Modeling and Analysis of Computer Security Incident Data

Edward Condon (RE, IREAP), Angela He (MATH), Michel Cukier (RE, ISR,CS)

The Institute for Systems Research

Center for Risk and Reliability

Motivations

- Computer and network related security incidents have increasing financial consequences involving:
  - Repair and recovery of resources
  - Direct losses from theft of personal and/or proprietary information
  - Reputational damage (negative impact on stock prices) or reduced consumer confidence (loss of business)
- Modeling security incident data can be a useful tool for:
  - Gaining insights into underlying factors and dynamics affecting how incidents evolve over time
  - Provides a baseline for future comparison
    - Did a new intrusion prevention system (IPS) make a difference?
    - Did a change in polices have the expected impact?

Why Examine Software Reliability Growth and Time Series Models?

- Both types models are well established in require few parameters
- Software Reliability Growth (SRG) models—view computer security incidents on a network as analogous to software failures in a system
  - Defects in software systems detected and repaired -VS-
  - Machines involved in an incident removed and "cleaned"
    - Reinstalled, Patched, reconfigured, etc.
  - Examined G-O, S-shaped, K-stage, and Duane SRG models
- Time Series (TS) models attempt to account for random fluctuations over time using historical data
  - Explored auto-regressive moving average (ARIMA) models

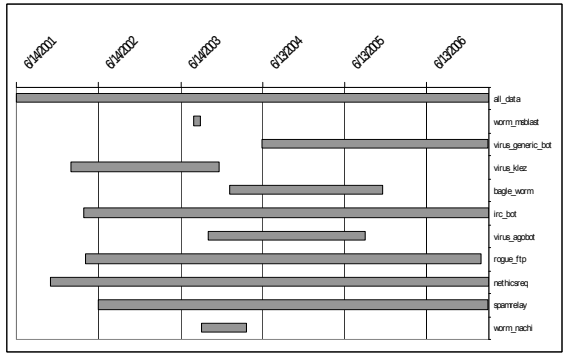
Defining an Incident

- Often a distinction is made between an "event" and an "incident"
  - An "event" is something that attracts notice or appears abnormal
  - An "incident" is an event that is verified and attributable to a security failure
- Three sources of events
  - An intrusion detections system (IDS)
  - Reports from system administrator (both internal and external)
  - Reports from users
- Determination of incident by UMD network security staff
  - Port-scans, packet and traffic analysis, on-site verification
  - After verification, network access typically blocked until machine cleaned

Summary of Incident Data

Incidents	# of Incidents	Start Date	End Date	# of Days	# per Day
All data	11966	6/14/2001	3/14/2007	2100	5.7
worm_msblast	2133	8/12/2003	9/12/2003	32	66.7
virus_generic_bot	1940	6/8/2004	3/12/2007	1008	1.9
virus_klez	1118	2/12/2002	12/2/2003	659	1.7
bagle_worm	849	1/20/2004	11/28/2005	679	1.3
irc_bot	690	4/8/2002	3/14/2007	1802	0.4
virus_agobot	589	10/13/2003	9/13/2005	702	0.8
nethicsreq	509	11/12/2001	3/14/2007	1949	0.3
rogue_ftp	500	4/19/2002	2/7/2007	1756	0.3
spamrelay	429	6/11/2002	3/9/2007	1733	0.2
worm_nachi	317	9/16/2003	4/1/2004	199	1.6

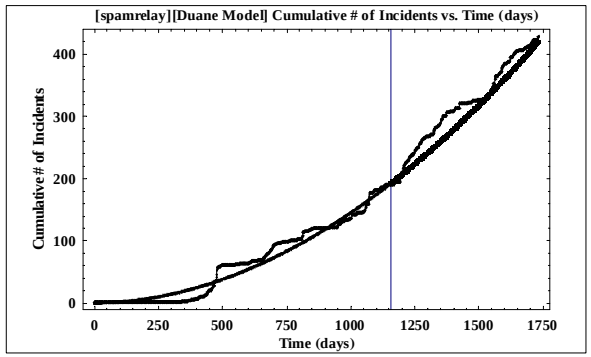
Timeline of Incident Data (top 10)



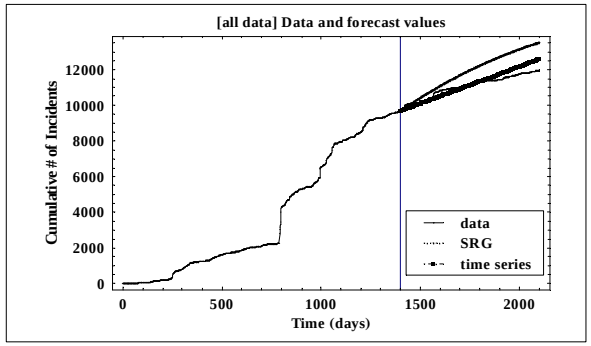
Incident Data

- Provided by the Office of Information Technology (OIT) at the University of Maryland
- Consists of 11,966 verified security incidents
- Recorded over a period of 5.75 years (from June 14, 2001 to March 14, 2007)
- The number of incidents detected in a single day varies from 0 to 580
- An average of 5.7 incidents per day
- The incidents were already categorized into 51 different types

Model and Forecast for "spamrelay"



SRGM and TS Forecasts for "all" data



Conclusions

- Applying the models provides insights about the different incident types and the overall number of incidents
- Some incident types are more distinct and well-defined
  - Two examples:
    - Incidents resulting from a unique cause (such as "worm_msblast")
    - Incidents with easily indentifable behavior (such as "spamrelay")
  - These types are easier to identify, patch or prevent
  - SRG models tend to be better suited for these incident types
- The Overall total number of incidents results from of a combination of unrelated factors
  - Not all incident types attributable to a known cause or exhibit well-defined behavior
  - TS models are better able to describe this evolution