Adapting Econometric Models, Technical Analysis and Correlation Data to Computer Security Data

Spyros K. Kollias,
Vasileios Vlachos,
Alexandros Papanikolaou and
Vassilis Assimakopoulos

Financial Forecasting

- Mainly based on processing past data
- A temporal process
  - Time series analysis is suitable!
  - e.g. ARMA, ARIMA, GARCH, ACD models
- Has successfully been applied to other fields in order to forecast, e.g.:
  - Wheat area and production (Pakistan)
  - Traffic flow
  - Next-day price of electricity
  - Air pollution levels

Exploiting Publicly Available Computer Security Data

- WildList (malware lists)
- DShield (various attacks)
- Social networks
  - Security-related accounts (e.g. AV vendors)
  - Users' posts
- Underground world
  - Price for stolen credit card numbers
  - Price for e-mail addresses for sending spam
    - More expensive => Harder to get => Better security
    - Real-life financial transactions!

Exploiting Publicly Available Computer Security Data (cont.)

- Will the publicly-available data suffice?
  - If not, use posts from social networks, search engines' query data, etc.
  - The developed tool will complement existing systems
- Creating the Analogy
  - Treat all threats as a stock index, e.g.:
    - Observed attacks over time ~ Price
    - Volume information will be lost
  - Categorise threats according to their nature
    - Better approach as it produces/preserves extra info
    - Sum of attacks ~ Volume
    - Severity of attacks ~ Price
- Depending on the amount of available information, the use of all applicable models may be impeded.
- Flexibility and ability to adapt new parameters

Will it Work ?

- Efficient Market Hypothesis (EMH)
  - The time series reflect info about their object
    - Weak: Only past, publicly-available data.
    - Semi-strong: Up-to-now info and the price changes accordingly.
    - Strong: All information – public and “insider” info
- Security sector time series follow the strong EMH.
  - No “insider” information
  - All threats are real and have already occurred

Example Scenario

- Process available virus info from WildList
- Create analogy
  - Top/Mid/Bottom-Cap(ital) Company ~ Top/Mid/Bottom-Durable Virus
  - Durability of virus:
    - How many lists has it appeared in? How long for?
- Top-durable virus
  - Known to most security-related applications
- Freshly-introduced virus
  - May cause significant damage in early stages
  - May be easy to confront if not very “intelligent”