilte Sensoren müssen entwickelt, ausgebracht und betrieben werden, wobei sich unterschiedlichste Aufzeichnungssysteme zur Fusion und Aggregation von Sensordaten ausprägen werden. Eine Herausforderung liegt in der automatisierten Konfiguration und der selbstorganisierenden Adaption der Betriebsparameter an aktuelle Umgebungseigenschaften, sowie der Integration von Ereignissen aus anderen Systemen (z.B. Performanzmessungen zur Anomalieerkennung oder Helpdesk-Daten zur Reduktion von False-Positive und False-Negative Meldungen). Für die problemlose In-

tegration von Sensoren und Analysestationen ist der standardisierte Datenaustausch relevant.

In Abbildung 1 sind einige der zentralen Komponenten und deren Interaktion eines IT-Frühwarnsystems skizziert. Auf unterster Ebene werden netzwerknahe Sensoren wie Netzwerkmonitore oder Dienstgütemessungen Daten über das aktuelle Netzwerkverhalten bzw. Inhalte der Verkehrsströme liefern. Diese werden in einer zweiten Sensorebene, welche aus Intrusion Detection Systemen (IDS) besteht, vorverarbeitet. Der Daten- und Konfigurationsaustausch zwischen den Ebenen könnte z.B. durch die standardisierten Formate IPFIX (IP Flow Information Export) und Netconf (Network Configuration) vorgenommen werden. Wesentliche Voraussetzung für den Betrieb in einer verteilten Umgebung ist die automatische Adaption und Konfigurationsänderung der Netzwerksensoren. Selbstorganisationsmechanismen erlauben eine kontextfreie (kein unnötiger globaler Status) und dennoch effiziente Arbeitsweise. Verbunden sind die Sensoren mit einer (semi-)automatischen Ereigniskorrelation, welche ankommende Informationen unterstützt durch dezentrale Ereignisfilter (z.B. Nutzermeldungen, Virenfilter) analysiert. Essentiell ist eine inhärente Integration von adäquaten Gegenmaßnahmen. Beispielhaft sind Firewallsysteme oder Wurm- und Virenfilter gezeigt, die direkt von der Ereigniskorrelation angesteuert werden. Für den Datenaustausch zwischen diesen Komponenten bietet sich das ebenfalls standardisierte IDMEF (Intrusion Detection Message Exchange Format) an.

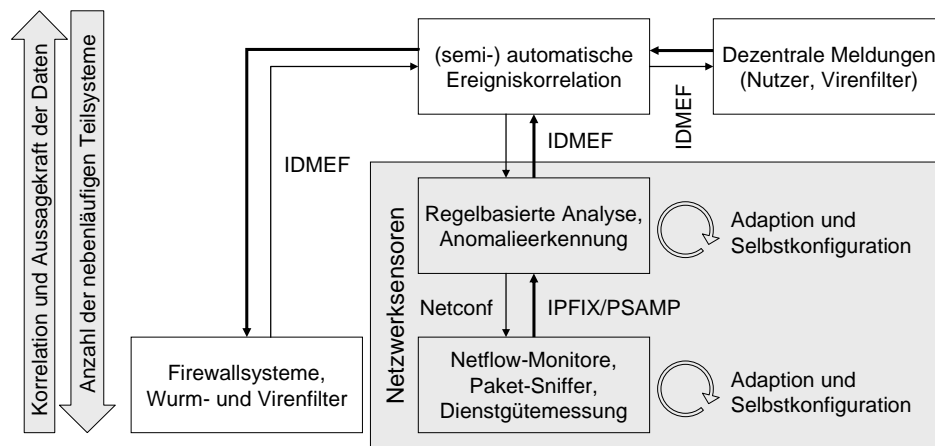


Abbildung 1. Architektur eines IT-Frühwarnsystems

Im folgenden werden Ausschnitte aus aktuellen Forschungsarbeiten präsentiert, welche sich nahtlos in die Kernthemen Sensorik und Auswertung eines

IT-Frühwarnsystems integrieren. In Abschnitt 2 werden Aktivitäten im Bereich Monitoring und Messen beschrieben gefolgt von Arbeiten zur rechtlich Relevanz im Umfeld der Speicherung und Übermittlung von potentiell personenbezogenen Meßdaten in Abschnitt 3. Die Selbstorganisation von Netzwerksensoren und Methoden zur schnellen und effizienten Adaption der Betriebsparameter wird in Abschnitt 4 diskutiert. Die Korrelation von Ereignissen und die abschließende Reaktion auf Verdachtsmomente ist Inhalt von Abschnitt 5. Zusammengefaßt werden die Arbeiten in Abschnitt 6.

2 Monitoring und Messen

Das HISTORY Projekt (High-speed Network Monitoring and Analysis) beschäftigt sich mit der Entwicklung einer Architektur, Methoden und Werkzeugen für die verteilte Analyse von Netzwerkverkehr [4]. Im Vordergrund steht dabei die Arbeit an Netzwerksensoren und an der analytischen Auswertung von Monitoraten. Im Rahmen der IETF sind wir an der Standardisierung der Aufzeichnungsmethodik und Datenübertragung zwischen Sensoren und Analysestationen beteiligt. In den Arbeitsgruppen IPFIX, PSAMP und NSIS arbeiten wir an Verfahren, welche die Interoperabilität verschiedener Werkzeuge durch standardisierten Datenaustausch garantieren. Inhalte dieser Arbeit sind Netflow-basiertes Monitoring (IPFIX), statistisches Sampling von einzelnen Paketen (PSAMP) [1] und die dynamische Aggregation von Daten mehrerer Sensoren [3]. Das wesentliche Ziel der Forschungsaktivitäten im Rahmen von HISTORY ist es, Methoden zu entwickeln, welche die Verarbeitung von großen Datenmengen bzw. hohen Datenraten in Echtzeit auf einfachen Standard-PCs erlaubt. Visualisierungstechniken und Anonymisierungsmethoden (siehe Abschnitt 3) runden das Bild einer visionären Umgebung für Netzwerkmonitoring und -analyse ab. Die entwickelten Werkzeuge sind als Open-Source verfügbar und wurden bereits in verschiedenen Anwendungen im Bereich Angriffserkennung [2], Accounting [12] und Traceback erfolgreich getestet.

Besonders hervorgehoben werden soll das in diesem Rahmen entwickelte Werkzeug Vermont (Versatile Monitoring Toolkit), welches die Aufzeichnung und Aggregation von Paketen im Gigabit-Bereich mit Standard-PCs erlaubt. Vermont stellt in diesen Bereichen eine Referenzimplementierung der IETF für die Protokolle IPFIX und PSAMP dar. Zusätzlich flossen die aktuellen Arbeiten im Bereich dynamische Datenaggregation [3] bereits ein und konnten erfolgreich getestet werden [21].

Für die Konfiguration von Vermont sind zwei Alternativen vorgesehen. Zum einen wurde bereits eine Netconf-basierte Lösung implementiert [18].

Alternativ soll das Pfad-gebundene Signalisierungsprotokoll Metering-NSLP [10, 11] basierend auf dem PSVP-Nachfolger GIMPS eingesetzt werden.

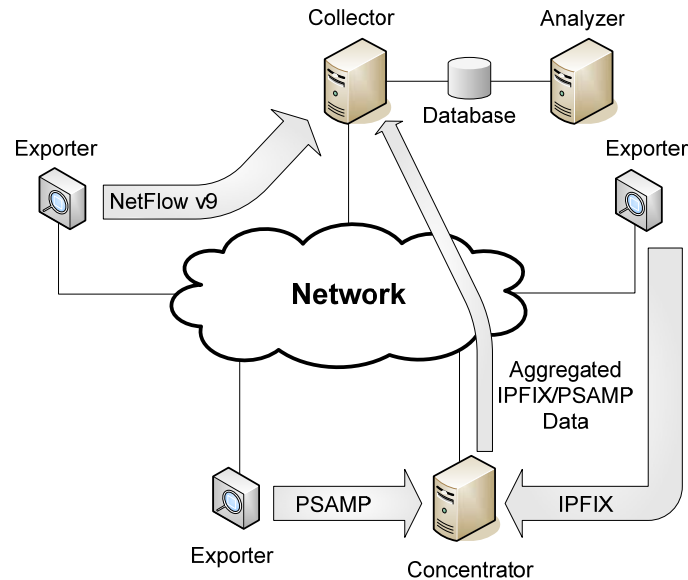


Abbildung 2. Überblick über die HISTORY-Architektur

Die gesamte Architektur ist in Abbildung 2 gezeigt. Mehrere verteilte Monitore sammeln Samples von Paketen bzw. Netflow-Datensätze und exportieren diese an einen übergeordneten Kollektor für eine spätere Analyse. Die hierarchische Strukturierung erlaubt sowohl den Einsatz von sogenannten Konzentratoren zur Datenverdichtung als auch eine effiziente Lastverteilung.

Die Messung der Netzwerkperformanz ist Schwerpunkt des DFN-Labors am RRZE [16]. Die Leistungsmessung erfolgt aktiv durch Meßstationen, die an wichtigen Vermittlungsstationen plaziert sind. Sie erzeugen künstlichen Monitor-Verkehr und bestimmen fortlaufend Latenzen und Jitter auf den dazwischen liegenden Streckenabschnitten. Für genaue Zeitmessungen sind sie mit GPS-Antennen ausgestattet.

Im Moment werden Installationen betrieben, die Meßdaten aus folgenden Bereichen zur Verfügung stellen: Deutschland (WiN, mit ca. 50 Stationen), Europa (GEANT, mit ca. 20 Stationen) und USA (mit zwei Stationen). Die empfindliche Meßtechnik liefert z.B. Interpretationsmöglichkeiten für Probleme bei Videoübertragungen [19]. Neben normalen meist tageszeitlichen Schwankungen treten manchmal Anomalien auf, die auf außergewöhnliche Ursachen schließen lassen. So entpuppte sich der unübersehbare Delay-

Anstieg auf einer Kernnetz-Strecke am 3.5.04 um 4 Uhr als der Beginn der Ausbreitung des Sasser-Wurms [14], wie in Abbildung 3 gezeigt. Auch Flash-Crowd-Effekte an ftp-Servern machten sich auf diese Weise bemerkbar. Richtig eingesetzt bzw. interpretiert, liefern diese Meßstationen sehr frühe Hinweise auf sich anbahnende Effekte.

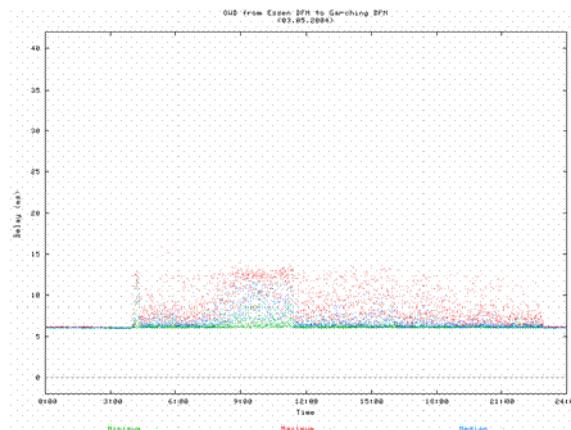


Abbildung 3. Erkennung von Netzwerkanomalien: Sasser-Wurm

3 Anonymisierung von Sensordaten

Aus rechtlichen Gründen zum Schutz der Privatsphäre ist es nicht möglich, Monitordaten unverändert in einem verteilt betriebenen Verbund von Analysestationen zu verarbeiten. In der deutschen Rechtsprechung ist dies zwar insofern eine Grauzone, als daß konkrete Gesetze fehlen und auch sonst die Meinungen stark divergieren. Dennoch geht man von einer hohen Bedeutsamkeit der Anonymität einzelner Internetbenutzer aus, die Vorrang vor allgemeinen Schutz- bzw. Betriebsbedenken hat. In einer gemeinsam mit dem Institut für Rechtswissenschaft der Universität Tübingen angefertigten Studie wurden die Grauzonen der deutschen Rechtsprechung untersucht und ausführlich dargestellt [13]. Ziel der Arbeit war es, die derzeitige Rechtslage zu untersuchen, Möglichkeiten der Anonymisierung zu entwickeln und ein regelbasiertes Anonymisierungsmodul zu implementieren. Zu Beginn wurden die aufzuzeichnenden Daten auf ihre Sensitivität untersucht mit dem Ergebnis, daß allein die IP-Adresse geeignet ist, einen Bezug zum Nutzer herzustellen, sowie daß der Zielport Auskunft über die Art des in Anspruch genommenen Dienstes geben kann. Nach der Feststellung der Zulässigkeit der Verarbeitung wird aus den datenschutzrechtlichen Grundsätzen der Datenvermeidung und -

sparsamkeit deren Modalitäten abgeleitet. Demnach sind die erhobenen Daten frühestmöglich zu pseudonymisieren, zu anonymisieren oder zu löschen, wenn auch dann der Zweck noch erreicht werden kann. An die rechtliche Begutachtung schlossen sich der Entwurf und die Diskussion verschiedener Algorithmen zur Anonymisierung an. Diese umfaßten die Speicherung der IP-Adressen in verkürzter Form oder deren Hashwertes, deren Verschlüsselung oder Abbildung mittels einer injektiven und netzwerkpräfixerhaltenden Funktion. Dabei wurde festgestellt, daß nur mittels der Verkürzung der Adressen ein datenschutzrechtlicher Gewinn erzielt wird, da ansonsten nur eine Pseudonymisierung der Adresse erfolgt, welche selbst bereits Pseudonym ist, oder der erlaubnisbedürftige Umwandlungsvorgang außerhalb des Zwecks der Erlaubnistatbestände läge. Im gleichen Kontext entstand eine Bibliothek zur Policy-gestützten Anonymisierung von Meßdaten [13].

4 Selbstorganisation von Netzwerksensoren

In einer verteilten Umgebung von Netzwerksensoren ist es eine wesentliche Anforderung, den Betrieb und die Wartung der Sensoren zu automatisieren. Zentrale, womöglich händische Eingriffe sind vermutlich nur bedingt möglich bzw. erfordern eine teure zusätzliche Netzwerkinfrastruktur, um die Zuverlässigkeit der Steuerung und Konfiguration zu gewährleisten. Die weitgehende Selbstorganisation der einzelnen Sensoren scheint die optimale Lösung der geschilderten Fragestellungen zu sein.

Seit längerer Zeit arbeiten wir an Methoden zur selbstständigen automatisierten Adaption von Betriebsparametern der Netzwerksensoren mit dem Ziel, die Meßgenauigkeit des Gesamtsystems zu erhöhen und Überlastsituationen zu vermeiden. Es hat sich gezeigt, daß die adaptive Anpassung von Parametern für einzelne Systeme durch einfache Regelschleifen möglich und sehr effizient ist [5-7]. Abbildung 4 zeigt ein einfaches Modell mit den wichtigen Parametern einer lokalen Regenschleife. Hervorzuheben sind die sogenannten Blacklist und Whitelist Tabellen. Diese repräsentieren Firewall-Konfigurationen bzw. als legitim erkannte Datenströme. Dieses Model erlaubt es nun, die Verhaltensweise von Monitoren, Angriffserkennungssystemen und Firewalls zu untersuchen, während man die Eingangsparameter, d.h. die Netzwerklast sowie die Anzahl und Aggressivität von potentiellen Angreifern verändert. Es hat sich gezeigt, daß die Möglichkeit der Manipulation der Wartezeit in den genannten Listen ausreicht, um die Arbeitsweise der Systeme an beliebige Netzwerklast anzupassen und dennoch jeweils eine maximale Erkennungsrate zu erzielen bei gleichzeitiger Vermeidung von Überlastsituationen. Detaillierte Berechnungsmodelle sind in [6, 7] zu finden.

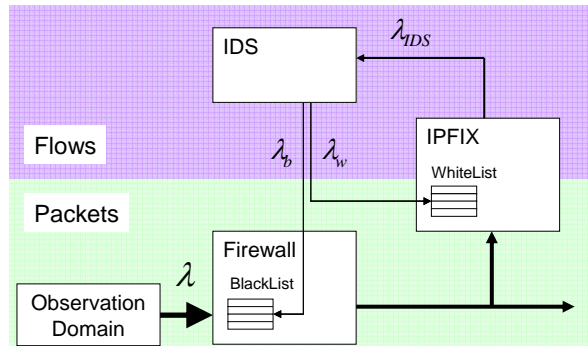


Abbildung 4. Modell für eine dynamische Rekonfiguration von Netzwerksensoren

Im nächsten Schritt sollen Cluster von Einzelsystemen durch spontane Ad-hoc-Methoden gebildet werden, die eine noch genauere Kalibrierung erlauben. Das Ziel ist es, Parameteranpassungen möglichst *vor* dem meßbaren Ereignis vorzunehmen. Dies kann nur durch Korrelation von Sensordaten verschiedener benachbarter Systeme gelingen. Allgemeine Studien zu Selbstorganisationsmechanismen in anderem Kontext wurden z.B. in [8, 9] durchgeführt, wobei vor allem biologisch inspirierte Verfahren hohe Erwartungen wecken und intensiv untersucht werden [17].

5 Korrelation und Reaktion

Die Korrelation von Sensorinformationen wurde in verschiedenen Projekten untersucht. Erwähnenswert ist die verteilte Angriffserkennung mit CATS (Cooperative Attack Detection Systems), welche auf eine breite Installation von Monitoringstationen aufsetzt und einen adaptiven Regelkreis mit den verteilten Netzwerksensoren beschreibt [2].

In einer Studie wurde die Zusammensetzung von Sicherheitsvorfällen untersucht. Das Ergebnis hierbei war, daß der größte Anteil aller erfaßten Sicherheitsvorfälle Malware ausmachte [15]. Durch die automatisierte Verbreitung von Sicherheitsvorfällen kann eine sehr große Anzahl von Systemen in kurzer Zeit infiziert werden. Um dem effizient zu begegnen müssen auch zur Bearbeitung der Sicherheitsvorfälle automatisierte Verfahren eingesetzt werden, um eine ressourcenschonende und zeitnahe Auflösung dieser Sicherheitsvorfälle zu erreichen. Um dies zu unterstützen wurde ein Vorfallsmanagement entwickelt und evaluiert. Die werkzeuggestützte Verarbeitung unterschiedlicher Sensordaten wurde im Rahmen der Entwicklung eines Vorfalls-

managementsystems untersucht [22]. Nach erfolgreicher Erkennung von Angriffen und Vorfällen ist eine adäquate Reaktion erforderlich. Wir sehen in der automatisierten Konfiguration von Firewalls und Access-Listen in Routern eine effiziente Methode, zeitnah auf Probleme zu reagieren [20].

6 Zusammenfassung

Die aufgezeigten Forschungsaktivitäten des Lehrstuhls für Rechnernetze und Kommunikationssysteme sowie des Regionalen Rechenzentrums der Universität Erlangen-Nürnberg stellen einzelne Bausteine dar, die im Kontext eines großflächigen Internet-Frühwarnsystems unmittelbar eingesetzt werden können. Durch vielfältige Tests im praktischen Einsatz sowie im Labor als auch durch Untersuchungen an Simulationsmodellen konnten die einzelnen Aspekte einer massiv-verteilten Analyse- und Erkennungsinfrastruktur analysiert und verifiziert werden. Die zentralen Eigenschaften im Fokus der Untersuchung sind dabei das effiziente Monitoring und Messen, die Selbstorganisation und Adaption von Netzwerksensoren an veränderliche Umgebungszustände, sowie die automatisierte Ereigniskorrelation und Anomalieerkennung.

Vernetzung: Der Lehrstuhl für Rechnernetze und Kommunikationssysteme arbeitet eng mit dem Regionalen Rechenzentrum zusammen. Beide sind durch Kooperationsprojekte mit dem Wilhelm-Schickard-Institut für Informatik der Universität Tübingen und dem Verein zur Förderung eines Deutschen Forschungsnetzes (DFN), sowie die EU Projekte Diadem Firewall und MUPBED (Multi-Partner European Testbeds for Research Networking) in einem nationalen und internationalen Forschungsumfeld vernetzt.

Literatur

1. T. Dietz, F. Dressler, G. Carle, B. Claise, and P. Aitken, "Information Model for Packet Sampling Exports," Internet-Draft, draft-ietf-psamp-info-03.txt (2005)
2. F. Dressler, G. Münz, and G. Carle, "CATS - Cooperating Autonomous Detection Systems," Proceedings of 1st IFIP International Workshop on Autonomic Communication (WAC 2004), Poster Session, Berlin, Germany (2004)
3. F. Dressler, C. Sommer, and G. Münz, "IPFIX Aggregation," Internet-Draft, draft-dressler-ipfix-aggregation-02.txt (2005)
4. F. Dressler and G. Carle, "HISTORY - High Speed Network Monitoring and Analysis," Proceedings of 24th IEEE Conference on Computer Communications (IEEE INFOCOM 2005), Poster Session, Miami, FL, USA (2005)
5. F. Dressler, "Adaptive network monitoring for self-organizing network security mechanisms," Proceedings of IFIP International Conference on Telecommunica-

tion Systems, Modeling and Analysis 2005 (ICTSM2005), Dallas, TX, USA (2005) 67-75

6. F. Dressler and I. Dietrich, "Simulative Analysis of Adaptive Network Monitoring Methodologies for Attack Detection," Proceedings of IEEE EUROCON 2005 - The International Conference on "Computer as a Tool", Belgrade, Serbia and Montenegro (2005) 624-627
7. F. Dressler, "Adaptive Re-Configuration of Network Monitoring Applications," Proceedings of Dagstuhl Seminar 06011: Perspectives Workshop: Autonomic Networking, Schloss Dagstuhl, Wadern, Germany (2006)
8. F. Dressler, "Self-Organization in Ad Hoc Networks: Overview and Classification," University of Erlangen, Dept. of Computer Science 7, Technical Report 02/06 (2006)
9. F. Dressler, "Self-Organization in Autonomous Sensor/Actuator Networks," 9th IEEE/ACM/GI/ITG International Conference on Architecture of Computing Systems - System Aspects in Organic Computing (ARCS'06), Frankfurt, Germany, Tutorial (2006)
10. A. Fessi, G. Carle, F. Dressler, J. Quittek, C. Kappler, and H. Tschofenig, "NSLP for Metering Configuration Signaling," Internet-Draft, draft-dressler-nsis-metering-nslp-03.txt (2005)
11. A. Fessi, C. Kappler, C. Fan, F. Dressler, and A. Klenk, "Framework for Metering NSLP," Internet-Draft, draft-fessi-nsis-m-nslp-framework-02.txt (2005)
12. U. Foell, C. Fan, G. Carle, F. Dressler, and M. Roshandel, "Service-Oriented Accounting and Charging for 3G and B3G Mobile Environments," Proceedings of 9th IFIP/IEEE International Symposium on Integrated Network Management (IFIP/IEEE IM 2005), Poster Session, Nice, France (2005)
13. F. Haibl, "Erstellung einer Funktion zur regelbasierten Anonymisierung von Verbindungsdaten im Internet," Pre-Master's Thesis (Studienarbeit), Wilhelm-Schickard-Institute for Computer Science, University of Tuebingen (2005)
14. P. Holleccek, "Interner Bericht des DFN-Labors," University of Erlangen (2004)
15. J. Kaiser, "IT-Sicherheit im Nebel," Proceedings of GUUG Frühjahrsfachgespräch, Munich, Germany (2005)
16. B. Koenig and S. Kraft, "Das DFN-Labor - Qualitätsversicherung im Wissenschaftsnetz," *DFN-Mitteilungen*, vol. 70 (2006) 13-15
17. B. Krüger and F. Dressler, "Molecular Processes as a Basis for Autonomous Networking," *IPSI Transactions on Advances Research: Issues in Computer Science and Engineering*, vol. 1 (2005) 43-50
18. G. Münz, A. Antony, F. Dressler, and G. Carle, "Using Netconf for Configuring Monitoring Probes," Proceedings of IEEE/IFIP Network Operations & Management Symposium (IEEE/IFIP NOMS 2006), Poster Session, Vancouver, Canada (2006)
19. S. Naegele-Jackson, H. Kerscher, R. Kleineisel, and P. Holleccek, "IPPM Measurements of the German Research Network G-WiN and their Application to Vi-

- deoconferencing Services," Proceedings of 8th IASTED International Conference on Internet & Multimedia Systems & Applications (IMSA 2004) (2004)
20. F. Prester, "Security within Networks with Ease and One-Command-Philosophy," Proceedings of Terena 2006, Catania, Sizilien (2006)
 21. C. Sommer, "Implementation of a Netflow Concentrator," Pre-Master's Thesis (Studienarbeit), Department of Computer Sciences, University of Erlangen-Nuremberg (2005)
 22. A. Vitzthum, "Implementierung eines Vorfallsmanagementsystems für IP-Sicherheit," Pre-Master's Thesis (Studienarbeit), Department of Computer Sciences, University of Erlangen-Nuremberg (2006)