I·C·S
The Institute for Cyber Security

C·SPECC
Center for Security and Privacy
Enhanced Cloud Computing

NSF

Characterize and Quantify Cyber Attack Pattern by Granger Causality

Van Trieu

Richard Garcia, Shouhuai Xu, Yusheng Feng

University of Texas at San Antonio

UTSA
Computer Science

Motivation and Goal

Motivation:

Cybersecurity Statistics[1]

- 3 billion Yahoo accounts were hacked (2016)
- Damage related to cybercrime is estimated to hit $6 trillion (2021)

1. IDENTIFY
2. PROTECT
3. DETECT
4. REACT
Cyber-Security

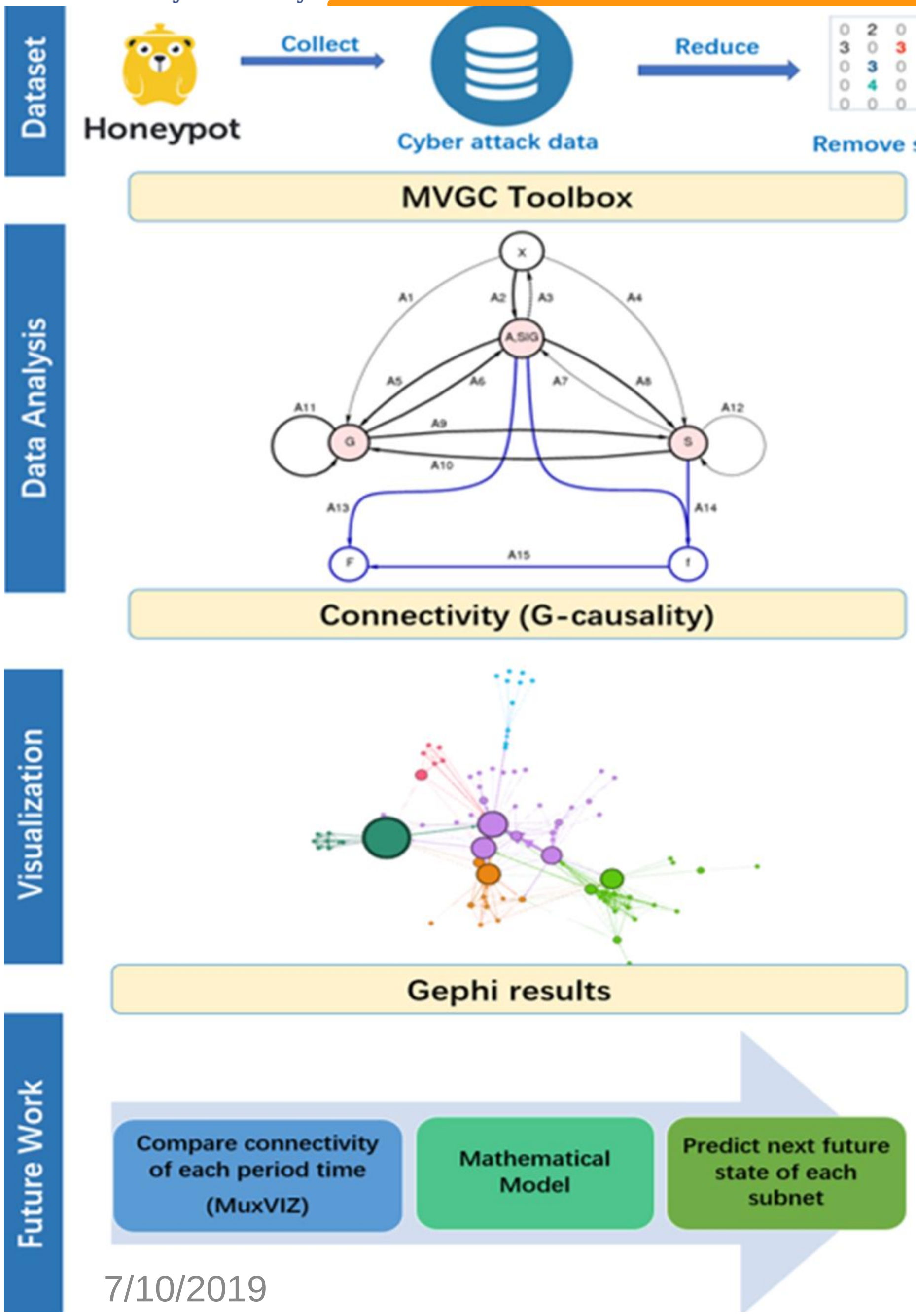
Enhance the performance of alert correlation and minimize damage from attacks is necessary.

Goal:

Study the phenomenon of the time series data in IPv4 address space utilization based on Granger Causality.

Apply the learned phenomenon into Long Short Term Memory model to improve prediction in dynamic time series data.

Framework



0	2	0	0
3	0	3	0
0	3	0	0
0	4	0	0
0	0	0	5

0	2	0	0
0	0	4	0
0	1	0	0
0	3	0	0
0	0	0	6

Data: /8 subnet
/16 subnet

X Granger Cause Y
Linear Regression
Sum Square Error

- *Dependent variable and lagged values*
- *Independent variables and lagged values*

Hypothesis F-test

- *Network of subnets*
- *Communities of subnets (Clusters)*

Characterize and identify the future state

Granger connectivity between subnets

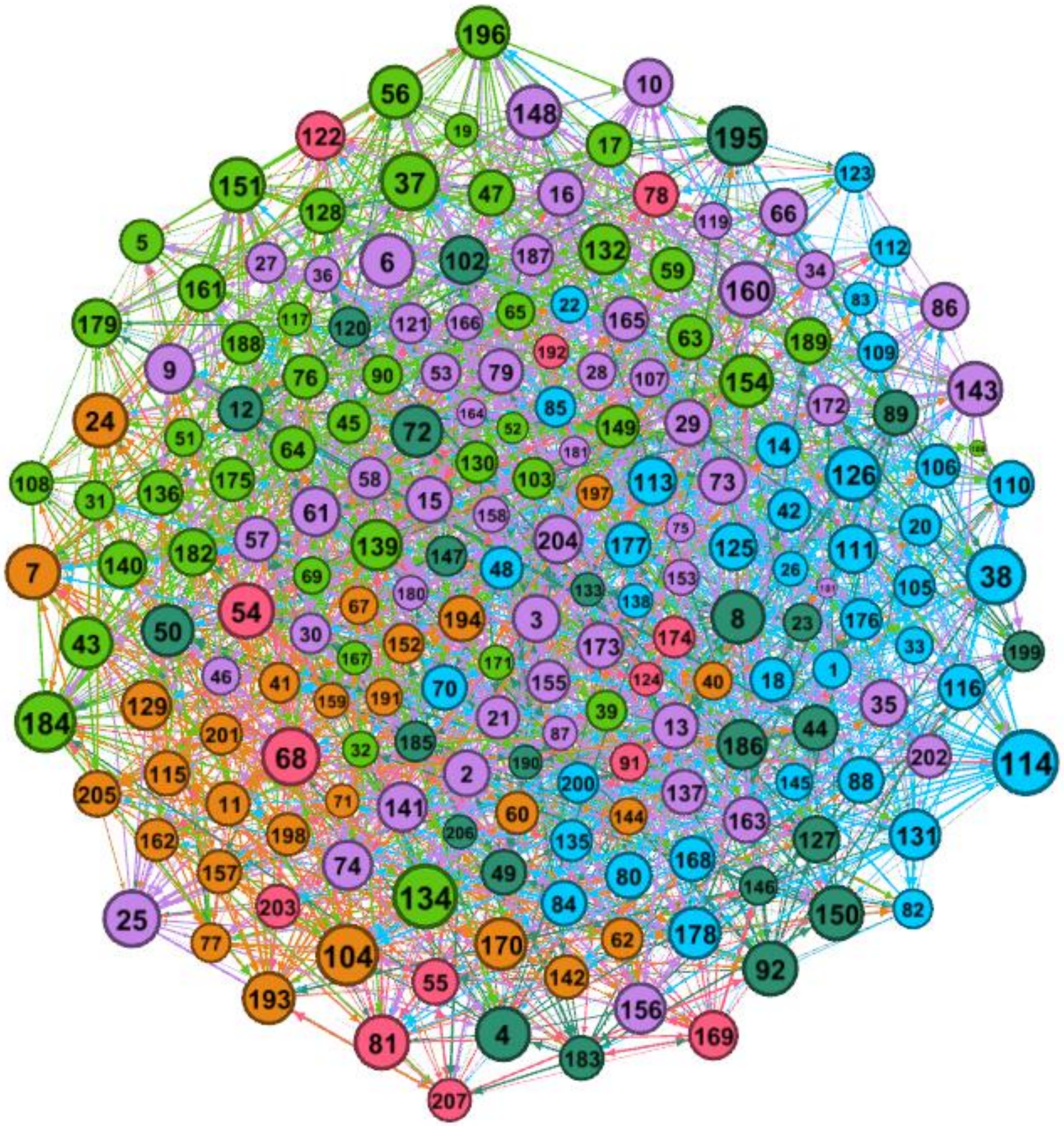
Network of /8 subnet (96 days)

Granger Causal results:

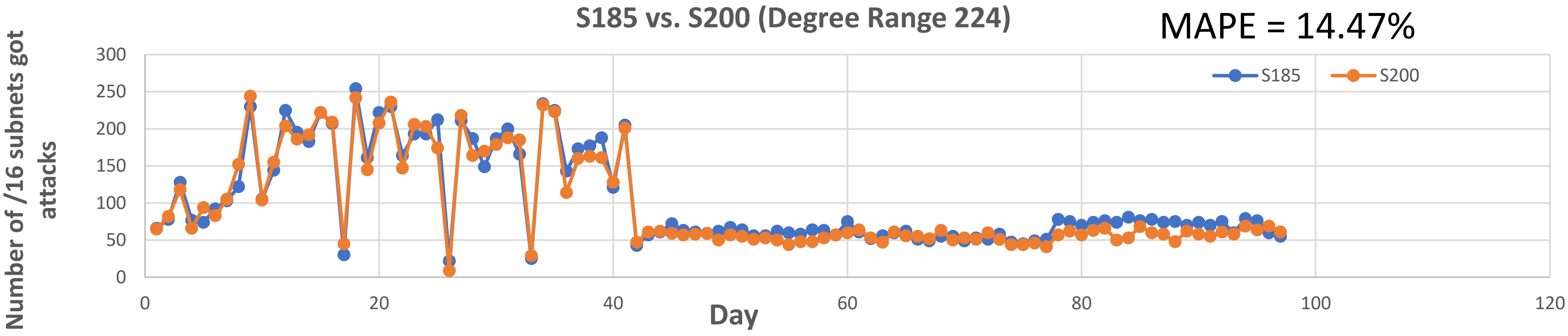
- 199/207 nodes
- 2592 edges

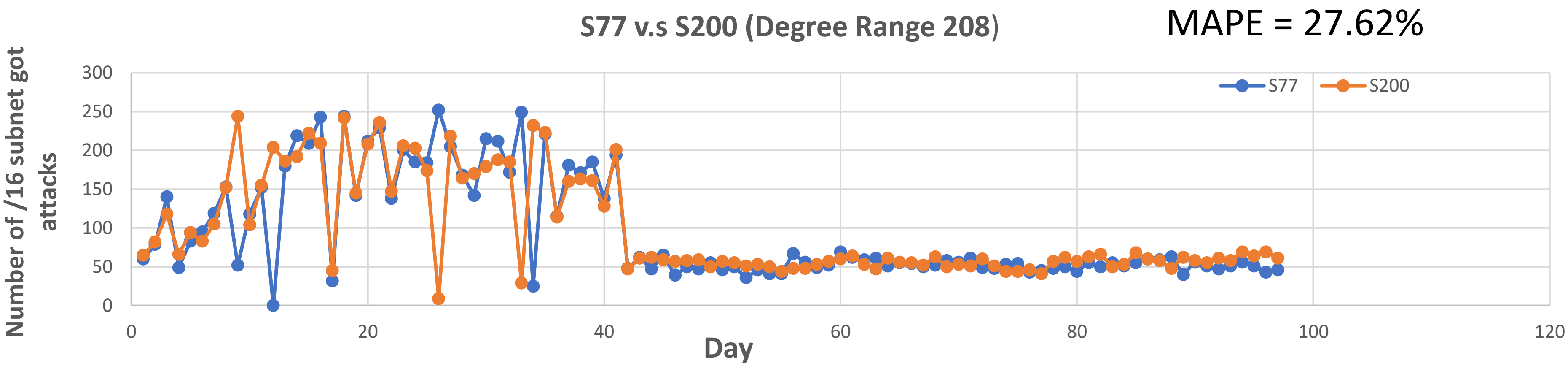
Seven different communities:

🟪	3	(20.1%)
🟩	5	(19.6%)
🟦	1	(17.59%)
🟧	2	(15.08%)
🟫	4	(14.57%)
🟥	6	(8.04%)
🟩	0	(5.03%)



Network Phenomenon

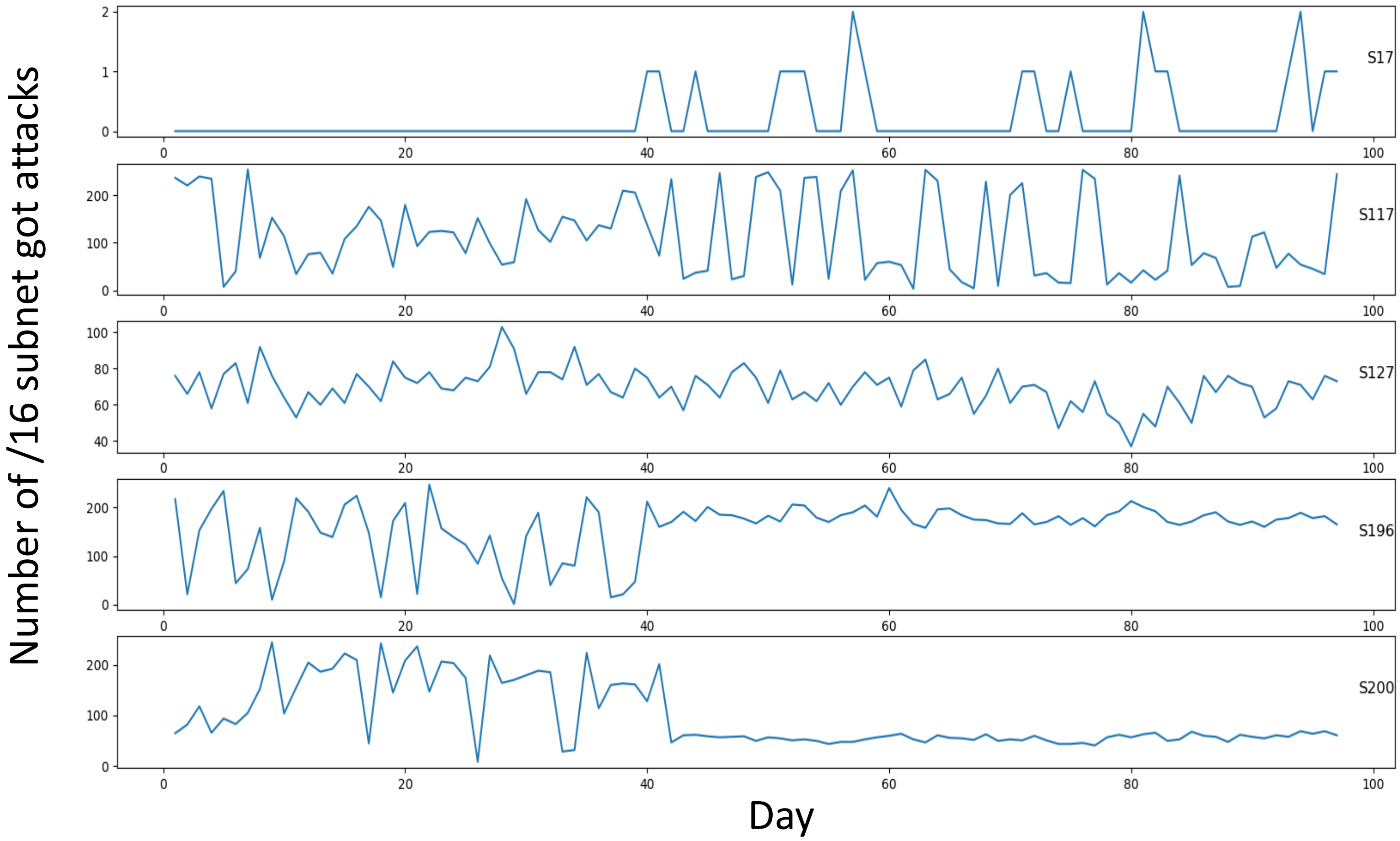




<u>**Insight 1**</u>: The higher the number of degree range in nodes, the similar in pattern of their time series data.

Network Phenomenon

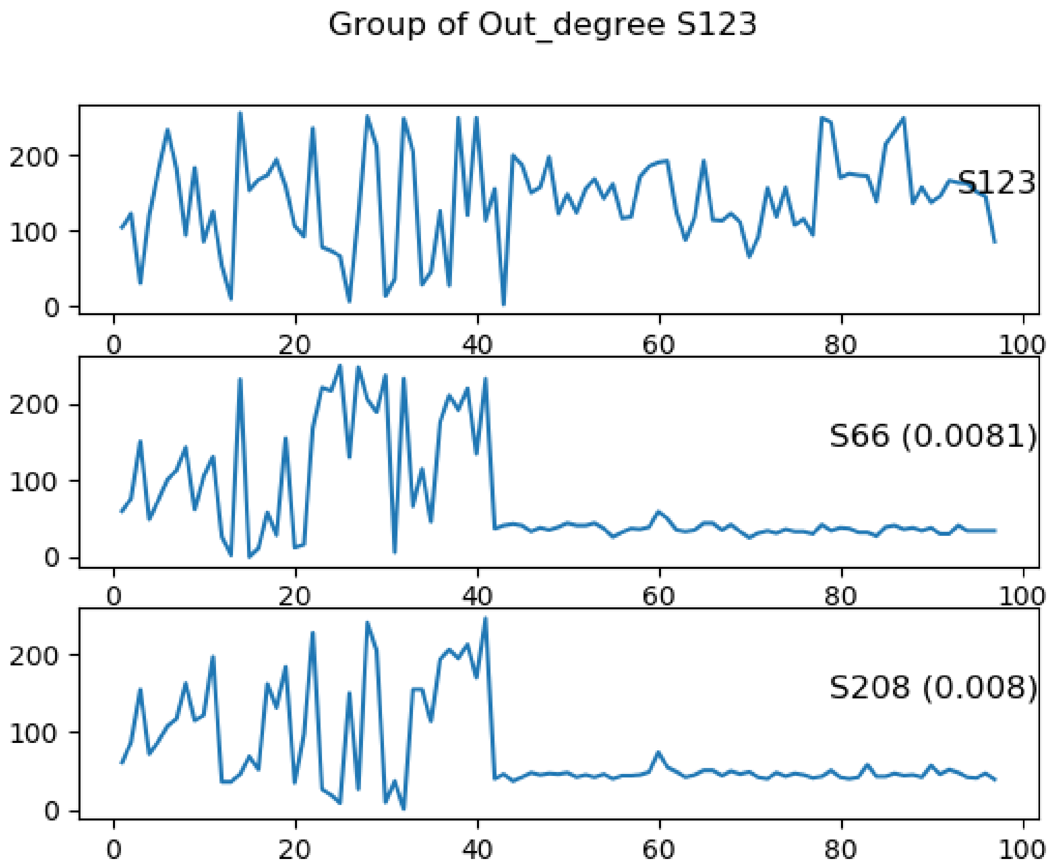
Five Standard Patterns

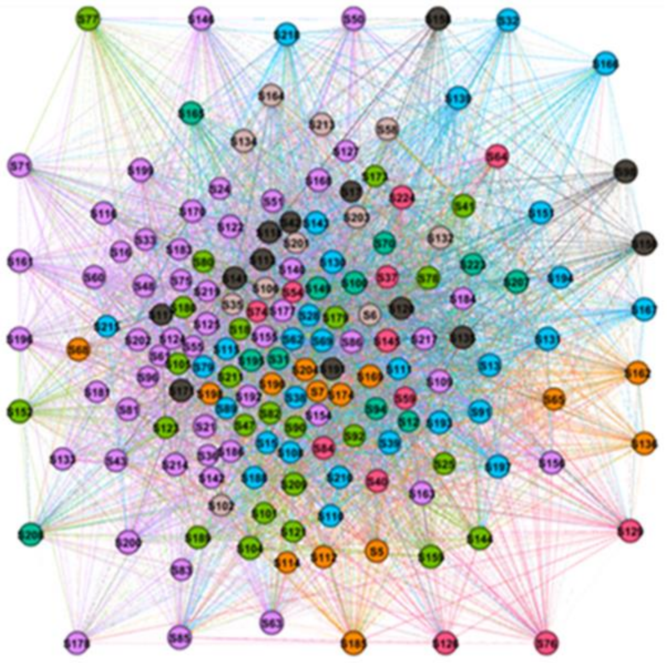


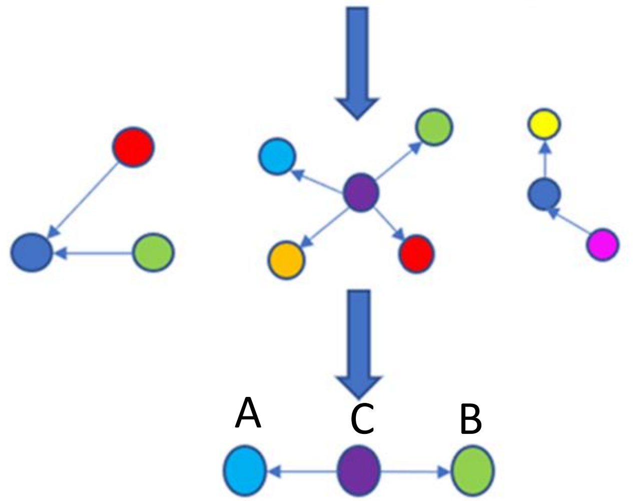
Percent	Degree Range
6.12	0 - 25
8.36	26 - 80
16.31	81-150
29.47	151-180
39.74	181-224

Insight 2: The nodes with highest degree range have the most popular pattern in the data set

Network Phenomenon

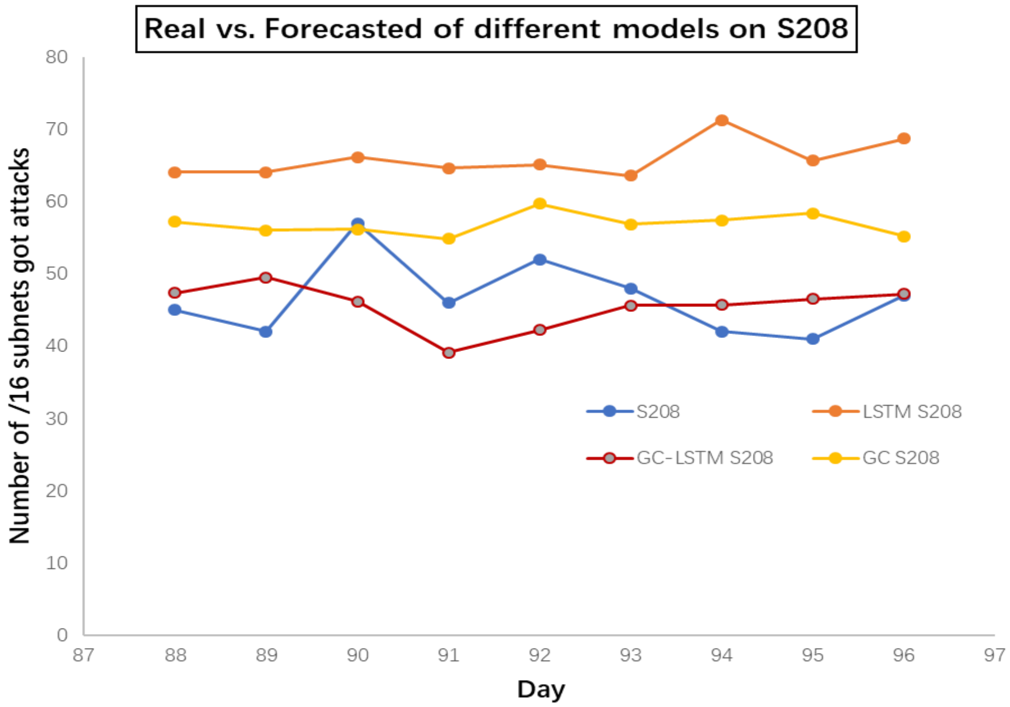


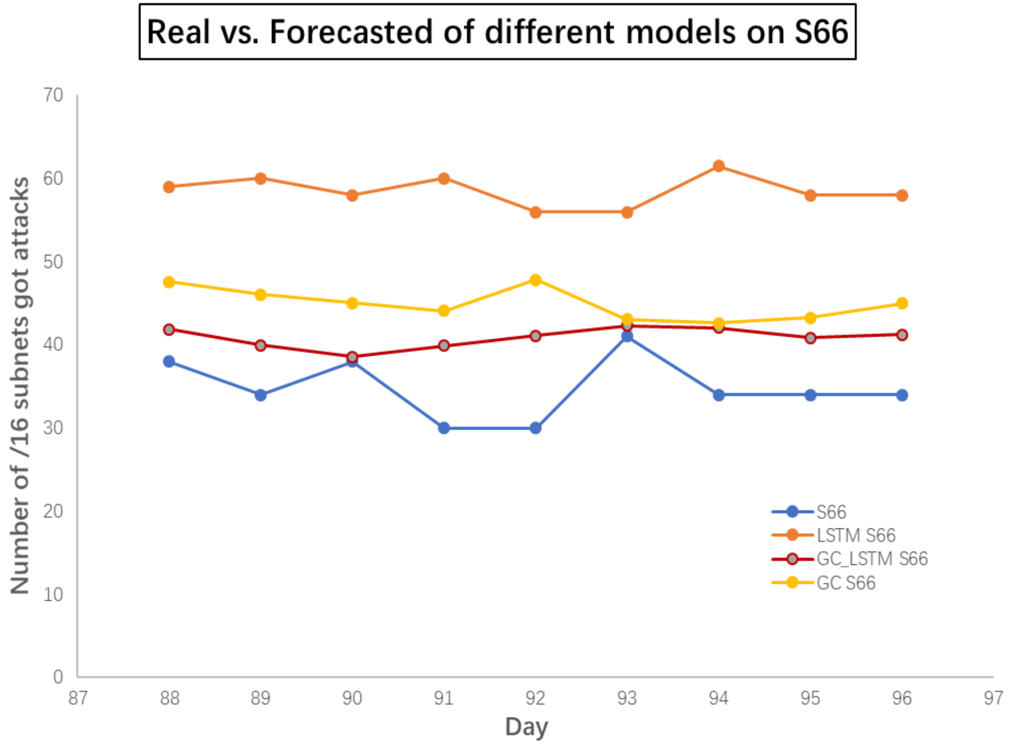




Insight 3: VAR of nodes (VAR between nodes), which have same source and similar p value, can give a better prediction compare to the original VAR (VAR between source and node)

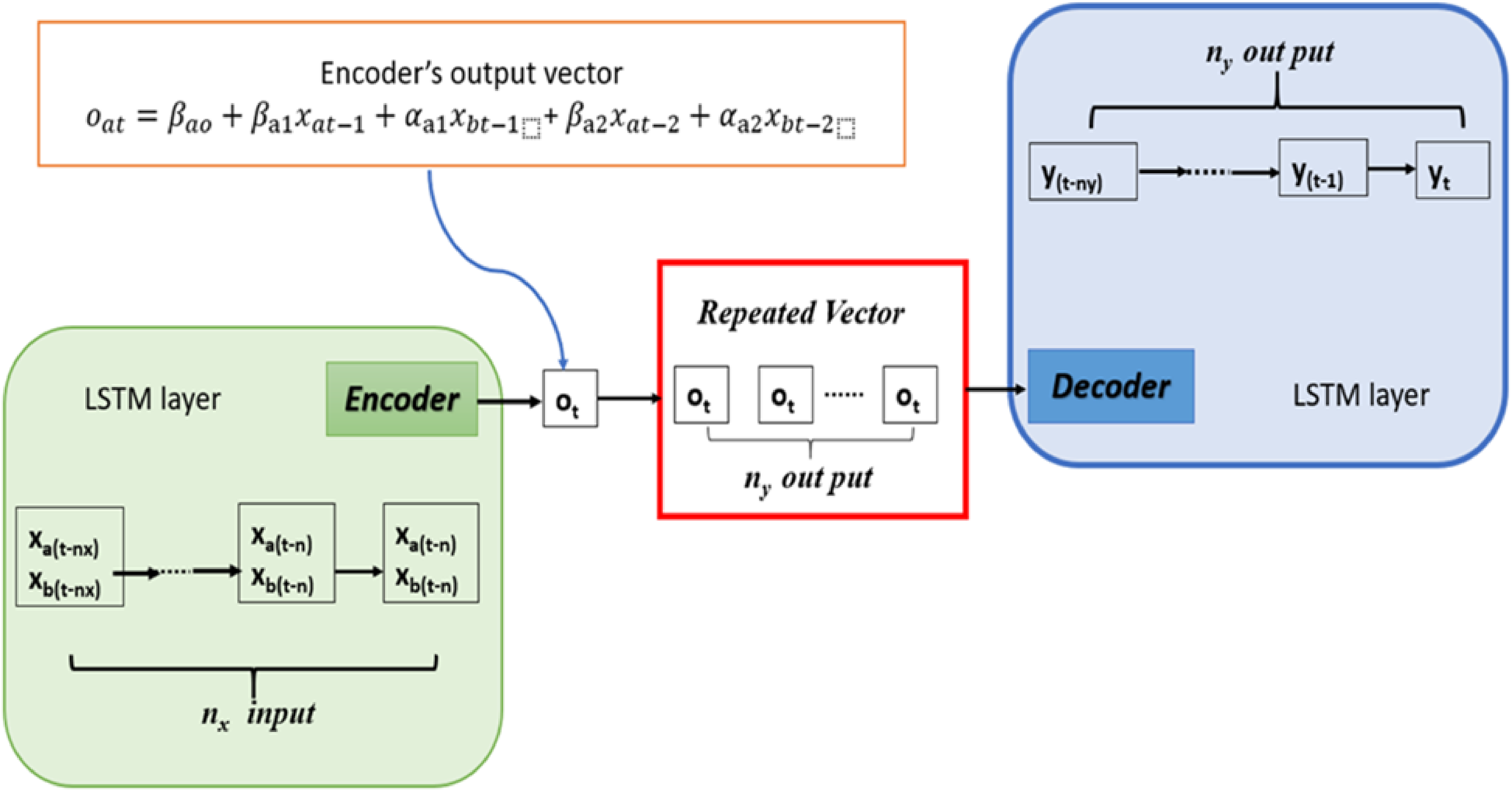
Statistical Comparison


Real vs. Forecasted of different models on S208


Real vs. Forecasted of different models on S66

	MAPE		RMSE	
	GC(S123)	GC	GC(S123)	GC
S66	69.88831	38.54982	43.44967	26.960052
S208	52.74795	31.48487	40.06773	19.413949

LSTM (Encoder-Decoder)



Encoder's output vector

$$o_{at} = \beta_{ao} + \beta_{a1}x_{at-1} + \alpha_{a1}x_{bt-1} + \beta_{a2}x_{at-2} + \alpha_{a2}x_{bt-2}$$

Encoder-Decoder LSTM summary model