# Automatically Tracing Information Flow of Vulnerability and Cyber-Attack Information through Text Strings

*Neil C. Rowe, Eric J. Sjoberg,*

*and Paige H. Adams*

U.S. Naval Postgraduate School

ncrowe@nps.edu

June 2008

# Overview

- Quick defenses to new cyberattacks are critical.
- Vulnerability and attack warnings get disseminated through several CERT sites, MITRE sites, vendor sites, etc.
- But there's little systematic analysis of this information dissemination:
  - Who's copying who, and who originates defense?
  - How fast can a new attack be handled today?
  - Are there bottlenecks in information dissemination?
- We are developing data-mining techniques to analyze this data.
- CVE numbers help track the same vulnerability, but are only used on the most formal pages.
- We are also starting to correlate vulnerability information with observed alerts and packets.

# CMU-CERT provides general info about vulnerabilities

US-CERT Technical Cyber Security Alert TA05-312A -- Microsoft Windows Image Processing Vulnerab - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://www.us-cert.gov/cas/techalerts/TA05-312A.html

National Cyber Alert System

Technical Cyber Security Alert TA05-312A

Archive

## Microsoft Windows Image Processing Vulnerabilities

Original release date: November 08, 2005
Last revised: --
Source: US-CERT

### Systems Affected

- Microsoft Windows 2000
- Microsoft Windows XP
- Microsoft Windows Server 2003

For more complete information, refer to Microsoft Security Bulletin MS05-053

### Overview

Microsoft has released updates that address critical vulnerabilities in Windows graphics rendering services. A remote, unauthenticated attacker exploiting these vulnerabilities could execute arbitrary code or cause a denial of service on an affected system.

### I. Description

The Microsoft Security Bulletin for November 2005 addresses multiple buffer overflows in Windows image processing routines. Viewing a specially crafted image from an application that uses a vulnerable routine may trigger these vulnerabilities. If this application can access images from remote sources, such as web sites or email, then remote exploitation is possible.

Further information is available in the following US-CERT Vulnerability Notes:

VU#680649 - **Microsoft Windows Graphics Rendering Engine buffer overflow vulnerability**

Internet

Start 10:44 AM

Microsoft Outlook Web A... | A:\ | C:\ali;s4675 | Microsoft PowerPoint - [i... | US-CERT Technical Cy...

# CERT Vulnerability Notes have more details

US-CERT Vulnerability Note VU#300549 - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://www.kb.cert.org/vuls/id/300549

Vulnerability Notes Database

Search Vulnerability Notes

Vulnerability Notes Help

Information

View Notes By

Name

ID Number

CVE Name

Date Public

Date Published

Date Updated

Severity Metric

Other Documents

Technical Alerts

## Vulnerability Note VU#300549

### Microsoft Windows Graphics Rendering Engine buffer overflow vulnerability

**Overview**

Microsoft Windows Graphics Rendering Engine contains a buffer overflow that may allow a remote attacker to execute arbitrary code on a vulnerable system.

**I. Description**

The Microsoft Windows Graphics Rendering Engine supports a number of image formats including Windows Metafile (WMF) and Enhanced Metafile (EMF). The Windows Graphics Rendering Engine fails to properly validate WMF and EMF image files. This may allow a remote attacker to manipulate memory allocation routines to create an under-sized buffer. When data is copied to this buffer, a heap-based buffer overflow may occur.

Note that according to public reports, this vulnerability may also affect the Graphical Device Interface (GDI) subsystem.

**II. Impact**

By persuading a user to open a specially crafted WMF or EMF image file, an attacker may be able to execute arbitrary code with the privileges of the user.

**III. Solution**

**Apply an update**

Microsoft has addressed this issue in Microsoft Security Bulletin MS05-053.

Internet

Start 10:47 AM

# CVE provides index numbers on vulnerabilities

CVE-2005-2123 (under review) - Microsoft Internet Explorer

Address: http://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2005-2123

Common Vulnerabilities and Exposures

The Standard for Information Security Vulnerability Names

Home | Get CVE | About CVE | News and Events | Editorial Board | Advisory Council | Compatible Products

## CVE-2005-2123

(under review)

*Additional information is available from the National Vulnerability Database (also sponsored by US-CERT).*

**This is a candidate for inclusion in the CVE list, which standardizes names for security problems. It must be reviewed and accepted by the CVE Editorial Board before it can be added into CVE. Therefore, this candidate may be modified or even rejected in the future.**

| | |
|---|---|
| Name | CVE-2005-2123 (under review) |
| Status | Candidate |
| Description | Multiple integer overflows in the Graphics Rendering Engine (GDI32.DLL) in Windows 2000 SP4, XP SP1 and SP2, and Server 2003 SP1 allow remote attackers to execute arbitrary code via crafted Windows Metafile (WMF) and Enhanced Metafile (EMF) format images that lead to heap-based buffer overflows, as demonstrated using MRBP16::bCheckRecord. |
| | • MISC:http://www.eeye.com/html/research/advisories/AD20051108b.html<br>• MS:MS05-053<br>• URL:http://www.microsoft.com/technet/security/Bulletin/MS05-053.mspx<br>• CERT-VN:VU#300549<br>• URL:http://www.kb.cert.org/vuls/id/300549<br>• BID:15352<br>• URL:http://www.securityfocus.com/bid/15352 |

# Security Focus / Bugtraq collects news on vulnerabilities

SecurityFocus - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://www.securityfocus.com/archive/1

SecurityFocus

» About » Advertising » Contact

Learn how to secure your Web site with the strongest SSL encryption possible. VeriSign® SSL Services.

VeriSign Secured

Get a Free guide to Web site security.

[E-Mail Address]

Submit

Symantec ThatCon

Threat level definition

Home | Bugtraq | Vulnerabilities | Mailing Lists | Security Jobs | Tools | Search:

News
Infocus
Foundations
Microsoft
Unix
IDS
Incidents
Virus
Pen-Test
Firewalls
Columnists
Mailing Lists
Newsletters
Bugtraq
Focus on IDS
Focus on Linux
Focus on Microsoft
Forensics
Pen-test
Security Basics
Vuln Dev
Vulnerabilities
Jobs
Job Opportunities
Resumes
Job Seekers

BugTraq

Mode: Threaded Go

Expand all | Post message

(Page 1 of 528) 1 2 3 4 5 6 7 8 9 10 11 Next >

- [KDE Security Advisory] kjs encodeuri/decodeuri heap overflow 2006-01-19
  Dirk Mueller (mueller kde org)
- phpXplorer file inclusion bivosecurity.be 2006-01-18
  ilz0 bsdmail com
- iDefense Security Advisory 01.17.06: Cisco Systems IOS 11 Web Service CDP Status Page Code Injection Vulnerability 2006-01-17
  labs-no-reply@idefense.com (labs-no-reply idefense com)
- iDefense Security Advisory 01.17.06: EMC Legato Networker nsrexecd.exe Heap Overflow Vulnerability 2006-01-17
  labs-no-reply@idefense.com (labs-no-reply idefense com)
- iDefense Security Advisory 01.17.06: EMC Legato Networker nsrd.exe DoS Vulnerability 2006-01-17
  labs-no-reply@idefense.com (labs-no-reply idefense com)
- Re: Microsoft knew about the WMF flaw for years 2006-01-17
  Steven M. Christey (coley mitre org)
- Oracle Database 10g Rel. 1 - SQL Injection in SYS.KUPV$FT_INT 2006-01-17
  ak red-database-security com
- Oracle Database 10g Rel. 1 - SQL Injection in SYS.KUPV$FT 2006-01-17

SCAN YOUR WEBSITE

VeriSign SSL Certificates secure e-commerce transactions.

Master of Science in Information Assurance - Online -

Customize

Internet

Start

10:57 AM

# Example Bugtraq vulnerability description

SecurityFocus - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://www.securityfocus.com/archive/1/415463/30/1050/threaded

BugTraq

Back to list | Post reply

Simple PHP Blog: Multiple XSS Vulnerabilities Nov 02 2005 12:14PM
enji infosys tuwien ac at

```
==========================================================
Simple PHP Blog: Multiple XSS Vulnerabilities
==========================================================
Technical University of Vienna Security Advisory
TUVSA-0511-001, November 2, 2005
==========================================================

Affected applications
---------------------

Simple PHP Blog (www.simplephpblog.com)

Versions 0.4.5 and prior.

Description
-----------

1.) preview_cgi.php, part 1

There is a cross-site scripting (XSS) vulnerability in the file preview_cgi.php on line 126:
The variable $entry is echoed,
but hasn't been sufficiently sanitized before. When logged in, this issue can be tested
with the following URL:

http://your-server/path-to-sphpblog/preview_cgi.php?entry=foo"><script>a
lert(document.cookie)</script>

The fields "your-server" and "path-to-sphpblog" in the given URL have to be adjusted
accordingly.

2.) preview_cgi.php, part 2

There is another cross-site scripting vulnerability in the file preview_cgi.php on line 129:
The variable $temp_subject is echoed,
```

SCAN YOUR WEBSITE

VeriSign SSL Certificates secure e-commerce transactions.

Master of Science in Information Assurance - Online -

Customize your degree:

- Emergency Management
- Business Continuity Planning
- Digital Investigations

# Bugtraq description example, second screenful

SecurityFocus - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://www.securityfocus.com/archive/1/415299/30/1050/threaded

Home | Bugtraq | Vulnerabilities | Mailing Lists | Security Jobs | Tools | Search:

News
Infocus
Foundations
Microsoft
Unix
IDS
Incidents
Virus
Pen-Test
Firewalls
Columnists
Mailing Lists
Newsletters
Bugtraq
Focus on IDS
Focus on Linux
Focus on Microsoft
Forensics
Pen-test
Security Basics
Vuln Dev
Vulnerabilities
Jobs
Job Opportunities
Resumes
Job Seekers
Employers
Tools
RSS
News
Vulns
Product Search

BugTraq

Back to list [ Post reply

SQL IN FORUM.PHP Oct 30 2005 12:03PM
ABDUCTER_MINDS YAHOO.COM

```
Class: Input Validation Error
CVE: CVE-MAP-NOMATCH
Remote: Yes
Discovered BY ABDUCTER & Explict BY DEVIL-00
ABDUCTER_MINDS (at) S4A (dot) CC [email concealed] (OR) ABDUCTER_MINDS (at)
YAHOO (dot) COM [email concealed]
Vulnerable:powered by oaboard 1.0
//////////////////////////////////
info:- FOR INFORMATION VISIT http://oaboard.myserver.at/oaboard/forum.php
//////////////////////////////////
discussion: THERE IS SQL IN FORUM.PHP
************************************
EXPLIOTS AND EXAMPLE
====================
//-------1---------//
http://WWW.VICTIM.COM/oaboard/forum.php?modul=topics&channel=[SQL]
http://WWW.VICTIM.COM/oaboard/forum.php?modul=topics&channel=-99%20UNION
%20SELECT%20null,password%20FROM%20pw99_user%20WHERE%20id=1
//-------2---------//
http://WWW.VICTIM.COM/oaboard/forum.php?modul=posting&topic=[SQL]&channe
l=3
http://oWWW.VICTIM.COM/oaboard/forum.php?modul=posting&topic=30%20UNION%
20SELECT%20null,username,null,password%20FROM%20pw99_user%20WHERE%
20id=1
/*&channel=3
************************************
CREDITS S4A.CC FOR ALL GEEKS
FOR AL ARAB
HACKER PAL
MY LOVE (NONO)
```

[ reply ]

SCAN YOUR WEBSITE

VeriSign SSL Certificates secure e-commerce transactions.

Master of Science in Information Assurance - Online -

Customize your degree:

- Emergency Management
- Business Continuity Planning
- Digital

Done | Internet

Start | Microsoft Outlo... | A:\ | C:\jalcs4675 | Microsoft Powe... | SecurityFocus... | tengpics.doc - ... | 1:40 PM

# Bugtraq links to further details on other sites

The Magic of magic byte - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://www.securityelf.org/magicbyte.html

Links Customize Links Free Hotmail Windows Free AOL & Unlimited Internet Windows Media personal RealPlayer Enterprise

Here is the **Triple Headed program** which has 3 different 'execution entry points', depending on the extension of the file (exe, html or eml) – just change the extension and the SAME file will be executed by (at least) THREE DIFFERENT programs! (Thanks to contributing author Wayne Langlois from www.diamondcs.com.au for providing this very nice program).

The source (asm):

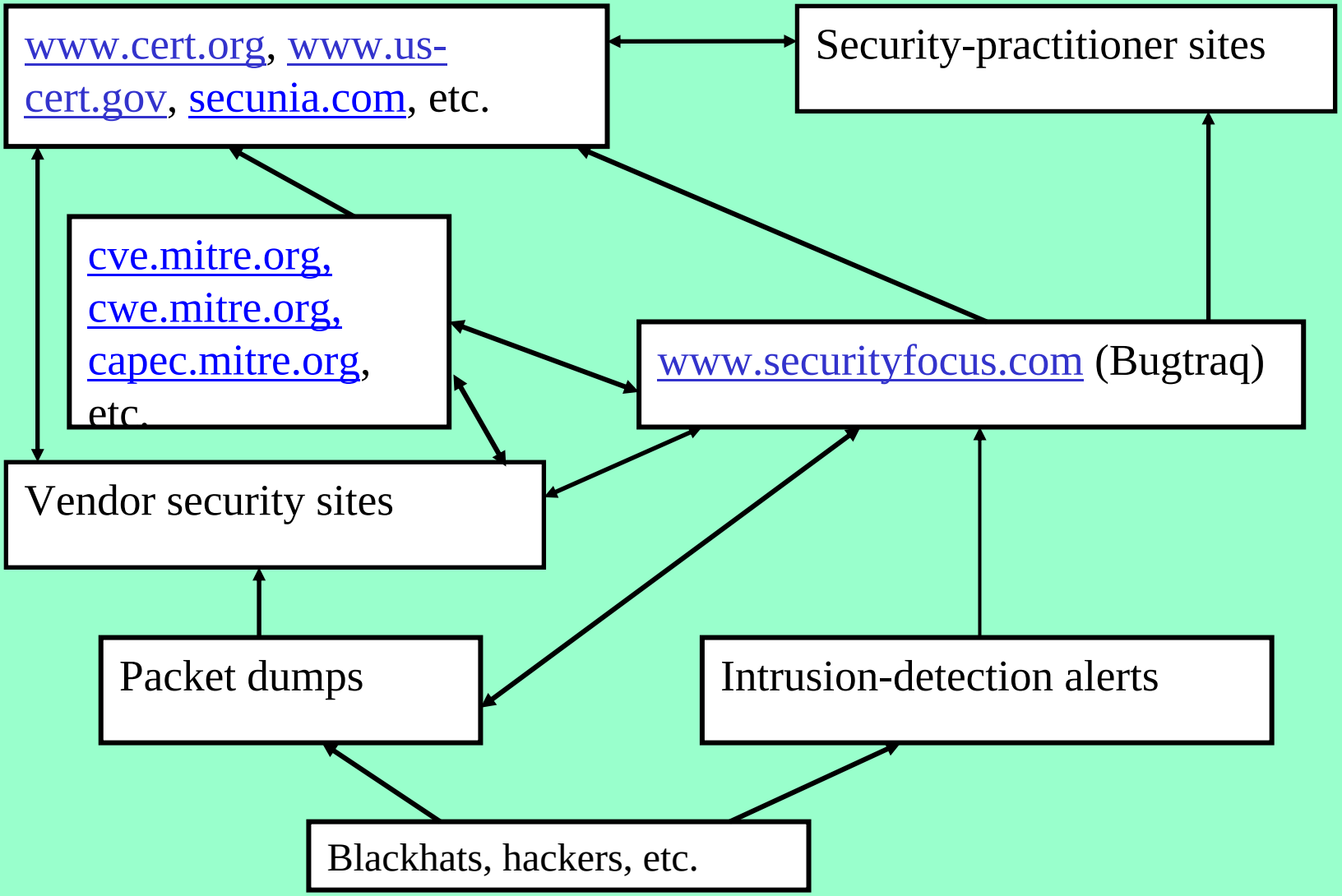
```
; EXE-HTM Hybrid
; by Wayne Langlois (www.diamondcs.com.au)
;
include 'e:\dev\asm\fasm162\include\win32ax.inc'
;
_HTML db 'Reply-To: <wayne@local>',0Dh,0Ah,'From: <wayne@local>',0Dh,0Ah,'To: <wayne@local>',0Dh,0Ah,'Subject: My subject',0Dh,0Ah,\
    'Content-Type: multipart/alternative; boundary="--Boundary",0Dh,0Ah,\
    0Dh,0Ah,\
    '----Boundary',0Dh,0Ah,\
    0Dh,0Ah,\
    '<html><body>',0Dh,0Ah,\
    '<script language="vbscript">Msgbox "Hello from the HTML component!"</script>',0Dh,0Ah,\
    '</body></html>',0Dh,0Ah,\
    '----Boundary',0Dh,0Ah,\
    'Content-Type: text/html',0Dh,0Ah,\
```

Done Internet

Start 1:49 PM

Microsoft Outlo... A:\ C:\jalcs4675 Microsoft Powe... The Magic of ... tengpics.doc - ...

# Major data flows of vulnerability information



# First approach to tracking flow of information security information: Keyword matching of Web pages

- Collect security keywords related to alerts.
- Send a subset of them to a browser site like Google.
- Collect the URLs of the top matches.
- Retrieve the pages of the top matches.
- Compare words of each sentence on a page to each sentence on another page.
- Exclude structural “stop words” (e.g. “the”, “in”, “then”, “system”).
- Find all very-close matches between sentences.

# Classification of exact sentence matches

- Normal routine copying of pages, as when a Web site collects important papers on security.
- Common authorship on sites, e.g. www.demboo.info/Carbon-cheats.htm and www.mulax.info/Sims-cheats.htm both say “For walkthroughs, cheats and tips call 09067 53 54 55."
- Acknowledged citation, particularly if the text is quoted or indented.  This can be distinguished by words such as "says“ and "stated".
- Plagiarism.
- "Boilerplate", formalized statements for some legal or policy objective, e.g. "Additional Information: For the most up-to-date information regarding these vulnerabilities, please visit the CERT/CC Vulnerability Notes Database at: http://www.kb.cert.org/vuls/".

# Classification of inexact sentence matches

- Common authorship, e.g. "IGN is the ultimate Spider-Man: The Movie resource for trailers, screenshots, cheats , walkthroughs" versus "IGN PS2 is the ultimate resource for PlayStation 2 trailers, screenshots, cheats, walkthroughs".
- Acknowledged citation.
- Acknowledged paraphrase, e.g. "According to the US-CERT there is publicly available exploit code for multiple vulnerabilities in Sun Java Runtime Environment (JRE)" .
- Unacknowledged paraphrase.
- Boilerplate, e.g. “Further information is available in the following US-CERT Vulnerability Note".
- Accidental similarities, e.g. "An attacker could use a specially crafted web page to exploit the vulnerability and take control of a system, warned Danish security firm Secunia" and "An attacker could exploit these vulnerabilities by using specially crafted network traffic, by convincing you to click on a specially crafted URL, or by convincing you to open a specially crafted Office document".

# Calculate similar sites from similar sentences

**For keywords “vulnerability”, “ICMP”, “packets”, “flags”, and “footprinting”:**

34.966: www.ecst.csuchico.edu to www.yolinux.com

34.943: www.e-infomax.com to www.ecst.csuchico.edu

34.943: www.ecst.csuchico.edu to www.uni-kiel.de

34.943: www.linuxdig.com to www.ecst.csuchico.edu

31.713: docs.mandragor.org to www.ecst.csuchico.edu

31.696: www.ecst.csuchico.edu to www.arameya.com

28.713: www.e-infomax.com to www.uni-kiel.de

28.713: www.linuxdig.com to www.e-infomax.com

28.713: www.linuxdig.com to www.uni-kiel.de

28.710: www.e-infomax.com to www.yolinux.com

28.710: www.linuxdig.com to www.yolinux.com

28.710: www.uni-kiel.de to www.yolinux.com

28.426: www.cs.wisc.edu to www.ecst.csuchico.edu

27.397: docs.mandragor.org to www.arameya.com

26.796: docs.mandragor.org to www.e-infomax.com

26.796: docs.mandragor.org to www.linuxdig.com

26.796: docs.mandragor.org to www.uni-kiel.de

26.793: docs.mandragor.org to www.yolinux.com

26.793: www.e-infomax.com to www.arameya.com

26.793: www.linuxdig.com to www.arameya.com

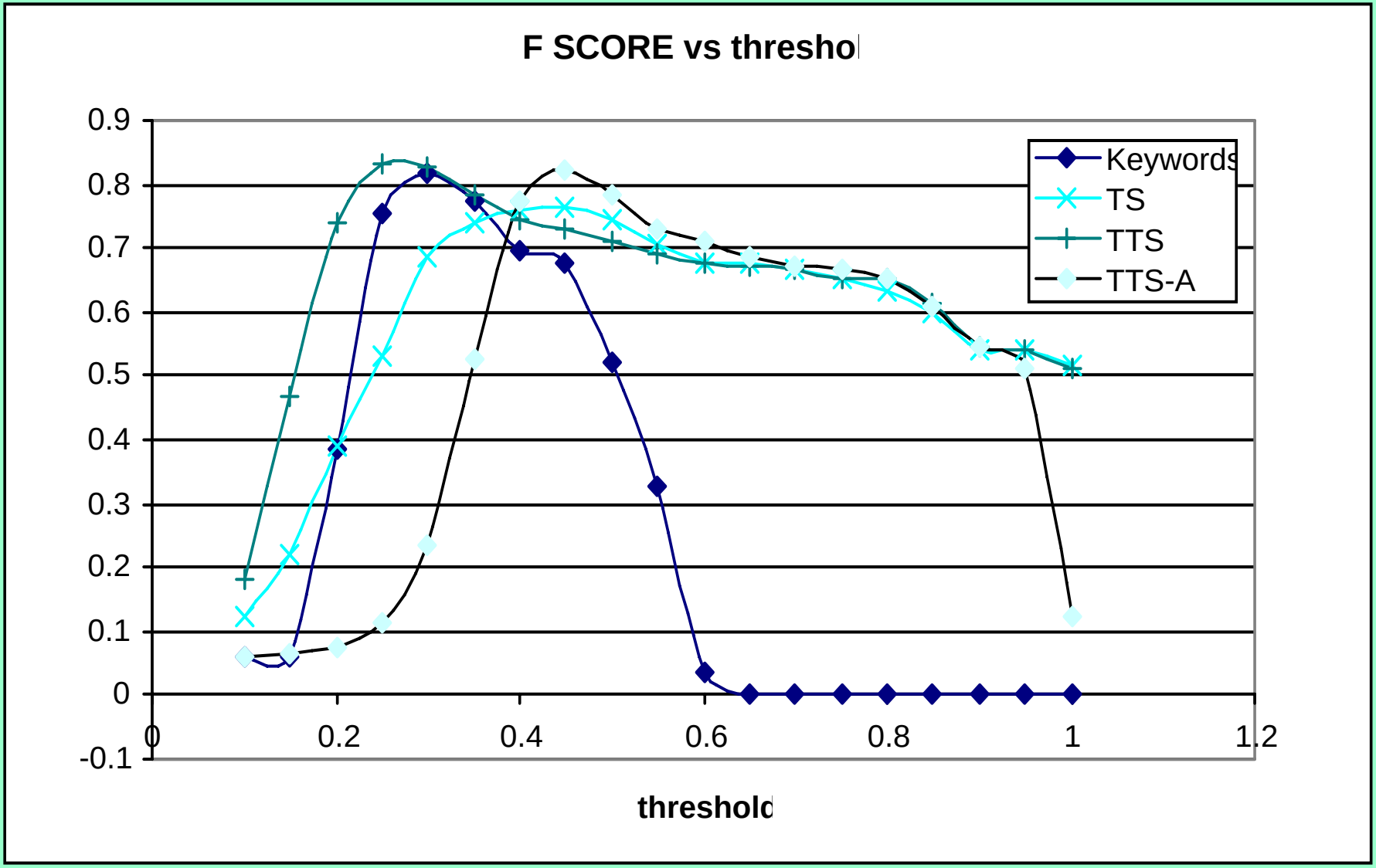
26.793: www.uni-kiel.de to www.arameya.com

26.788: www.arameya.com to www.yolinux.com

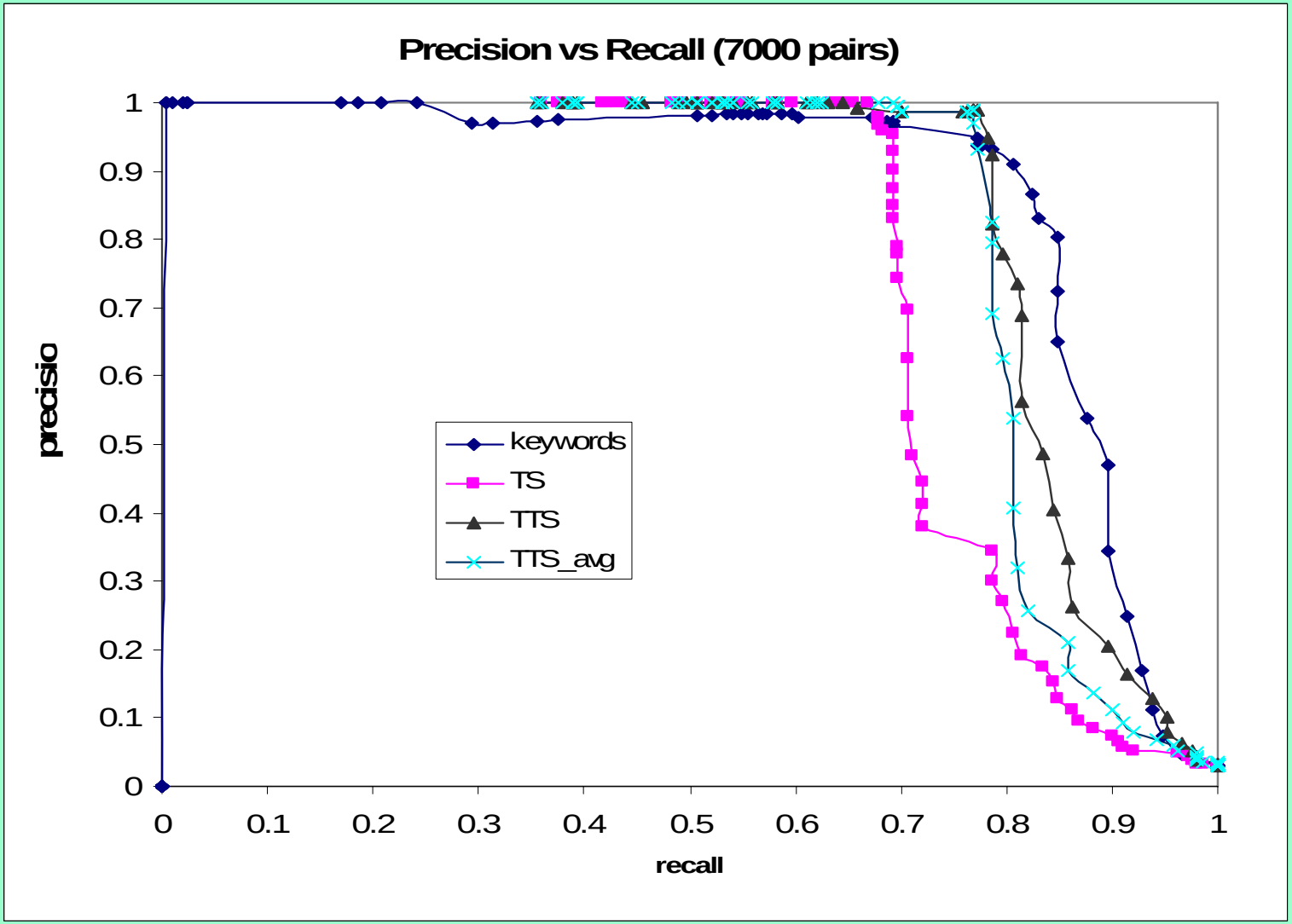
# Deeper analysis of page similarity

- We can do better on comparing sentences if we know the parts of speech used.
- This suggests using a tagger (we used the Brill one).
- We compared (1) matching keywords only, (2) matching tags only (ignoring tag order), (3) matching both keywords and tags, (4) extending 3 to include bigram matches.
- Methods (3) and (4) performed the best, but (4) was not significantly better than (3), hence bigrams are not necessary for good performance.

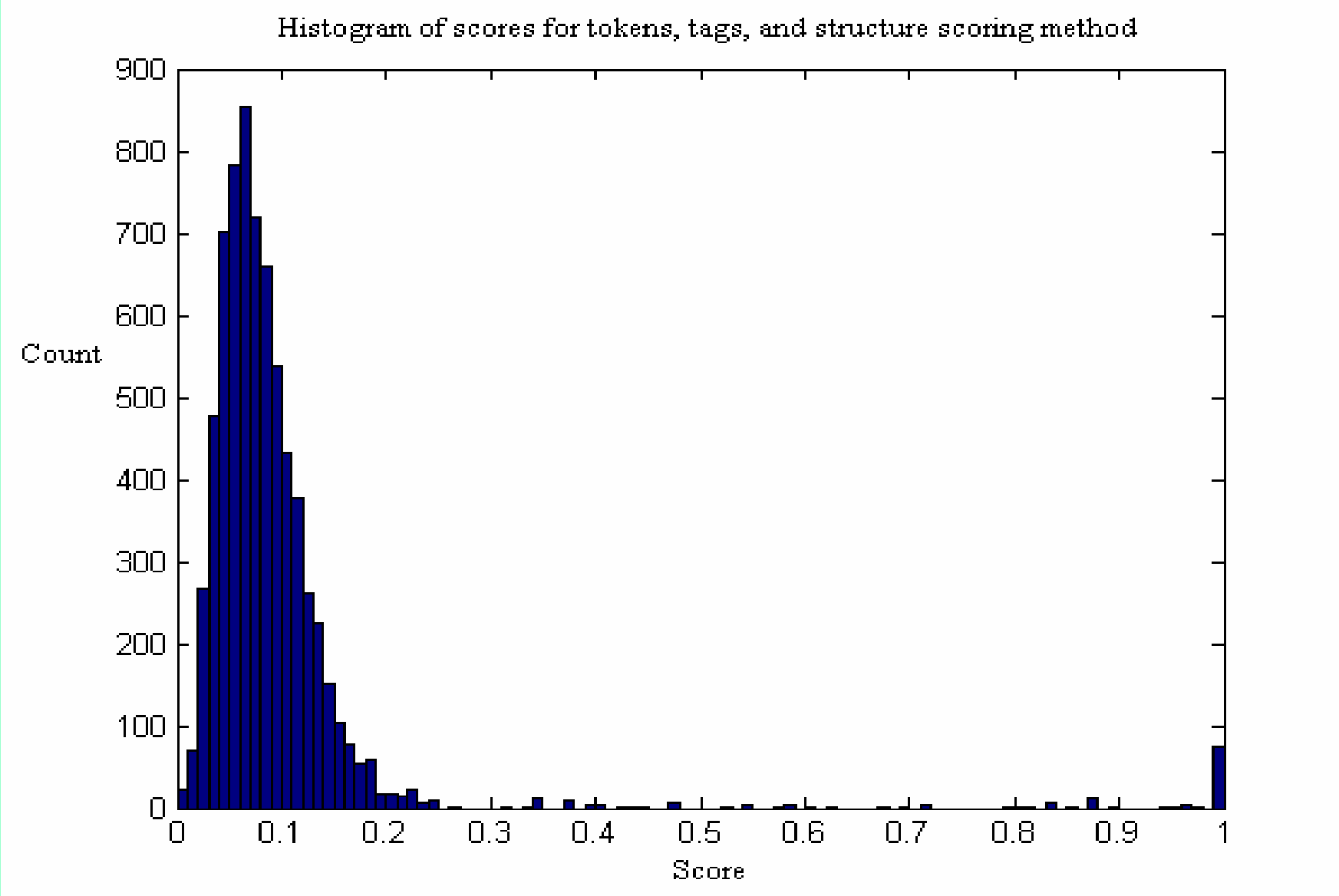
# Performance is better with tagging



# Another view of the data



# The distribution of random sentence matches



# Match failures suggest future directions

"One strain of scam email makes the bogus claim that recipients have won one of the much sought after devices in a bid to trick prospective marks into visiting a malware loaded site."

"Email recipients are sent a bogus email informing them that they have won a new iPhone, in reality the email contained malware designed to subvert and compromise the user's computer."

"NanoScan is a rapid, light scanner that currently detects over 750,000 active viruses, spyware, Trojans and other malware within just one minute."

"Panda Software has launched the mini, customisable version of NanoScan, the instant virus scanner from Panda Software, designed to detect active malware on a PC in less than one minute."

"According to a 2005 FBI Cyber Crime Study, 90 percent of small businesses had at least one cyber security incident within the past year."

"In fact, of the 500 companies that responded to a recent FBI survey, 90 percent said they'd had a computer security breach, and 80 percent of those said they'd suffered financial loss as a result."

"Using ideological attraction, the Soviets successfully recruited many high-level spies."

"At that time period the Soviets recruited their spies using ideological motivation."

# Using more direct attack data

- Alerts and vulnerability notes are secondary-source information.
- More direct information would be the output of intrusion-detection systems.
- We can pull text related to specific alerts that are noticed on a system and find attack trends.
- Still more direct are packets themselves.
- Many of these contain text strings that can suggest attack trends.

# Example intrusion-detection system text

Consider Snort alert from our honeypot: *Date: 2007-09-12 Time: 15:46:56.148-07 Alert_code: 1394 Alert_description: SHELLCODE x86 NOOP IP_address_1: 89.26.217.22 Port_#_1: 4310 IP_address_2: 192.168.0.3 Port_#_2: 445 time_to_live: 118*

This is co-referenced with the text:

- **Message** SHELLCODE x86 NOOP
- **Summary** This event is generated when an attempt is made to possibly overflow a buffer. The NOOP warning occurs when a series of NOOP (no operation) are found in a stream. Most buffer overflow exploits typically use NOOPs sleds to pad the code.
  **Impact** This might indicate someone is trying to use a buffer overflow exploit. Full compromise of system is possible if the exploit is successful.
  **Detailed Information** This rule detects a large number of consecutive NOOP instructions used in padding code. It's not specific to a particular service exploit, but rather used to try and detect buffer overflows in general. It is common for buffer overflow code to contain a large sequence of NOOP instructions as it increases the odds of successful execution of the useful shellcode.
  **Affected Systems** Any x86 programs.
  **Attack Scenarios** An attacker uses a buffer overflow exploit which contains the following payload: 90 90 90 90 90 90 90 90 90 90 /bin/sh
  **Ease of Attack** Simple.
  **False Positives** High, This event may be generated by applications such as ftp and http when binary data is being transferred. A false Positive can be generated if the snort sensor detects text from an IRC client or any other application that passes data plaintext. The event is generated if Snort detects several (a) characters in a row - such as 'aaaaaaaaaa'.

# Example packet text strings

```
09/14-00:47:21.626361 131.120.18.41:53 -> 192.168.0.3:3559 UDP
   TTL:111 TOS:0x0 ID:15349 IpLen:20 DgmLen:145 Len: 117
47 59 81 83 00 01 00 00 00 01 00 00 02 67 63 06  GY...........gc.
5F 6D 73 64 63 73 08 55 53 4E 42 41 52 4F 4E 05  _msdcs.USNBARON.
6C 6F 63 61 6C 00 00 06 00 01 00 00 06 00 01 00  local...........
00 00 00 00 40 01 41 0C 52 4F 4F 54 2D 53 45 52  ....@.A.ROOT-SER
56 45 52 53 03 4E 45 54 00 05 4E 53 54 4C 44 0C  VERS.NET..NSTLD.
56 45 52 49 53 49 47 4E 2D 47 52 53 03 43 4F 4D  VERISIGN-GRS.COM
00 77 A1 C8 65 00 00 07 08 00 00 03 84 00 09 3A  .w..e...........:
80 00 01 51 80                                   ...Q.

09/16-22:43:13.038582 131.120.18.41:53 -> 192.168.0.4:1052
UDP TTL:111 TOS:0x0 ID:13512 IpLen:20 DgmLen:137 Len: 109
FF FA 81 83 00 01 00 00 00 01 00 00 08 64 6F 77  .............dow
6E 6C 6F 57 64 0D 77 69 6E 64 6F 77 73 75 70 64  nloWd.windowsupd
61 74 65 03 63 6F 6D 00 00 01 00 01 C0 15 00 06  ate.com.........
00 01 00 00 0E 10 00 35 03 6E 73 31 04 6D 73 66  .......5.ns1.msf
74 03 6E 65 74 00 06 6D 73 6E 68 73 74 09 6D 69  t.net..msnhst.mi
63 72 6F 73 6F 66 74 C0 23 77 A1 A5 3D 00 00 03  crosoft.#w..=...
84 00 00 02 58 00 09 27 C0 00 00 03 84           ....X..'.....
```

# The “MARB MEOW” strings

```
05 00 00 03 10 00 00 00 A8 06 00 00 E5 00 00 00  ................
90 06 00 00 01 00 04 00 05 00 06 00 01 00 00 00  ................
00 00 00 00 32 24 58 FD CC 45 64 49 B0 70 DD AE  ....2$X..EdI.p..
74 2C 96 D2 60 5E 0D 00 01 00 00 00 00 00 00 00  t,..`^..........
70 5E 0D 00 02 00 00 00 7C 5E 0D 00 00 00 00 00  p^......|^......
10 00 00 00 80 96 F1 F1 2A 4D CE 11 A6 6A 00 20  ........*M...j. 
AF 6E 72 F4 0C 00 00 00 4D 41 52 42 01 00 00 00  .nr.....MARB....
00 00 00 00 0D F0 AD BA 00 00 00 00 A8 F4 0B 00  ................
20 06 00 00 20 06 00 00 4D 45 4F 57 04 00 00 00   ... ...MEOW....
A2 01 00 00 00 00 00 00 C0 00 00 00 00 00 00 46  ................F
38 03 00 00 00 00 00 00 C0 00 00 00 00 00 00 46  8...............F
00 00 00 00 F0 05 00 00 E8 05 00 00 00 00 00 00  ................
01 10 08 00 CC CC CC CC C8 00 00 00 4D 45 4F 57  ............MEOW
E8 05 00 00 D8 00 00 00 00 00 00 00 02 00 00 00  ................
```

# Conclusions

- Information transmission is important with security alerts and other security intelligence.
- The transmission infrastructure has been built informally – and it mostly works.
- But there can be bottlenecks and redundancies.
- And an adversary could exploit weaknesses or attack the infrastructure itself.
- We need data on the infrastructure to make good decisions about it.
- We’re just starting this research.