

CYBER SECURITY DIVISION

Stucco

Situation & Threat Understanding by Correlating Contextual Observations

John Gerth, Stanford University

John R. Goodall, Oak Ridge National Laboratory

January 2014



Homeland
Security

Science and Technology



Team Profile



Need

[**] 1:234:56 IRC - Channel JOIN [**]

[Classification: A Network Trojan was detected]

09/04-17:11:45.456789

10.32.92.230:6667 -> 69.42.215.170:33982 {TCP}

TTL: 34 TOS:0x0 ID:3456 IpLen:20 DgmLen: 44 *****S*

Seq:

TcpL

[Xref

Provides a starting point...
but additional **context** is
necessary to determine impact

Gather information on traffic

```
lip  lipn      rip          ripn          d_port date      start end  fl  tb
-----
-----
mary 10.32.92.230 vmurzlic1.rz.uni-leipzig.de 139.18.17.138 6667
    2012.09.05 01:52 11:43 12  348
mary 10.32.92.230 vmurzlic1.rz.uni-leipzig.de 139.18.17.138 40600
    2012.09.05 00:02 22:07 775 26964
mary 10.32.92.230 undernet.awknet.com          69.42.215.170 33982
    2012.09.05 00:00 22:01 593 48088
```

Gather information on remote host

The screenshot shows a desktop environment with several browser windows. The primary window is ThreatSTOP, displaying the 'Check an IP address' page for IP 69.42.215.170. The page includes a search bar, navigation links, and a table of activity logs. A secondary window shows the Wikipedia page for 'Undernet', and a third window shows the ISC SANS website.

ThreatSTOP | Check an IP address

IP Info: 69.42.215.170 | SANS Internet Storm Center; Cooperative Network Security Community - Internet Security

Check an IP address

You can check IP addresses that appear in your log files against our extensive database of past and current malware by checking a single IP or submitting a device log. At any time you can call (760) 683-8121 or email sales@threatstop.com to learn more about how ThreatSTOP can fit into your network. To find about a single IP, please enter it below.

Research IP 69.42.215.170

First Identified	Most Recently active	Present in the following blockers:
2012-07-16 16:12:59 GMT	2013-08-28 10:23:30 GMT	ShadowServer
2012-07-16 16:12:59 GMT	2013-08-28 10:23:30 GMT	BOTNETS
2012-07-16 16:12:59 GMT	2013-08-28 10:23:30 GMT	BASIC
2012-07-16 16:12:59 GMT	2013-08-28 10:23:30 GMT	ADVANCED
2012-07-10 06:15:48 GMT	2012-07-10 06:15:48 GMT	DSshield Block List
2012-07-10 06:15:48 GMT	2012-07-10 06:15:48 GMT	COMMUNITY
2012-07-10 06:15:48 GMT	2012-07-10 06:15:48 GMT	ADVANCED

Dig info from google DNS

```
>>> Dig 9.8.1-p1 <<<> 88.8.4.4 -x 69.42.215.170
(1 server found)
;; global options: +cmd
;; Got answer:
;; ->HEADER<- opcode: QUERY, status: NOERROR, id: 26
;; flags: qr rd ra; QUERY: 1, ANSWER: 1, AUTHORITY: 0, ADDITIONAL: 0

;; QUESTION SECTION:
; 69.42.215.170. IN A
```

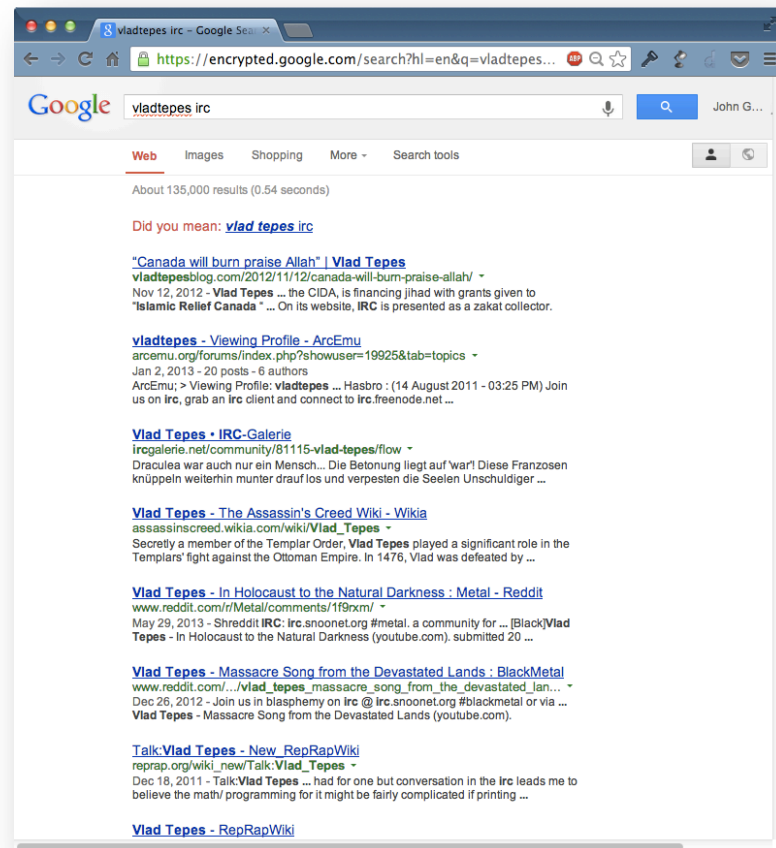

Gather information on processes

\$ top

command not found: top

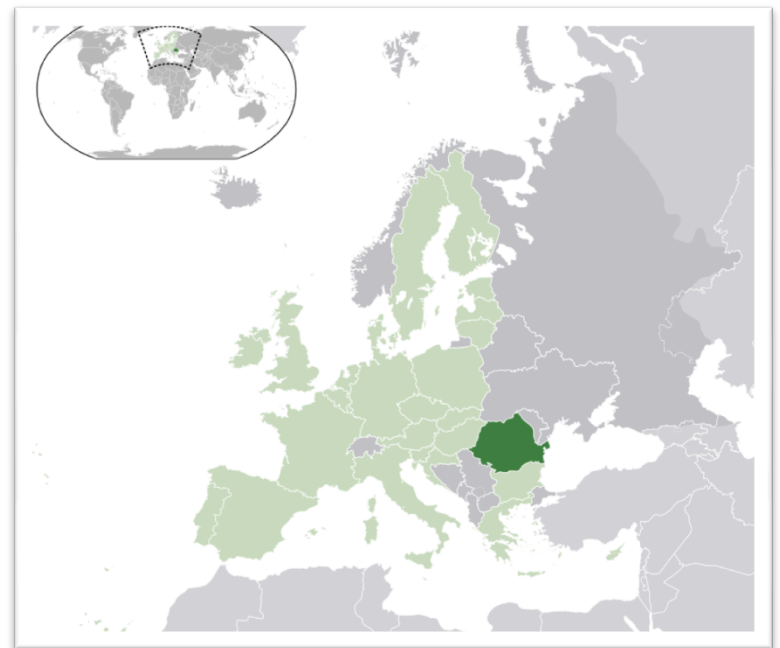
Combination of commands
(netstat ps lsof) shows two
binaries one with an outbound
IRC connection (vladtepes on
port 33982) and one offering
an IRC service (tiresias on port
40600)

Google: 135,000 results. Nothing useful.



Gather information on user logins

aanjneya pts/1	example-09-14108 Wed Sep 5 19:40 - 22:21 (02:40)
franklin pts/0	c-76-126-210-61. Wed Sep 5 18:57 - 19:57 (01:00)
lfyg pts/0	dn0a203a11.exnet Wed Sep 5 10:04 - 10:27 (00:22)
reehj pts/0	c-67-180-35-133. Tue Sep 4 19:51 - 19:52 (00:01)
rebrekm pts/2	kalo.exampled.ed Tue Sep 4 18:08 - 18:11 (00:03)
usoah pts/2	peter-pc.example Tue Sep 4 17:24 - 17:25 (00:00)
avvasm pts/1	dnab4043eb.examp Tue Sep 4 16:29 - 19:16 (02:47)
silakkok pts/0	dnab4046d9.examp Tue Sep 4 16:27 - 18:33 (02:05)
fred pts/0	79-116-146-15.rd Tue Sep 3 13:35 - 14:40 (01:05)
franklin pts/0	70.102.234.3 Tue Sep 4 06:41 - 06:41 (00:00)
cagatay pts/0	dn0a210425.exnet Mon Sep 3 18:32 - 18:33 (00:01)
gnauhccj pts/0	dn0a210240.exnet Mon Sep 3 14:36 - 15:39 (01:02)
srk pts/1	c-98-210-153-100 Mon Sep 3 08:51 - 09:03 (00:11)
msb pts/0	192-119-20-89.pa Mon Sep 3 08:20 - 10:35 (02:15)
fred pts/1	macbocon.example Sun Sep 2 22:39 - 23:57 (01:17)
fred pts/0	macbocon.example Sun Sep 2 21:11 - 22:39 (01:27)
fred pts/1	macbocon.example Sun Sep 2 18:07 - 19:23 (01:15)
fred pts/0	dn5221a5.exnet Sun Sep 2 16:05 - 18:26 (02:21)
thomasjm pts/1	dn0a208bad.exnet Sun Sep 2 15:11 - 17:12 (02:01)
fred pts/0	dn522169.exnet Sun Sep 2 13:17 - 16:00 (02:42)
alerim pts/0	bzq-84-110-37-10 Sun Sep 2 12:19 - 12:19 (00:00)
kbw5 pts/1	c-76-102-15-39.h Sat Sep 1 23:31 - 02:24 (02:53)
fred pts/0	c-67-180-21-231. Sat Sep 1 22:53 - 01:27 (02:34)
fred pts/1	c-67-180-21-231. Sat Sep 1 21:10 - 22:23 (01:13)
reehj pts/0	50-193-59-150-st Sat Sep 1 20:35 - 21:26 (00:51)
fred pts/2	c-67-180-21-231. Sat Sep 1 19:38 - 21:10 (01:31)
msb pts/1	dn5221c4.exnet Sat Sep 1 15:46 - 20:55 (05:08)



Gather information on user activity

Sep 4 13:37:06 mary su[1632]: Successful su for root by root

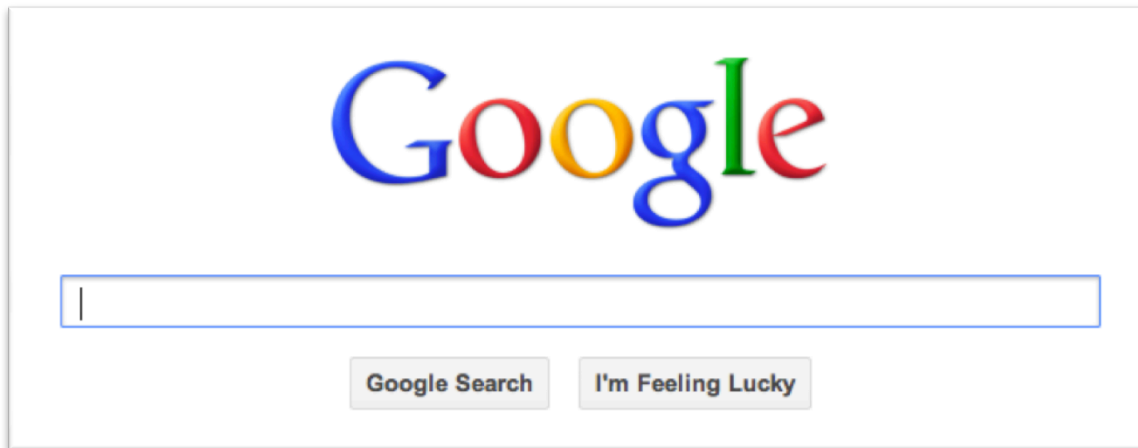
Sep 4 13:37:06 mary su[1632]: /dev/pts/0 root:root

Sep 4 13:37:06 mary su[1632]:
root by fred(u

Weird privilege escalation at a time
when user fred was logged in

Sep 4 14:04:31 mary su[1632]:
user fred

Sep 4 14:04:31 mary su[1632]: pam_unix(su:session): session closed for user root



Problem summary

- Endogenous data
 - Collect log files from multiple hosts
 - Run commands to identify ongoing relevant activity
 - Consult past incident logs for similar attacks
- Exogenous data
 - Search security sites on similar exploits and vulnerabilities
 - Collect information on remote IPs
 - Search blogs and mailing lists for similar events

Analysts need tools that support efficiently identifying, gathering, and synthesizing contextual data to understand and reason about events

Current approaches

- Current tools focus almost exclusively on endogenous data
- Current methods for obtaining context are manual and time-consuming
 - Endogenous data is scattered in a variety of systems
 - Exogenous data can be hidden deep in search results or on forums, in mailing lists, or within APIs
- Current methods are inefficient and take time away from deeper analytical investigation



Approach

Develop a platform to collect contextual data from endogenous and exogenous sources to organize the data into a **knowledge graph** of domain concepts that analysts and other systems can quickly find relevant information

Core components

- Continuous collection and processing of documents from endogenous and exogenous sources
- Domain Specific Language (DSL) for parsing and extracting domain concepts and relationships from structured data
- Natural language processing for extracting domain concepts and relationships from text documents
- Alignment methods for instantiating the knowledge graph
- API for programmatically accessing the graph
- Visualizations for exploring the graph to derive context

Benefits

- More time can be spent analyzing suspicious events and less time spent searching for relevant context
- Context can help analysts make better decisions
- Information can be made available more quickly
- Can perform analytics on the graph to learn new insights
- Public API can be used by other security systems
- Security community can leverage ontology, relevant data sources, labeled data sets and other projects
- Methods and tools may be useful to other domains

Current status

- Draft specification of domain ontology
- Ontology visualization and editing tool
- DSL to parse/transform structured documents into graph
- Proof-of-concept prototype of information extraction for unstructured data sources
 - Method to automatically tag security data to create labeled data sets for supervised learning
 - Complementary approaches for extracting entities based on entropy maximization and bootstrapping
- Demonstration of collecting and processing structured data sources within real-time pipeline

Open source projects

Numerous open-source projects on github.com/stucco

- Ontology: github.com/stucco/ontology
- Ontology editor/vis: github.com/stucco/ontology-editor
- Morph parser/transformer: stucco.github.io/morph/
- Security data sources: stucco.github.io/data/
- Labeled data: github.com/stucco/auto-labeled-corpus
- Demonstration: github.com/stucco/dev-setup



In 2014

- Plans for this year
 - Integrate NLP methods into processing
 - Fill out core functionality: alignment and UI
 - Research relationship extraction methods
 - Iterate on use case, data sources, collectors, extractors
- Technology Transition Activities
 - Publicize ideas to practitioner community

Questions

<http://stucco.github.io/>

John Gerth

Stanford University

gerth@cs.stanford.edu

650-725-3273

John Goodall

Oak Ridge National Laboratory

jgoodall@ornl.gov

865-446-0611

This project is funded by the U.S. Department of Homeland Security Science & Technology Directorate, the Dutch National Cyber Security Centre, and the Defence Research and Development Canada (DRDC) pursuant to the Agreement between the Government of the United States of America and the Government of Canada for Cooperation in Science and Technology for Critical Infrastructure Protection and Border Security. This material represents the position of the authors and not necessarily that of the funders.

The Department of Homeland Security sponsored the production of this material under DOE Contract Number DE-AC05-00OR22725 for the management and operation of Oak Ridge National Laboratory.