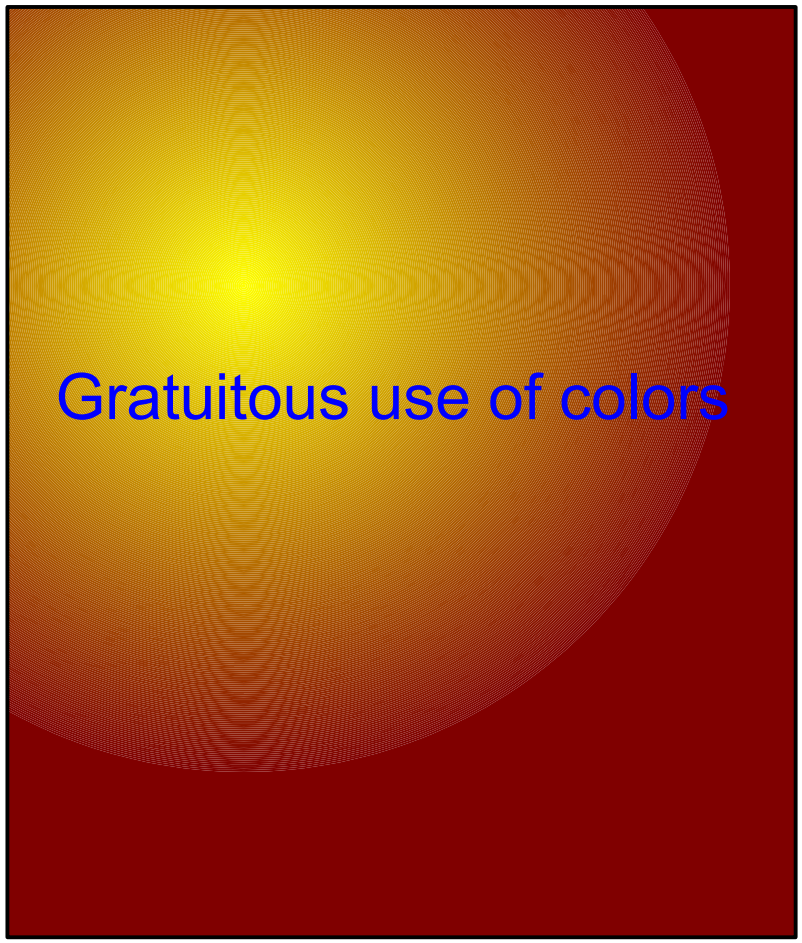
Bulk Email Forensics

HTCIA - Aug 22-25, 2009

Dr. Fred Cohen
President - California Sciences Institute
CEO – Fred Cohen & Associates

Outline

- Background of the speaker and subject
- Some of the forensic challenges
- Tools and techniques
- Claims and counterclaims
- Damage assessment
- Making and breaking a case
- Questions / Comments?



Your speaker

- Education:
  - B.S. Electrical Engineering (C-MU '77)
  - M.S. Information Science (Pitt '81)
  - Ph.D. Electrical Engineering (USC '86)
- Experience:
  - >30 years of information protection R&D, design, engineering, testing, implementation, and operation
  - >20 years since first digital forensics case
- CEO - Fred Cohen & Associates
  - Enterprise information protection architecture
  - Digital forensics for high-valued legal cases

CalSci

- President – California Sciences Institute
  - Starting doctoral classes in 2009-10
- M.S. And Ph.D. Program in National Security
  - Technical aspects of these fields
- M.S. In Advanced Investigation
- Ph.D. In Digital Forensics
  - The first Ph.D. program in Digital Forensics in the United States
- calsci.org

What does he know about the subject?

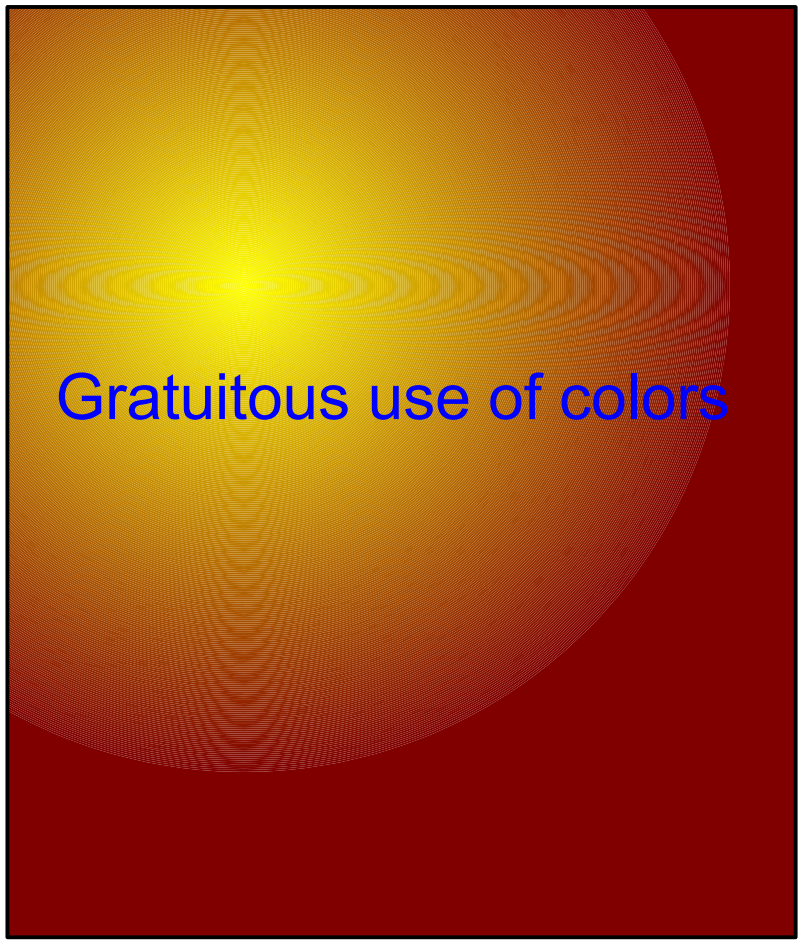
- Knowledge, skill, experience, training, or education
  - Federal Rules of Evidence 701-706 (email forensics)
- Knowledge, Skills, and Experience:
  - Created commercial email servers and processing, industry analyst analyzing these mechanisms, operates email processing systems for high volumes of email, teaches, lectures, etc. in the area, POST certified trainer in these areas, admitted to testify as an expert in Federal, State, and Local Criminal and Civil digital forensics matters including related to emails, published refereed and other articles, etc.
- Education:
  - B.S., M.S., and Ph.D. in relevant field

A case to exemplify the issues

- A case study will be used to show the issues
  - ASIS Internet Services vs. Optin Global, Inc., Case No. C-05-5124 JCS in the US District Court, Northern District of California
- The basics:
  - Plaintiff claims that Defendant sent or caused to be sent 12,576 unsolicited commercial emails to Plaintiff (ASIS Internet Services) – an ISP
  - Asks for $1,000/email in statutory damages ($12M+)
  - Asks for additional damages for lost time, bandwidth, expenses, and disk usage

Outline

- Background of the speaker and subject
- Some of the forensic challenges
- Tools and techniques
- Claims and counterclaims
- Damage assessment
- Making and breaking a case
- Questions / Comments?



Forensic challenges to bulk emails

- Challenge 1:
  - There's a lot of it
    - In another recent case, >300,000 emails asserted
  - This means that you can't examine each one
  - And yet any one can contain evidence of
    - Violation or non-violation of statutes
    - Spoliation of that email or the whole set of emails
    - The processes undertaken
    - The identity of senders, recipients, and others involved
    - The mechanisms used to process the evidence
    - Other characteristics of import to the legal matter
  - And any mistake can mean you lose credibility

More challenges

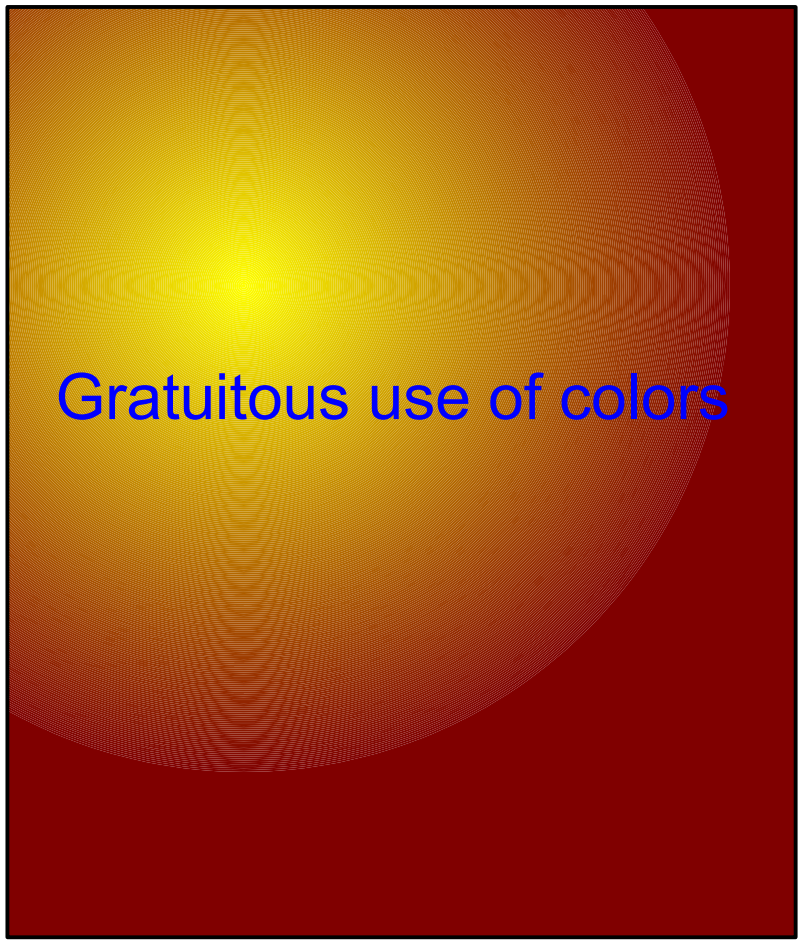
- Challenge 2:
  - Did I mention? It's a legal matter, not a technical one.
  - There are legal limits on what you can see, do, say.
  - There are limits on time, costs, cooperation, etc.
  - The issues in the case dominate the work you do.
  - There are deadlines (days in plenty of cases).
  - The other side is trying to get you (really!)
  - You can't tell a judge and jury about bits and bytes, protocols and RFCs, etc.
    - At least not very easily
    - And not without some clear explanations

More challenges

- Standard tools – our tools – aren't always right:
  - Are not designed to handle large volumes very well
  - Do not handle the legal issues on their own
  - Do not do the sorts of analysis you need to do
  - Do interpretation erroneously when they do it at all
- Neither are the “experts”...
  - In these cases, the sides may not tell the truth, the whole truth, or nothing but the truth
    - They may not even get close to the truth in many cases
    - They obfuscate, withhold information, and try to trick you
    - They may make up evidence, or at least they may not be able to show that they didn't make it up

Outline

- Background of the speaker and subject
- Some of the forensic challenges
- Tools and techniques
- Claims and counterclaims
- Damage assessment
- Making and breaking a case
- Questions / Comments?



A typical tool

- Here's a common tool (command):
  - grep -i "critical content" * | wc
- But the answer might be wrong, depending on the question:
  - How many “emails” contained “critical content”?
  - How many times did “critical content” appear?
    - Neither of these are answered by this command
- As a result, you need to be careful about characterizing your questions
  - Because you only have the tools you have and the tools you make - and they only answer certain questions.

Issues of definition

- What is an “email”?
  - I don't think so... not if you read the law carefully
  - The law has many peculiar definitions
    - You need to create your own definitions to describe the things you normally use in keeping with the legal definitions
- The case study:
  - 12,576 emails asserted
  - The RECORDS provided contained only 1,421 byte sequences that could reasonably be interpreted as “actual emails”

Dates and times

- How to reconcile legal dates/times with headers?
  - You cannot trust email header information unless you can prove it to be true (it is not “self authenticating”)
  - Plaintiff had no demonstrable proof of dates and times
  - Assuming the RECORDS were legitimate, an “anchor” was used to nail down dates and times
    - A Postini server that apparently processed the emails in question and left Received date and time stamps was also used by FCA, which used NTP at the time. An “anchor” email was found that linked FCA (and NTP) time to Postini time to Plaintiff time.
- The case study:
  - Of 1,421 actual emails, 242 met date and time constraints of the legal matter.

The UCE caused server delays!

- That's one of the common claims of damages
  - But is the claim true? How is it measured?
  - Some plaintiff's assert that the emails show the delays, but do they?
  - Emails arriving later
    - Get delivered earlier
  - Some delayed by days
    - Others by seconds
- Correlation is not Causality
  - In this case, they
    - didn't even have correlation

Arrival	Delay
06/27/02 07:33 AM	+0000-00-00 00:00:02
06/27/02 07:53 AM	+0000-00-00 00:00:06
06/27/02 09:11 AM	+0000-00-00 00:00:04
06/27/02 11:55 AM	-0000-00-00 00:00:03
06/27/02 02:41 PM	+0000-00-01 21:24:25
06/27/02 06:23 PM	+0000-00-01 13:06:42
06/27/02 08:12 PM	+0000-00-01 20:16:02
06/27/02 08:24 PM	+0000-00-01 13:09:01
06/27/02 09:12 PM	+0000-00-02 01:12:32

Table 1 - Extracted email arrival and delay times.

How much of an effect?

- How were emails processed?
  - Did the emails cause delays?
  - Or was the process flawed?
  - Let's look at the daily rates
  - Split out by hops
    - From final destination
- Hops in the delivery process
  - All hops are within Plaintiff's infrastructure
  - The “hop” depiction shows how many appeared where in the process on any given day
  - Note that more appear later in the process...
    - Hint: some were duplicated by the process!

Date	Final receipt (1)	Hop 2	Hop 3	Hop 4	Hop 5
10/01/03	4	4	3	2	0
10/02/03	9	9	9	9	0
10/03/03	8	8	8	8	0
10/04/03	6	6	6	6	0
10/05/03	11	11	10	10	0
10/06/03	11	9	8	7	0
10/07/03	23	20	19	18	1
10/08/03	11	11	11	11	0
10/09/03	12	9	6	6	0

Table 2 - Number of emails arriving at different hops by date

Not all emails can/should be sent

- Simple Mail Transfer Protocol (SMTP)
  - HELO all.net
  - 250 OK
  - Mail from: <fc@all.net>
  - 250 OK
  - Rcpt to: <&*to;>
- What is the proper answer?
  - If you said 250 OK – you are wrong!
  - If you let the content be sent – you are "inviting" it
- In the matter at hand
  - 133 actual emails were "invited" - leaving 109 to go

Some collections contain duplicates

- How many copies of the same email can be counted as emails in a case? 1?
  - Examples of how duplicated get made
    - Use of “forwarding” or “mirroring” in server applied to multiple UIDs leads to multiple copies of the same email getting sent to the mirror.
    - Crashes and restarts can create duplicates (back up email queue, send email, system crash, restore from backup, resend email). I have a 6-month later example!
- In the case study
  - 11 of the remaining actual emails were duplicates
  - And the actual email count drops still further
  - Will it reach 0?

Some strange email processes

- A tree of deliveries
  - Shows paths of reception
  - Based on Received headers
  - Note the path and quantities
  - Some emails went from
    - Mail.S.com to
    - Mail.J.com to
    - Mail.S.com to
    - Mail.S.com to
    - X.net
- How did the emails get routed in a circle?
  - 232 of them?

0 325802 X.net
1 325090 mail.S.com
2 325090 mail.S.com
3 215585 mail.J.com
4 232 mail.S.com
4 24 amail.J.com
3 109301 amail.J.com
4 5 mail.S.com ...

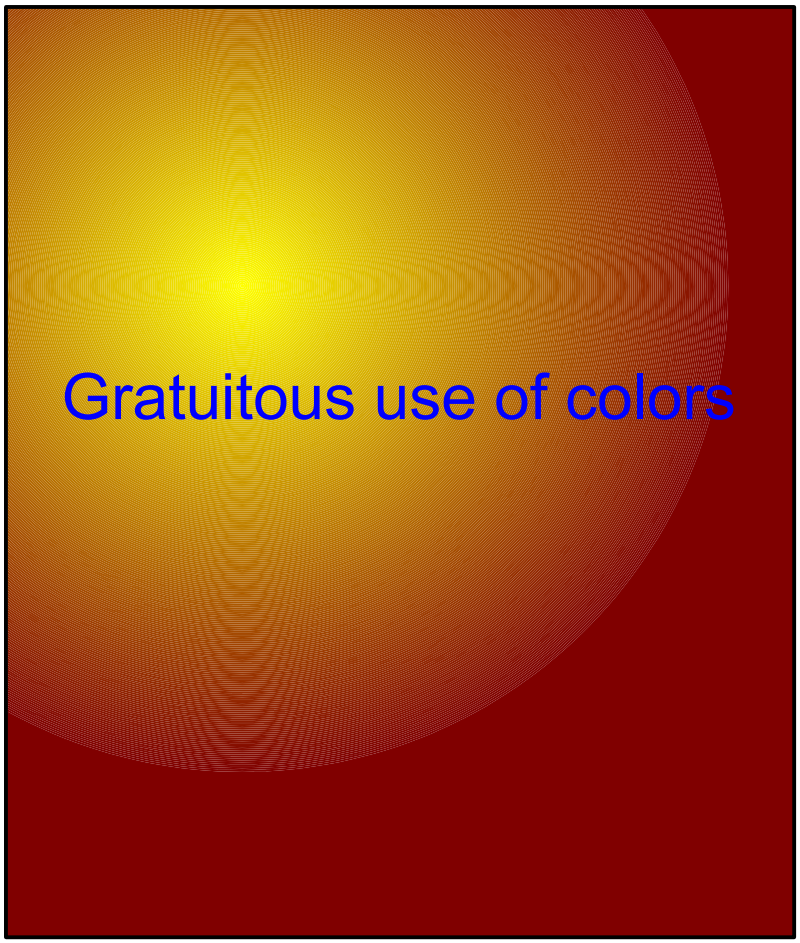
Figure 1 - A tree depiction of an email handling process

Back to the case study

- Out of 12,576 original claimed emails
  - Technical analysis led to 98 “actual emails”
  - The 98 “actual emails” resulted in 175 “emails”
- Statutory damages maximum went
  - From more than US$10 Million
  - To less than US$200,000
- Unfortunately, I don't get a % of loss reduction.
  - Fortunately I do provide my patent pending (not):
- High fees – No guarantees ® Service
- But there's more...

Outline

- Background of the speaker and subject
- Some of the forensic challenges
- Tools and techniques
- Claims and counterclaims
- Damage assessment
- Making and breaking a case
- Questions / Comments?



Some other considerations

- If users requested the emails, they are not "unsolicited"
  - How do we know the users didn't sign up for the emails?
  - All of the potentially affected users must be contacted and provide signed statements!
    - Or some other method may show they were unsolicited.
    - Then they need to be deposed by the other side to determine whether they might have "signed up" and not known that they did so.
    - And their memory can be tested – and – and – and...
- In the case study:
  - NO USERS SIGNED STATEMENTS

Is the "evidence" real?

- Plaintiffs making large-scale UCE claims:
  - Systems administration and programmers
  - Self-proclaimed activists against "spam"
  - Run their own very small client-reduced (less?) ISPs
  - Share information on how to file and pursue legal actions through group Web sites and messages
  - File scores of legal actions against a common list of UCE sources
  - Openly claim their goal to end UCE
  - Violate the privacy of their customers for evidence
- They are thus prime candidates to be vigilantes
  - Means, motive, opportunity, don't retain records well

Was an email fraudulent?

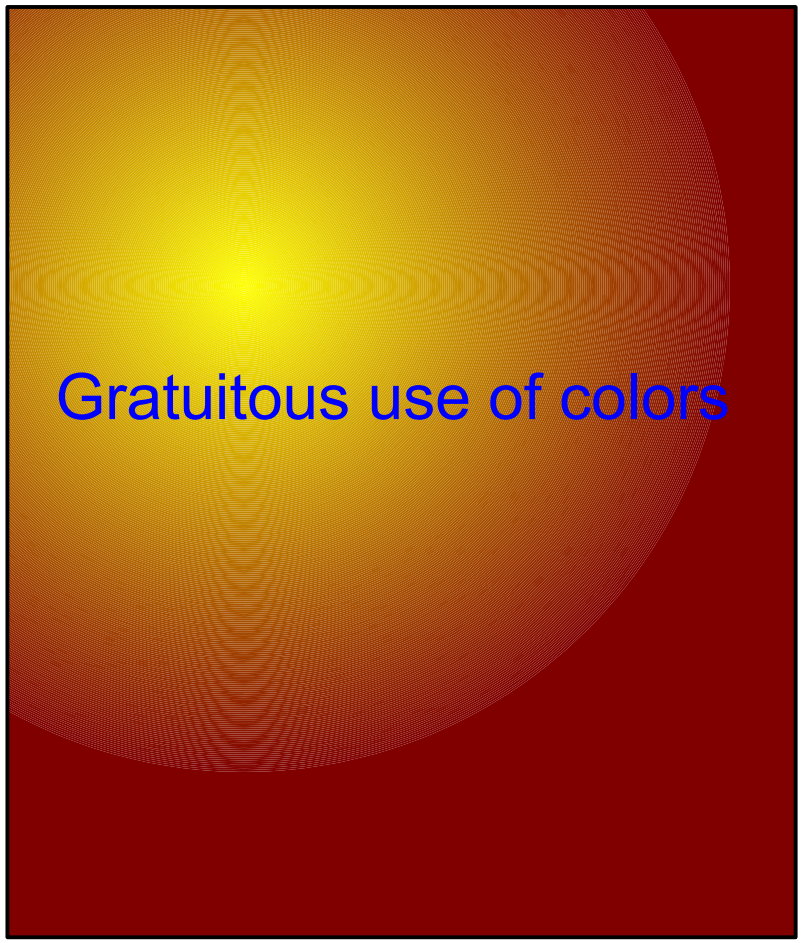
- Using multiple email addresses is “fraudulent”
  - But they themselves do the same thing
- Using multiple domain names is “fraudulent”
  - But they themselves do the same thing
- Using multiple email servers is “fraudulent”
  - But they themselves do the same thing
- Using the word “free” is fraudulent
  - But they don't actually look at the emails before suing
- Removal links didn't work
  - But they don't actually test out the removal links
- The content implies the source
  - But there is lots of evidence to the contrary

Were the claims consistent?

- Claims in bulk email cases come in bulk
  - Typically, they won't describe an actual claim unless forced to do so
- Courts have ruled that they need to provide evidence for each claim
  - So they create spreadsheets using automated tools (or humans) and try to assert which of their many claims are associated with each email
- In tens of thousands of emails, they make plenty of mistakes
  - Inconsistent claims (e.g., the identical Subject is fraudulent and not, etc.) are problematic for them

Outline

- Background of the speaker and subject
- Some of the forensic challenges
- Tools and techniques
- Claims and counterclaims
- Damage assessment
- Making and breaking a case
- Questions / Comments?

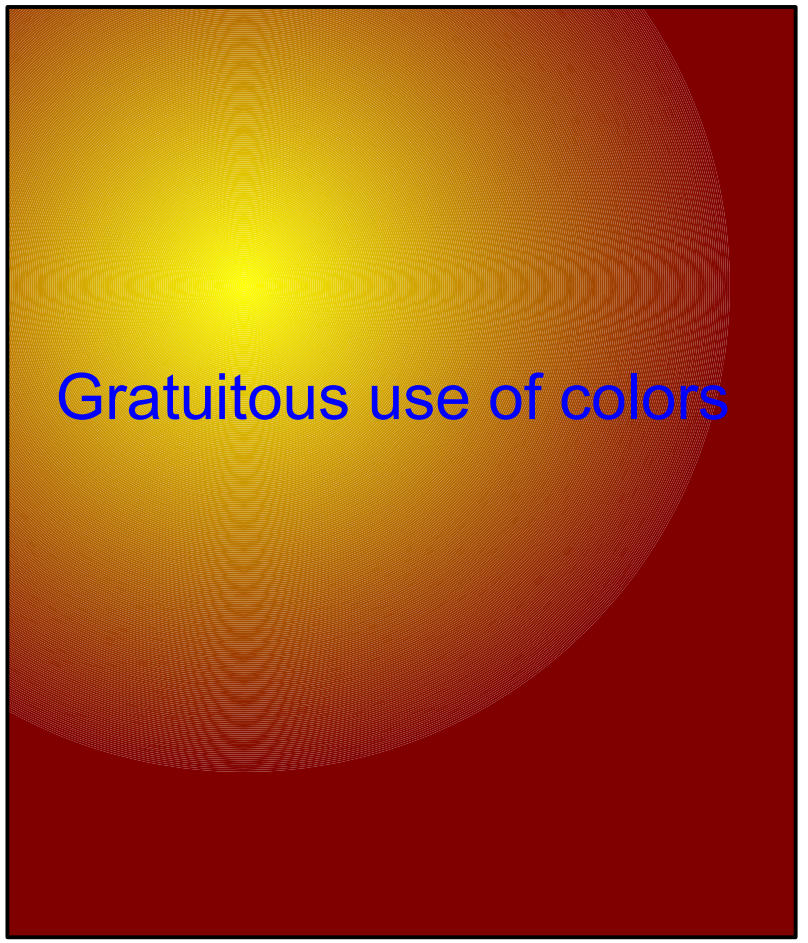


How much did the UCE cost?

- Really?
  - How much can you, as an expert, prove it actually cost in any given case? How do you prove it?
- Some key questions:
  - How much does bandwidth cost?
  - How much does disk space cost?
  - If the “evidence” fits on one CD-ROM, how much bandwidth and disk space was really required?
  - How many UCE emails can fit on a CD?
  - How much processing power does it take to process an email?
- In the example, no actual damages identified

Outline

- Background of the speaker and subject
- Some of the forensic challenges
- Tools and techniques
- Claims and counterclaims
- Damage assessment
- Making and breaking a case
- Questions / Comments?



Reducing the claims to zero

- The vanishing to vacuous emails defense
  - A “divide and conquer” approach
    - Find anomalies and use them to cut down the number of emails that fit the crime
    - Find more anomalies than you need and show that even if only some are accepted as anomalies, the case shrinks close to zero
- The spoliation approach
  - The more anomalies, the closer you are to spoliation. If we can't trust it all, why can we trust any of it?
- The no supporting evidence approach
  - Data retention and disposition policies and legal holds should be in place – or your evidence will go away!

Who actually sent these?

- Proof of who sent or had sent is problematic
  - Timely trace-back to original source
    - IP addresses, domain names, and ownership changes
    - Records are not always accurate or up to date
  - Contemporaneous records that stand up in court
    - Just because they are “normal business records” doesn't mean that they prove what they are asserted to prove
  - If they went through a “proxy” are they “forgeries”?
    - Assertions of “false, misleading, fraudulent, etc.” must be proven with real evidence and knowledge of how the Internet and systems on it actually operate

Is the evidence real/good?

- In case after case, plaintiffs have provided poor evidence that has clear indications of spoliation, fraud, and fabrication
  - Evidence should include all relevant records
  - A retention and disposition policy should exist and be followed
  - Legal holds should be placed on all relevant records as soon as the possibility of a legal action is known
  - Contemporaneous lookups and searches should be done or the value of them is reduced
  - Records from external sources should be secured to prove with independent sources what actually took place

Do they have an expert?

- All of these technical things with DFE require an expert to introduce and present the evidence, analysis, and interpretation
  - Knowledge, skills, training, and education required
  - A basis for claims must be made
  - They will be contested strongly – show your work
- In the example case,
  - The expert refused to sign their final report after being deposed, and would not take part, so they had no expert and could not argue against the opposing expert
  - As a result, many undisputed facts existed

Most such cases

- The results run one of three ways
  - Plaintiff wins a huge amount that can not be collected
    - Defendant never contested and left town
    - Likely a criminal who was committing frauds
  - Settlement for undisclosed amount ($10K - $200K)
    - Plaintiff sues again for another set of emails or passes the word to the rest of the group who also sue
  - *Defendant defends case and wins (example case)
    - Defense costs $250K-$800K (expert and defense lawyer make the money)
    - Plaintiff spends little money and sues again and again
    - Eventually, if too extreme, Defendant countersues...

Making the case

- In making such a case
  - A lot of time and effort are required
  - A real expert is required
  - Subpoenas and examinations are required
  - Depositions are required
- An example where the case was made
  - 2009-08 (settled)
  - Complainant sought permanent restraining order
  - Respondent claimed emails were not sent by them
  - Experts were needed to make/break the case

How to make the case

- Step 1: Get real evidence
  - From independent 3rd parties where possible
  - Via subpoena and independent ISP pulls
  - With process recorded / validated / attested to
- Method:
  - Unix-based periodic screen capture
  - Running MD5 checksum of last screen capture
  - Step by step retrieve each email from server
  - Verify downloads properly stored on CDs
  - ISP has no reason to forge records
  - Records are under ISP sole control (independent)

Make the case

- Step 2: Step-by-step header trace back
  - Try for records from each (subpoenas)
  - Path and content similarity analysis useful
  - Look for inconsistencies to detect flaws
  - Get to limited number of real sources
- Method:
  - Automated analysis of headers to get details
  - Manual review of results and grouping
  - Network lookups and records requests
- Note: most of the details will not be available!

Make the case

- Step 3: Ask the right questions
  - Depositions and records requests to get lots of small details that fill the gaps
    - What process did you use?
    - What company do you get ISP services from?
    - What are all your user IDs?
    - What are all your phone numbers?
    - Where were you on or about (date)?
    - Where do you work?
    - Have you ever used Kinkos computer services?
    - Do you use computers from Starbucks/libraries ...
- It's hard to lie consistently
  - Lots of little facts add up to in/consistencies

Make the case

- Step 4: Put it all together
  - These emails were retrieved from X by me
  - They headers recorded there as part of normal business records relied upon for day to day use indicate X got them from Y
  - The records at Y indicate that the account used was of user Z
  - User Z is an account Respondent admitted to
  - The terms of service (EULA) says people are responsible for use of their accounts
  - The IP address is listed as from the ISP Respondent uses in their home city

Yeah but...

- Step 5: Their defense
- It wasn't me...
  - Someone broke into my computer
  - The emails were forgeries
  - I didn't send them
  - The graphical images are not from my phone
- Any of these might be true...
  - But once you get over the threshold of the standard of proof, the burden shifts to the other side
  - They now have to prove and you have to refute

Yeah but...

- Someone broke into my computer
  - Give us a forensic image and show the presence of the break-in details to prove you claims – or fail to do so
  - Where is your evidence of the claim?
  - If they did, show that they also did the other things involved...
- The emails were forgeries
  - But the analysis shows that there were no detected inconsistencies.
  - What inconsistencies did you detect and what is the basis to claim they are forgeries?

Yeah but...

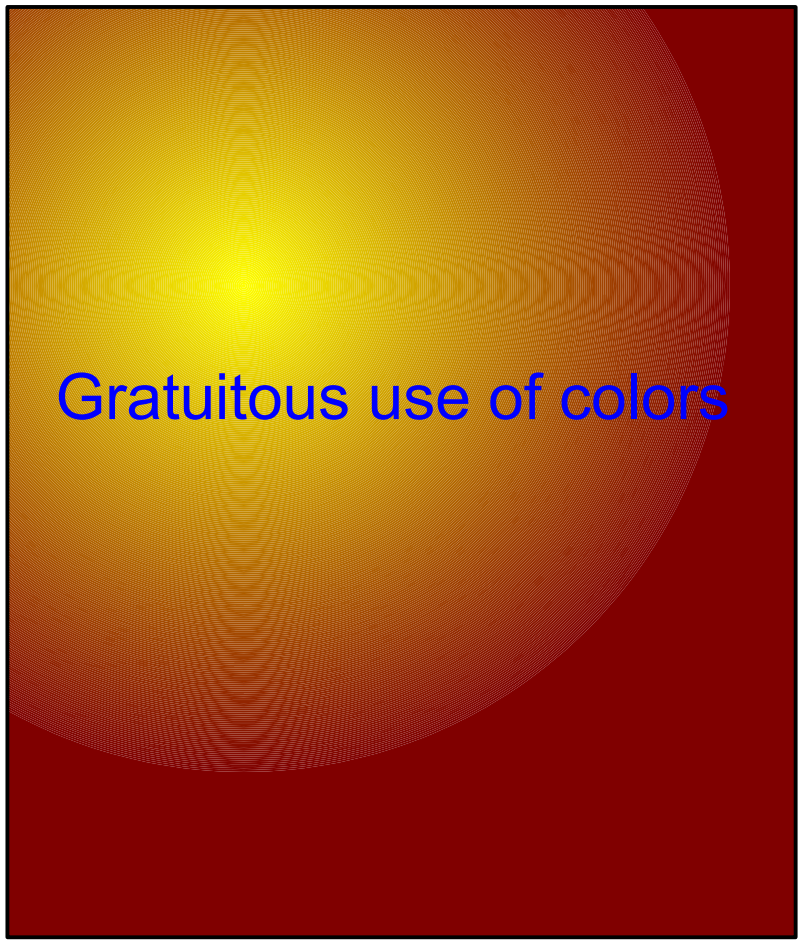
- I didn't send them
  - They were sent from your account and the EULA says you are responsible.
  - Can you prove you were not there when sent?
  - Can you prove you did not send them?
- The graphical images are not from my phone
  - Prove that they were not...
  - Suppose they weren't – so what?
  - Did you have any conversion software?
- The burden is now on them...

One hour of CPE

- This class provides 1 hour of CPE credits
- To put it in context:
  - A M.S. degree from CalSci focussed on digital forensics involves about 160 contact hours and a total of about 500 total hours of effort over a minimum of 1 year.
  - A Ph.D. degree from CalSci focussed on digital forensics involves about 500 contact hours and a total of almost 2,000 hours of total effort over a minimum of 3 years.
- Working on increasing volumes requires special skills and technology. Where will you get them?

Outline

- Background of the speaker and subject
- Some of the forensic challenges
- Tools and techniques
- Claims and counterclaims
- Damage assessment
- Making and breaking a case
- Questions / Comments?



Thank You

Questions?
Discussion?!

Dr.Cohen at Mac.Com

http://all.net/