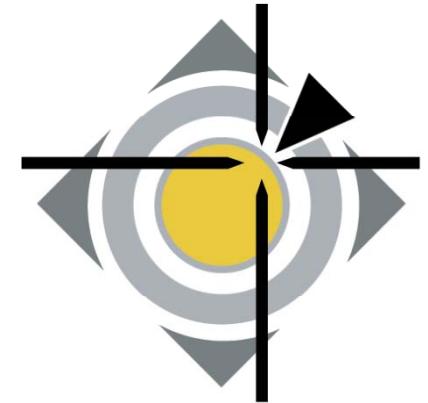


IMF 2008

Workshop-Day

Mannheim, Germany – 25.08.2008



Best Practices (Internet) Auditing

Andreas Rohr

Agenda

▶ **Audit in General**

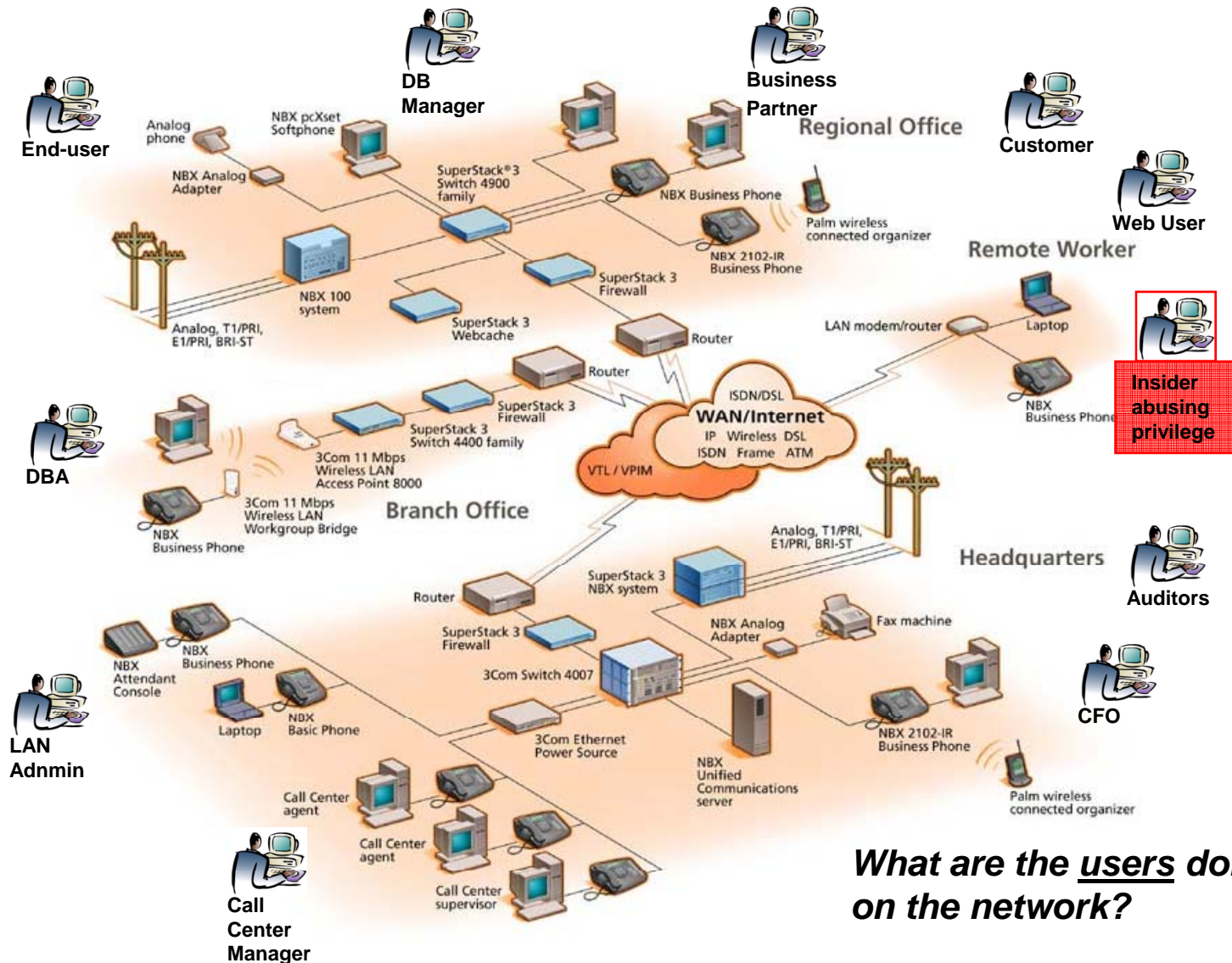
- Why Auditing
- Audit Sources
- Dealing with the amount of audit data

▶ **Internet-Audit (case study)**

- Private internet usage
- Statistics
- Calculate internet usage time consumption



Why Auditing?



What are the users doing on the network?

Why Auditing?

► Motivations to audit can be driven by:

- Requirements, stated by an (international) law: **Compliance** due to legal constraints (BASEL II, SOX, ...)
- **Policies**, that can not be technically enforced
- Interest in business/process related **indicators** (ex.: When do users/customers do what to which extend?)
 - Planning IT-infrastructure resources (operations management)
 - Optimized customer processes (resource allocation)
 - Recognize possible bottlenecks
- **IT-Security** interests

What is Auditing?

- ▶ Auditing generally deals with the question (two different definitions):
 1. **Who** has done **What**, **When**, **Why** and **Where**?
 2. **Who** did **What** type of action **on What**?
When did he do it and **Where**, **From Where** and **Where To**?

Types of Auditing

- ▶ **Incident/Case (driven) audits**
- ▶ **Regularly performed audits**
 - Asynchronous: post mortem to an event
 - Synchronous: (near) real-time after an audit-entry is generated
 - Trigger based audits

Audit Sources

▶ **Logs created by various systems and applications**

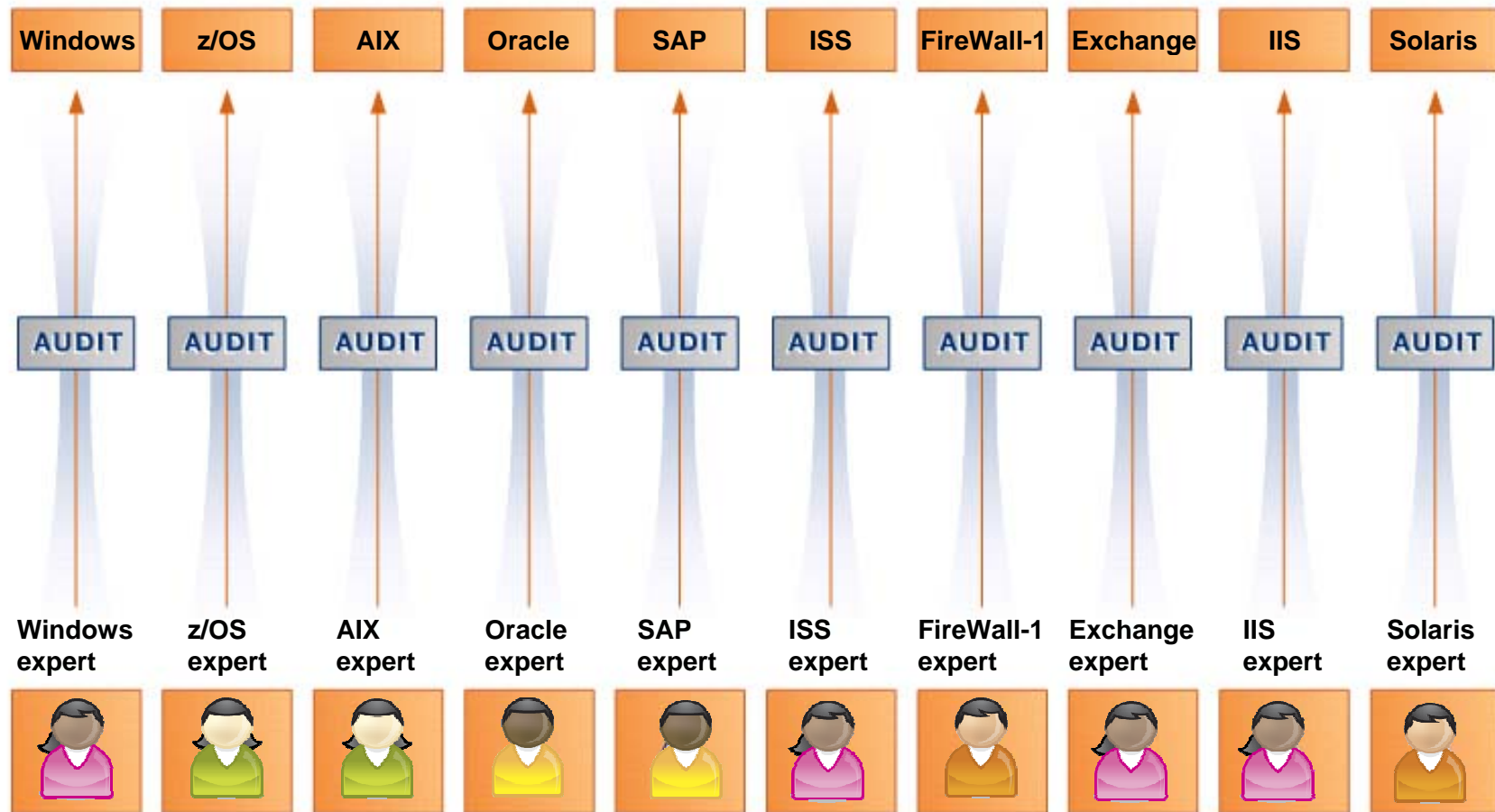
- Door entry systems (ID-cards, biometric authentication)
- Operating System Logging (syslog, system event logging, RACF, etc.)
- Database logging (transaction log, access log, ...)
- Network devices (Router logs)
- Security devices (Firewalls, IDS)
- Application specific logging (access-logs, VMWare Virtual Center Activity Log, custom logs defined by a administrator)



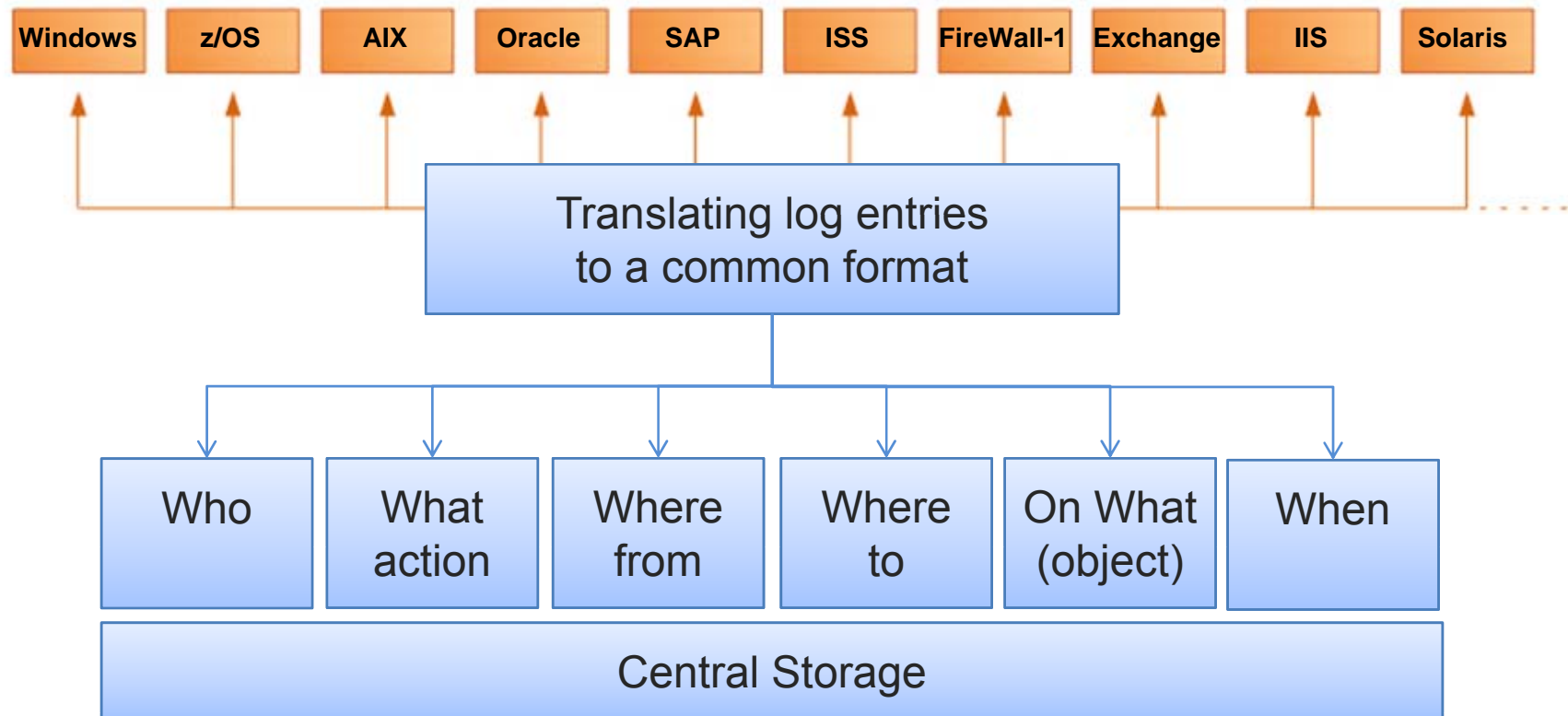
▶ **Cameras**

- On-the-fly object categorization of a video-stream (surveillance camera) → Meta-Data for any object that moves within the camera scope

Audit Sources



Audit Sources – Translation



Audit Sources – Translation

```
129.0.65.12 || landreasrohr || [16/
ermanwings.com/libraries/yui/calenc
97 || "http://www.germanwings.com/
SIE 6.0; Windows NT 5.0; .NET CLR
129.0.65.12 || landreasrohr || [16/
ermanwings.com/libraries/yui/calenc
|| "http://www.germanwings.com/inde
6.0; Windows NT 5.0; .NET CLR 1.1.
129.0.65.12 || landreasrohr || [16/
ermanwings.com/libraries/yui/yahoo-
|| 30580 || "http://www.germanwinc
ible; MSIE 6.0; Windows NT 5.0; .NE
129.0.65.12 || landreasrohr || [16/
ermanwings.com/style/compact3.css I
wings.com/index.de.shtml" || "Mozi
.NET CLR 1.1.4322; .NET CLR 2.0.50
129.0.65.12 || landreasrohr || [16/
ermanwings.de/images/maincontent/cc
|| 461 || "" || "Mozilla/4.0 (compe
For Help, press F1
```

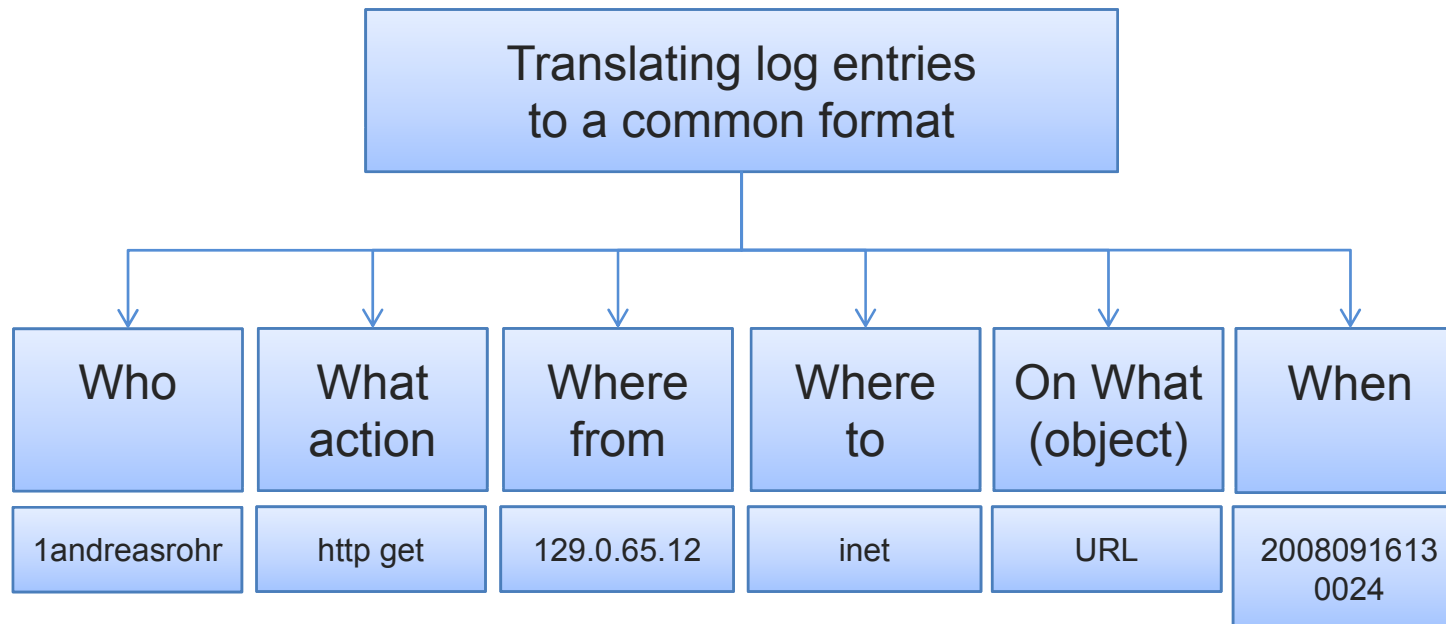
Proxy Farm access log

Windows Event log (binary)

The screenshot displays two overlapping windows from a Windows operating system. The background window is a file explorer titled 'audit.internet' located at 'D:\Org2_rohr\audit\ADS\DC-eventlog'. It shows a directory tree with various files including 'ads.vbs', 'ads_extract_eventlog.php', 'ARo_Bw_Auth.php', 'ARo_Bw_Ldap.php', 'class.ARo_Bw_LDAP.php', 'class.audit.internet.php', 'class.audit.internet.statistics.20080109', 'class.audit.internet.statistics.peruser.php', 'class.audit.internet.statistics.php', 'convert.php', 'dbimport_ads.php', 'extract_internet_usage.php', 'extract_IP.php', 'get_log_files.php', 'get_noexpire_password_option.vbs', 'get_radio_groups_and_user.vbs', 'get_uncategorized.php', 'get_uncategorized_uidbased.php', 'harryelpel.php', 'internet_harry_elpel.php', 'internet_harry_elpel2.php', 'ldap_unknown_users.php', 'logformat_convert.log', 'monthly_internet_statistics.php', 'monthly_internet_statistics_per_deptm', 'pages_for_a_user.php', 'parse_log_files_by_date.php', 'parse_log_files_by_date_new_test.php', 'parse_log_files_by_date_nonanonymou', 'pdftest.php', 'pdftest.php', 'queries.sql', and 'run_statistics.php'. The foreground window is a hex editor titled 'SysEvent.Evt' showing a binary dump of the Windows Event Log. The dump consists of a grid of hexadecimal values and their corresponding ASCII representations. The ASCII column shows fragments of text such as '\.....\...', '...I.A.S...', 'H.H.S.A.N.1.0.0.', 'K.U.V.1.0...C.i.', 's.c.o.W.o.r.k.s.', '..B.M.V.G.\C.i.', 's.c.o.W.o.r.k.s.', '..1.2.9...0...1.', '..1.3.7...%2.', '1.4.7.4.8.3.6.8.', '6...%2.1.4.7.', '4.8.3.6.8.6...1.', '2.9...0...1...1.', '..C.i.s.c.o. .S.', 'B.1...1.2.9...0.', '..1...1.3.7...V.', 'i.e.t.u.a.1...2.', '..W.i.n.d.o.w.s.', '-.A.u.t.h.e.n.t.', 'i.f.i.z.i.e.r.u.', 'n.g. .f.ü.r. .a.', '1.1.e. .B.e.n.u.', 't.z.e.r. .v.e.r.', 'w.e.n.d.e.n...%', '%2.1.4.7.4.8.3.', '6.8.8...%2.1.', '4.7.4.8.3.6.8.5.', '..%2.1.4.7.4.', '8.3.6.8.5...P.A.', 'P...%2.1.4.7.', '4.8.3.6.8.5...3.', '6...%4.1.3.2.', 'u...e...%', and 'LfLevS..\(JG)\(JG'.

Audit Sources – Translation (ex.)

```
aro.txt - WordPad
File Edit View Insert Format Help
129.0.65.12 || 1andreasrohr || [16/Sep/2008:13:00:24 +0200] || "GET http://www.g
ermanwings.com/libraries/yui/calendar/assets/calendar.css HTTP/1.0" || 200 || 57
97 || "http://www.germanwings.com/index.de.shtml" || "Mozilla/4.0 (compatible; M
SIE 6.0; Windows NT 5.0; .NET CLR 1.1.4322; .NET CLR 2.0.50727)" ||
```



Aggregation / Correlation

- ▶ **Hundreds of thousands of audit/log entries**
- ▶ **Different audit sources**
- ▶ **User behavior (detection of normal vs. abnormal behavior measured to a certain baseline)**

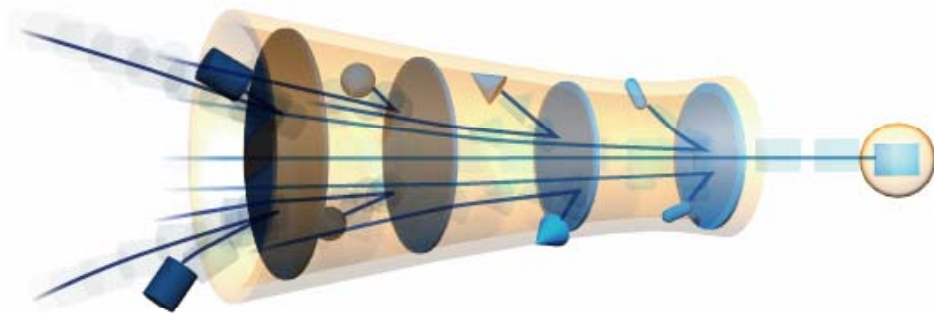
 **Aggregation / Correlation is needed!**

Statistical Analysis

- Detecting non typical behavior

Rules-based Correlation

- Detecting non typical behavior
- Detecting misuse
- Enforcing security policies



Accountability of Logs/Audit Sources

▶ Spoofing of basically any attribute that is used by an auditor:

- ensuring that a log entry is not tricked/spoofed by the trigger (triggering user)
 - ID spoofing (who)
 - IP-Address spoofing (from where)
 - etc...
- Timestamp service for logs / log entries (PGP, company PKI)

▶ Cross-correlation of different sources

- Door entry systems / time registration systems
- DNS log
- OS login (LDAP log)



Auditing Acceptability

Required Skills for an Auditor

- ▶ **Coheasive view on what and how systems / applications are tracking/logging**
- ▶ **At least one programming (scripting) language for automation (ksh, perl, php, vbscript, etc.)**
- ▶ **Forensic knowledge (capabilities) to cross-check results (if they are plausible to the rest of the data)**

Internet-Audit

▶ Case study of the FMoD Germany

▶ Assumptions:

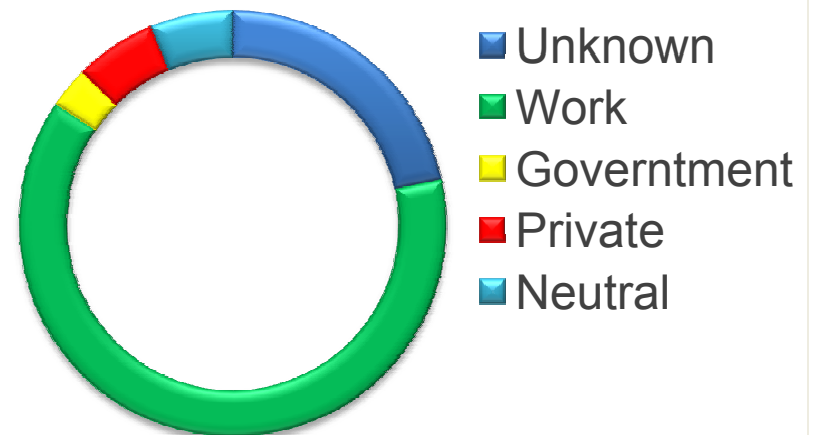
- Internet usage is not permitted for private purposes
- Authentication for internet access (proxy farm authenticates with LDAP directory)
- MOU between MoD secretary of state and the employee committee
 - Audit on misuse (e.g. private usage) on a per user basis is permitted once a month
 - Anonymised /cummulated audit (statistics) on the department level and above; results in a executive summary report at the end of each month



Internet-Audit: Private Usage Def.

► Definition of private usage within the FMoD:

- Classification based on manually rating of sites:
 - Unknown (unclassified yet)
 - Work relevant
 - Special subnet (governmental intranet)
 - Private
 - Not classifiable (neutral: not private, but also not work relevant)
- Using categories of webwasher URL-filter database
- Threshold for mitigation between man power to classify web sites and precision: 20% Unknown

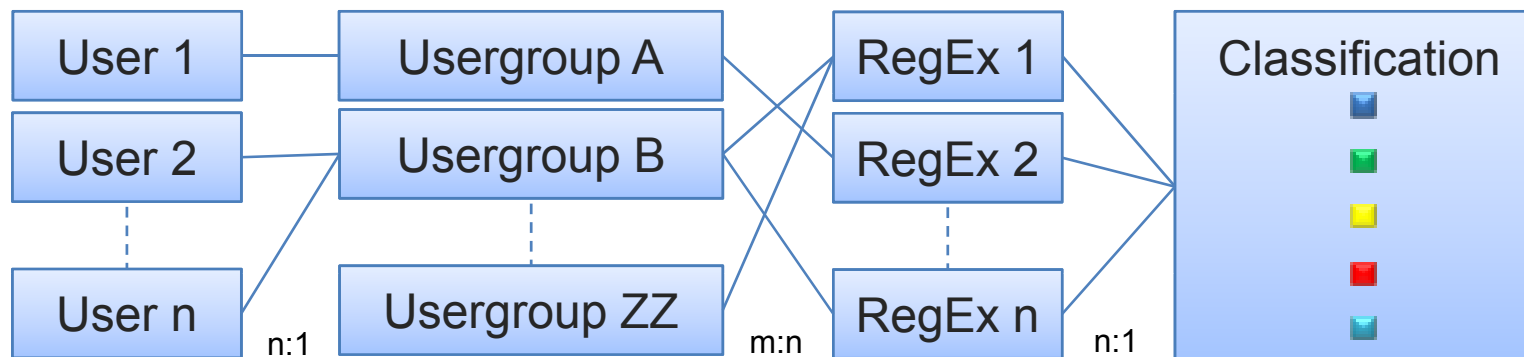


Fictitious distribution!

Internet-Audit: Private Usage Def.

► Definition of private usage within the FMoD (2):

- Sites are classified per second, third, ... level domain as RegEx:
 - **.*?google\.[de|com|fr] : Work**
 - **chat.google\.[de|com|fr] : Private**
 - **.*google-analytics\.com : Neutral**
 - **.*ivbb\.bund\.de : Government**
- All RegEx rules are related to user groups
- Default: all RegEx applies to each user group
- All RegEx rules are ordered; first match decides → like FW rules



Internet-Audit: Measuring time usage

▶ **Discussion with our legal advisors for court proof (plausible) results**

- For how long has a defendant used the internet for private purposes?

▶ **Problem:**

- http is a stateless protocol
- Within the access log there one line per http request
- How long does a user sits actively in front of a loaded webpage?
- Are there any disruptions (ex. phone calls, visitors)?



Internet-Audit: stateless requests

- ▶ How to translate stateless requests (as stated within the proxy farm's access log) to a stateful timeframe in sense of real usage?

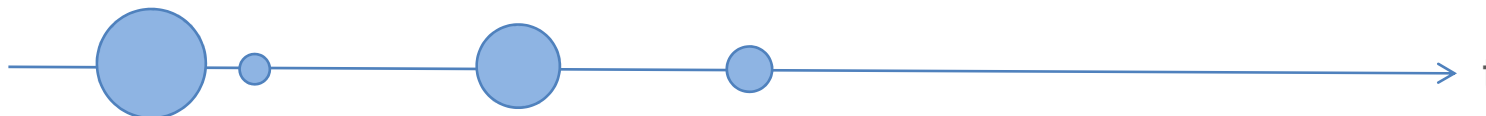
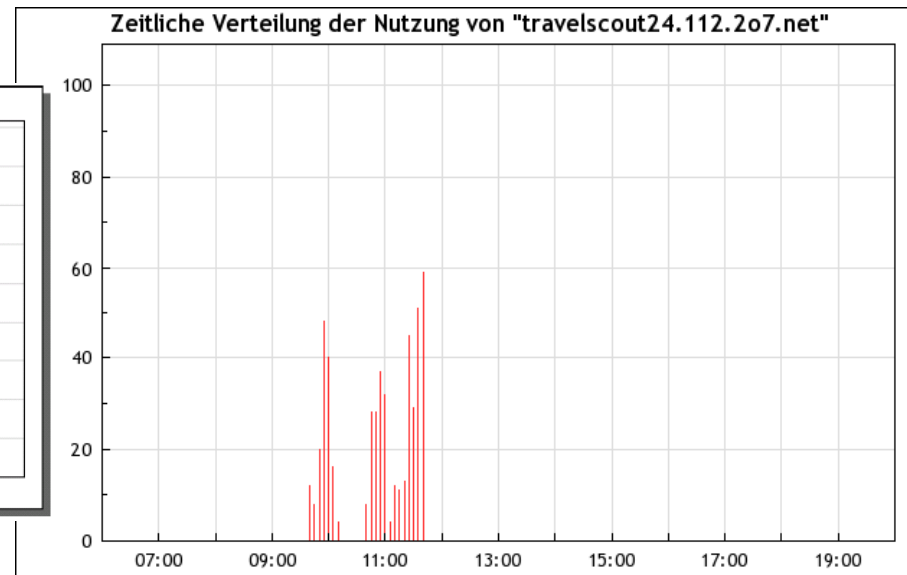
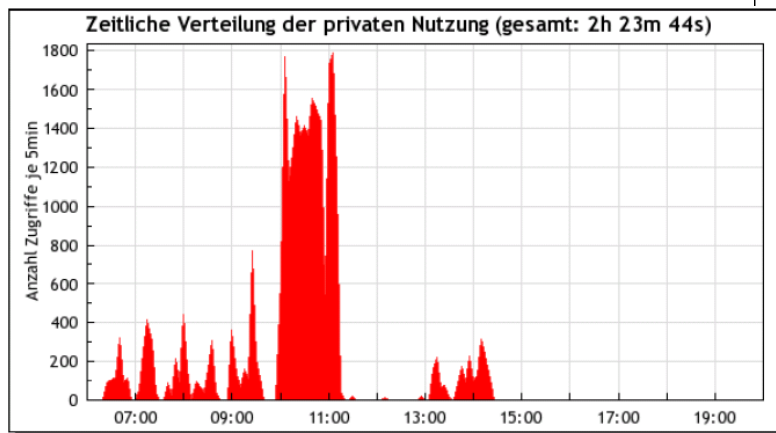
The screenshot displays a Windows desktop environment. On the left, a WordPad window titled 'aro.txt - WordPad' shows a list of access log entries. Each entry contains an IP address, a user identifier, a URL, and browser information. For example: '129.0.65.12 || landreasrohr | ermanwings.com/libraries/yui/...'.

On the right, a PSPad window titled 'PSPad - [D:\Org2_rohr\audit\internet\class.audit.internet.php]' shows PHP code. The code is a class method that processes log files. It includes comments in German: '// prüfen, ob Logfile lesbar und dann einlesen in ein array'. The code uses functions like `gzfile()` to read log files, `explode()` to parse log lines, and `array_key_exists()` to check for user mappings. It also includes error handling for malformed log lines.

```
76 }
77 // prüfen, ob Logfile lesbar und dann einlesen in ein array
78 if (is_readable($this->_log_localpath.$log_filename)) {
79     $users = array();
80     $categories = array();
81     $log_lines = gzfile($this->_log_localpath.$log_filename);
82     $line_counter = 0;
83     $cur_hour_to_check = $timeframe."00";
84
85     $w2d = array(); // website 2 department table
86     $ptc = array();
87
88     foreach($log_lines as $line) {
89         $line_counter++;
90         if ($line_counter % 10000 == 0) {
91             echo date("YmdHis").": ".$line_counter." lines parsed; time with
92 logfile: ".$cur_hour_to_check." \n";
93         }
94         $line = trim($line)." ";
95         $log_parts = explode(" || ", $line);
96         if (!(is_array($log_parts) && count($log_parts) == 9)) {
97             echo "ERROR: skip line (seems to have wrong log format): ".$line.
98 \n";
99         }
100         else {
101             $ip = $log_parts[0];
102             $user = trim(strtolower(str_replace(array("uid=", " ", "cn=", "ou=
103 People,dc=rubin,dc=bwvg,dc=bw,dc=de"), "", $log_parts[1])));
104             if (array_key_exists($user, $this->_user_mapping_user2department)
105 {
106                 $department = $this->_user_mapping_user2department[$user];
107             }
108             else {
109                 $department = $this->_departments["12"];
110             }
111             $department_id = $this->_department2id[$department];
```

Internet-Audit: stateless requests

- ▶ **How to translate stateless requests (as stated within the proxy farm's access log) to a stateful timeframe in sense of real usage?**
 - now we have a set of translated access log entries that belong to a certain user

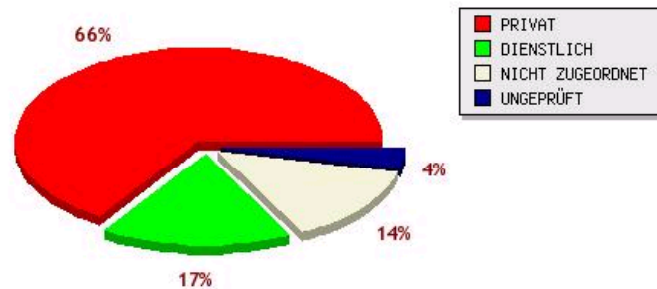


Internet-Audit: Determine time usage

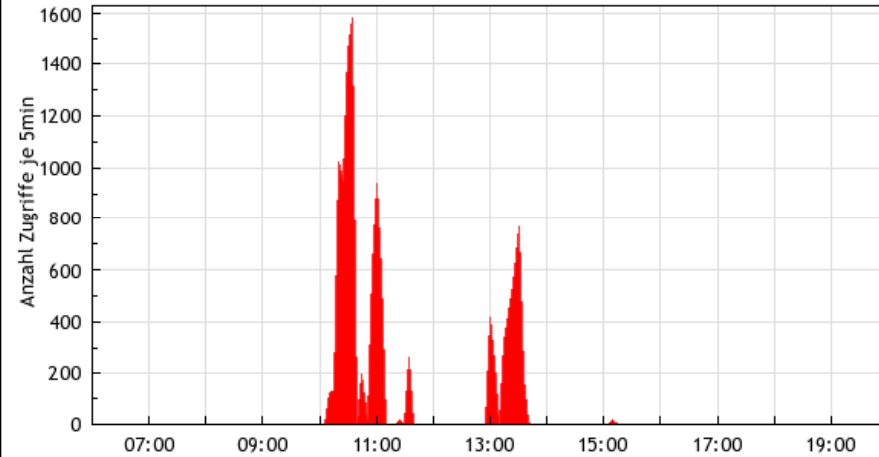
- ▶ **For calculating the time usage we developed a heuristic:**
 - Time usage consists of 1 or more intervals
 - Each interval starts with the first request within this interval
 - An interval spans over requests, where consecutive requests must be within a defined timeframe **TF** (of 1 minute)
 - If this timeframe **TF** is exceeded by two consecutive request the interval is closed and a new interval starts with the least of the two requests
 - To avoid marginal usage to be counted we only “charge” intervals with a length greater than 1 Minute
 - The calculation heuristic ignores the usage part after the last request of an interval

Internet-Audit: Determine time usage

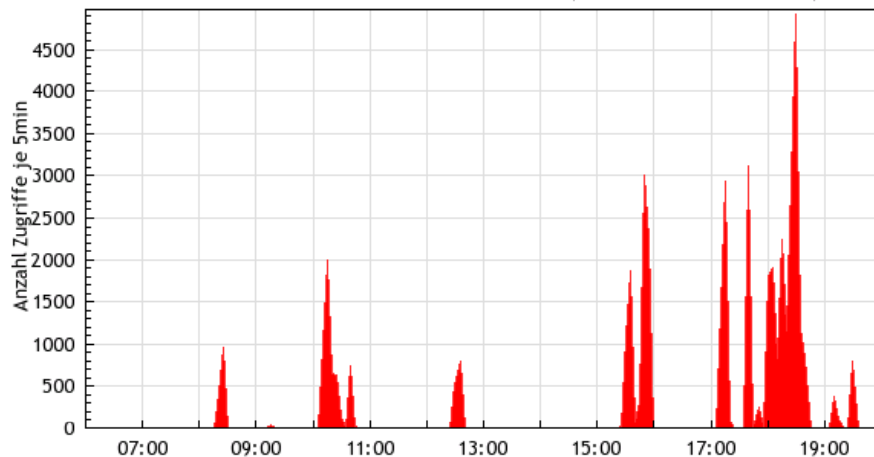
Prozentuale Verteilung der Internetnutzung



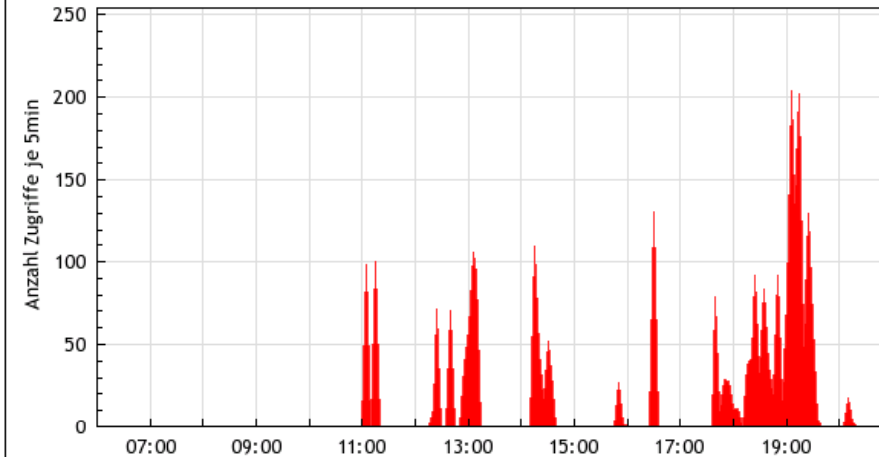
Zeitliche Verteilung der privaten Nutzung (gesamt: 49m 0s)



Zeitliche Verteilung der privaten Nutzung (gesamt: 1h 34m 52s)



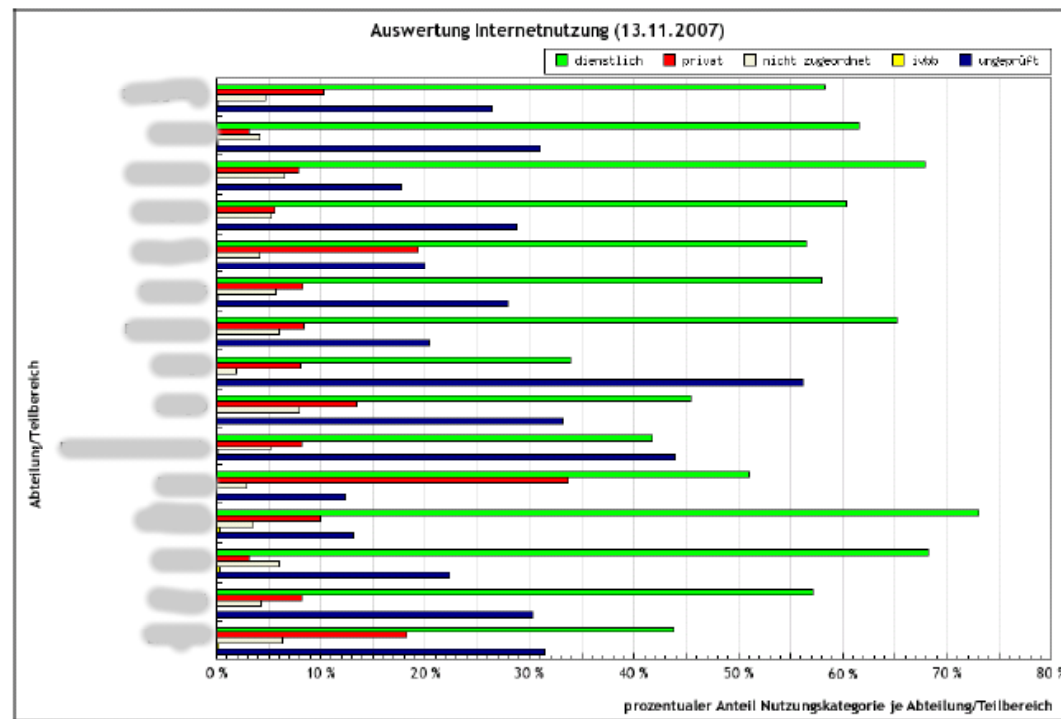
Zeitliche Verteilung der privaten Nutzung (gesamt: 55m 0s)



Internet-Audit: Anonymous Audit

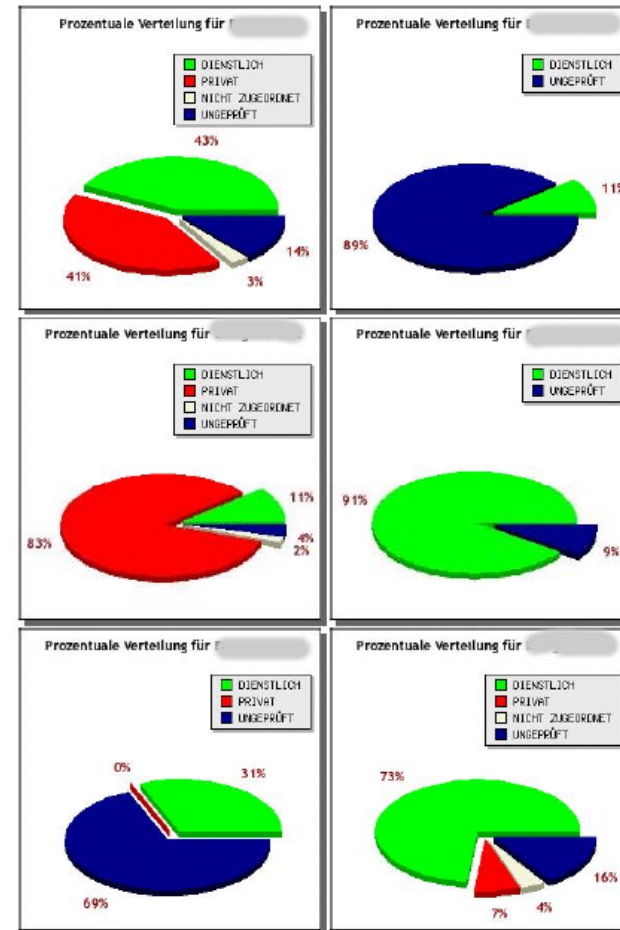
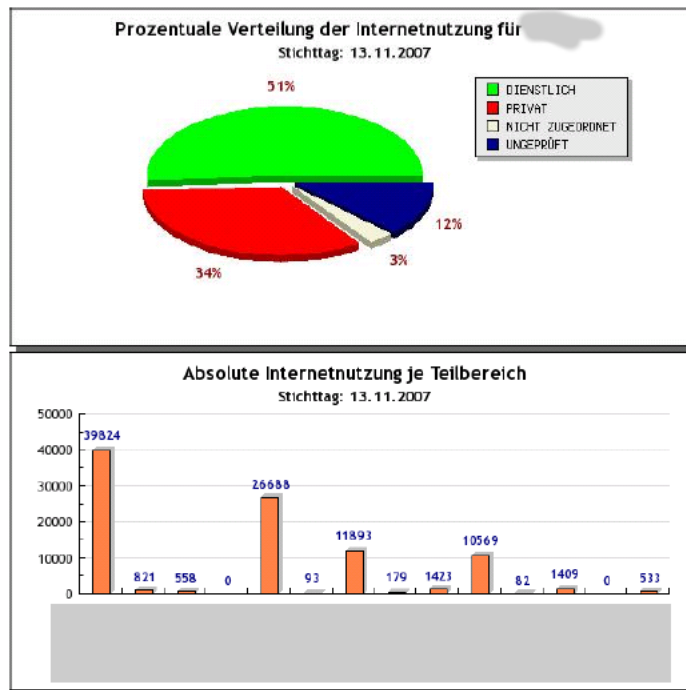
► We audit the internet usage once a day

- We cumulate the distribution of classes for each department
- According to the organisational structure we build the branch and division as well as the overall statistic



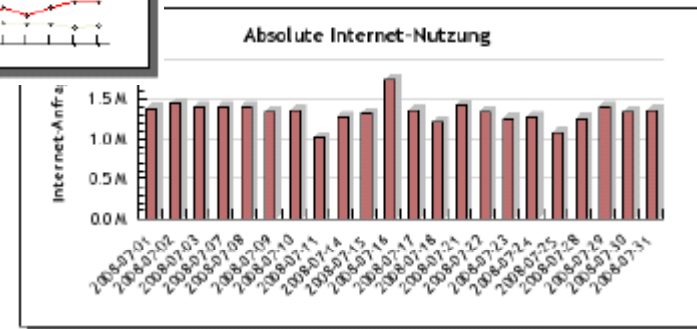
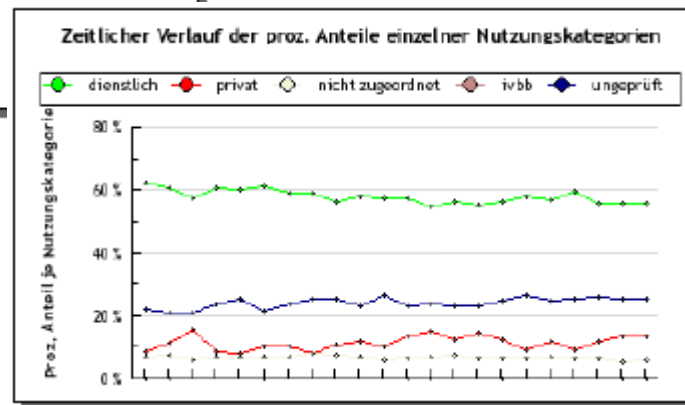
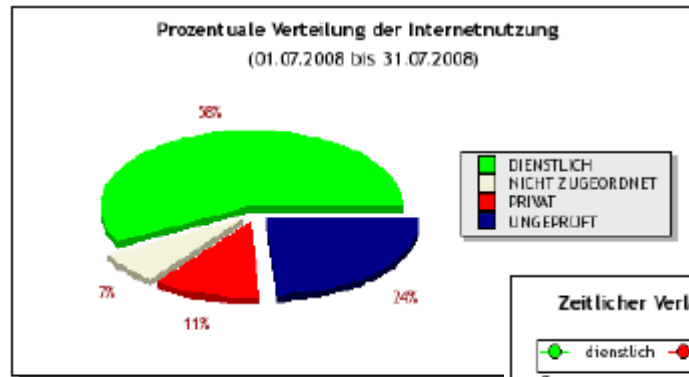
Internet-Audit: Anonymous Audit

► Division drill down to the departments



Internet-Audit: Anonymous Audit

► Executive summery (once a month)



Thanks for attention/participation



Questions ???



Contact:

Andreas Rohr
Federal Ministry of Defense
Fontainengraben 150
53125 Bonn



+49 228 12 9277



andreasrohr (at) bmv.g.bund.de



rohr (at) hsu-hh.de