Cybersecurity Dynamics:

A Foundation for the Science of Cybersecurity

Shouhuai Xu,PhD

Director, Laboratory for Cybersecurity Dynamics

Professor, Department of Computer Science

University of Texas at San Antonio

www.cs.utsa.edu/~shxu

@USTC

Acknowledgement

Mentors: Moti Yung, Gene Tsudik, Ravi Sandhu, Elisa Bertino, Mike Reiter

Collaborators: Yuzhong Chen, Jin-Hee Cho, Gaofeng Da, Pang Du, Yujuan Han, Taizhong Hu, Lei Hua, Zi-Gang Huang, Alex Kott, Ying-Cheng Lai, Tao Li, Xiaohu Li, Wenlian Lu, Yilun Shang, Jie Sun, Hao Wang, Maochao Xu, Ren Zheng, Deqing Zou, et al.

Students: Paul Parker, Li Xu, Zhenxin Zhan, Qingji Zheng, Moustafa Saleh, Marcus Pendleton, Richard Lebron-Garcia, Jose Mireles, Eric Ficke, Zhen Li, Jian Shi, Haoyu Chen, Zhi Li, Huashan Chen, Zheyuan Sun, Deqiang Li, Zong-Zong Lin, et al.

Openings:
Post-Doc and PhD Students

Openings for Post-Doc researchers (2-3 positions):

- Applied mathematics (e.g., Dynamical Systems, Control Theory, Game Theory, Uncertainty Quantification, Data Assimilation)
- Data analytics: Statistics (time series analysis), (Adversarial) Machine Learning (Deep Learning), etc.
- IoT security, IoM(edical)T security, Blockchains

Openings for PhD students (3-4 positions):

- Machine Learning (Deep Learing); IoT security; IoMT security; Blockchains; Systems/software security; metrics etc.

Top-down vs. Bottom-up Way of Thinking in Cybersecurity

- The often-adopted bottom-up way of thinking
  - Reductionism: Building-blocks → composition
- This talk: A top-down way of thinking
  - Treating cyberspace as complex systems
- Both are important!

Terminology

Throughout this talk:

Security = cybersecurity = cyber security

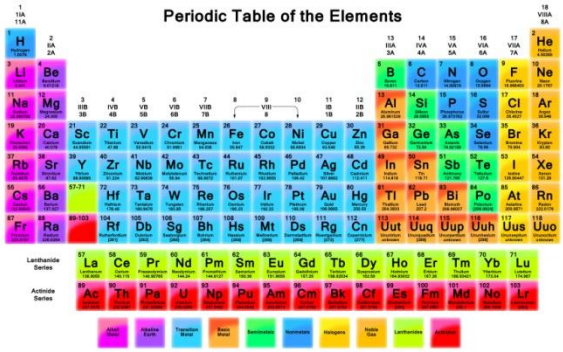
Cybersecurity Status Quo

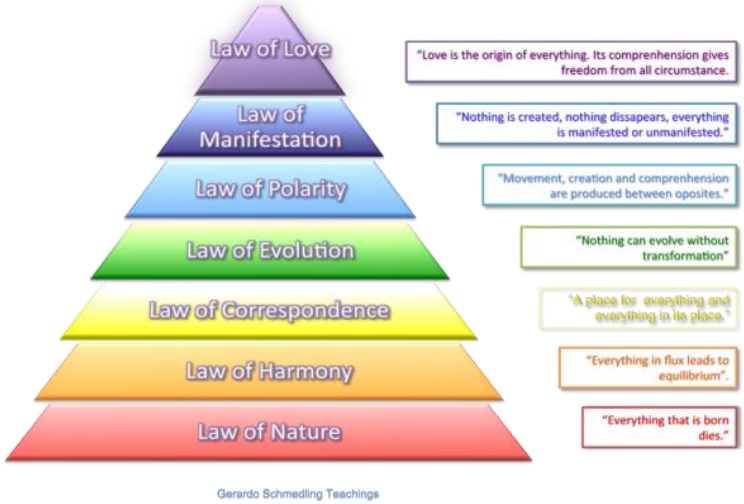
Many questions cannot be addressed by the existing body of knowledge, such as:

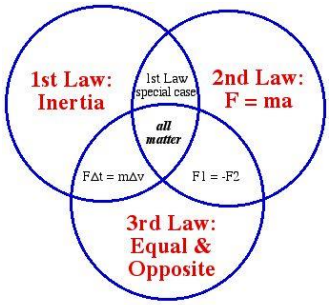
1. How should we quantify the power/capabilities of cyber attack/defense mechanisms?
2. What is the current cybersecurity situation in the cyber system I'm defending for?
3. How should I orchestrate my defense actions to optimize my enterprise cybersecurity (via collective use of Adaptive, Proactive, and/or Active defenses)?

Cybersecurity Status Quo (cont.):

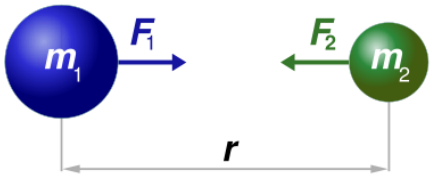
What are the Cybersecurity counterparts of these (or proving the lack of)?







Newton's Laws



$$F_1 = F_2 = G\frac{m_1 \times m_2}{r^2}$$

Why Cannot We Answer Them Yet?

Our understanding of cybersecurity is

- an Art (or heuristic), rather than a Science

Using other disciplines as analogies:

- Prior to Shannon's definition of "secure encryption"
- Prior to the invention of Telescope in Physics

A Brief History of the Term of "Science of (Cyber)Security"

- 2003: Adi Shamir's Turing Award Talk
  - Crypto is a Science, Security is a mess
  - Still true as of 2018
- 2008: the term of "Science of Security" emerged in the United States.
  - Workshops etc
- 2014: Symposium on Science of Security (HotSoS)
  - Kind of geared towards the Lablet projects though
- 2018: International Conference on Science of Cyber Security (SciSec), www.sci-cs.net
  - Mission: How can we elevate the Art of Cybersecurity to the Science of Cybersecurity, like how the Art of Crypto was elevated to the Science of Crypto?

The Term Is Wonderful!

But, how can we get things started,

conceptually

and

technically?

Talk Outline

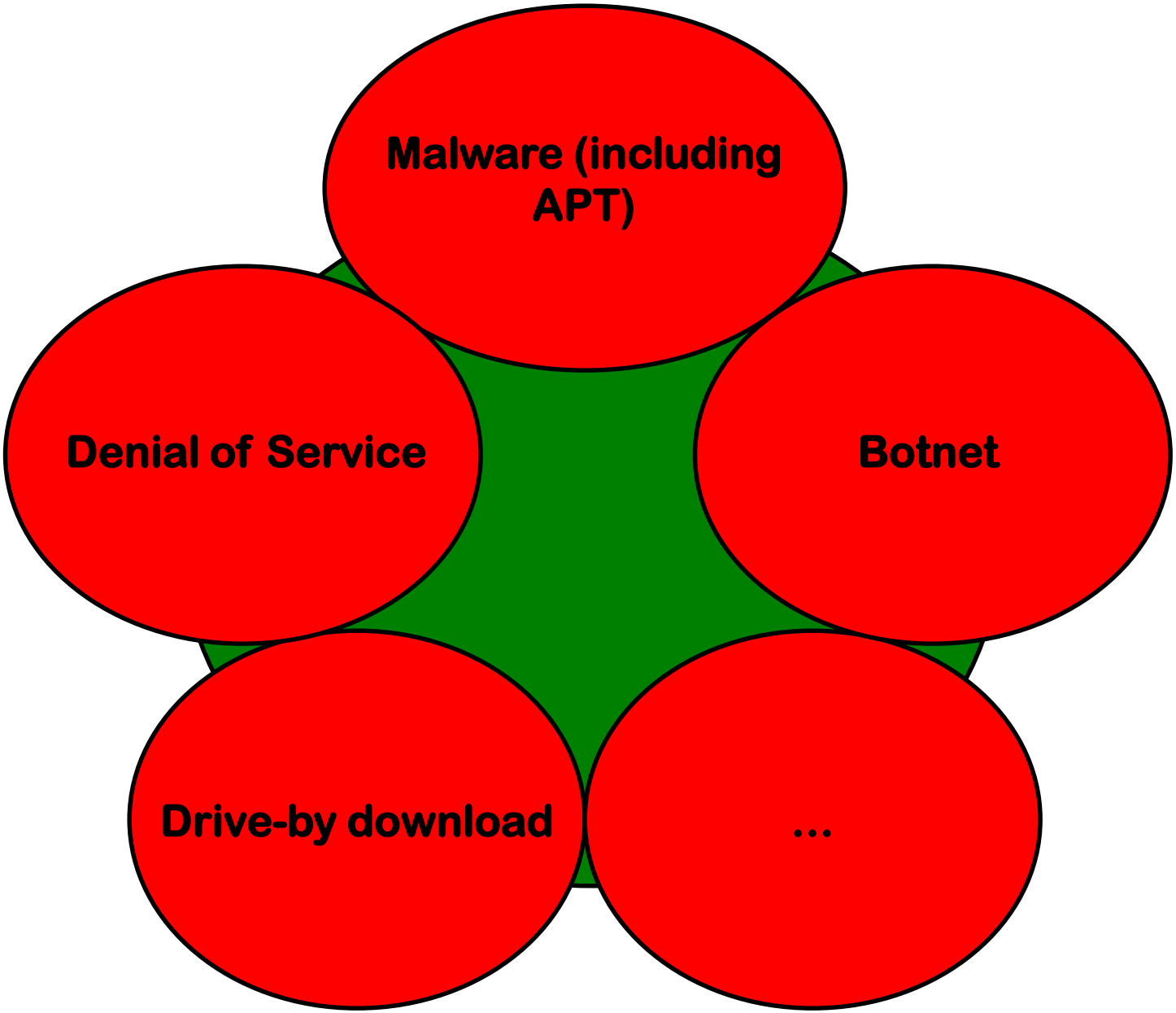
- The Cybersecurity Dynamics framework: Concept
- The x-axis: First-principle cybersecurity modeling
- The y-axis: Cybersecurity data analytics
- The z-axis: Cybersecurity metrics
- Conceptual clarifications
- Takeaway messages

Why the Cybersecurity Dynamics Approach?

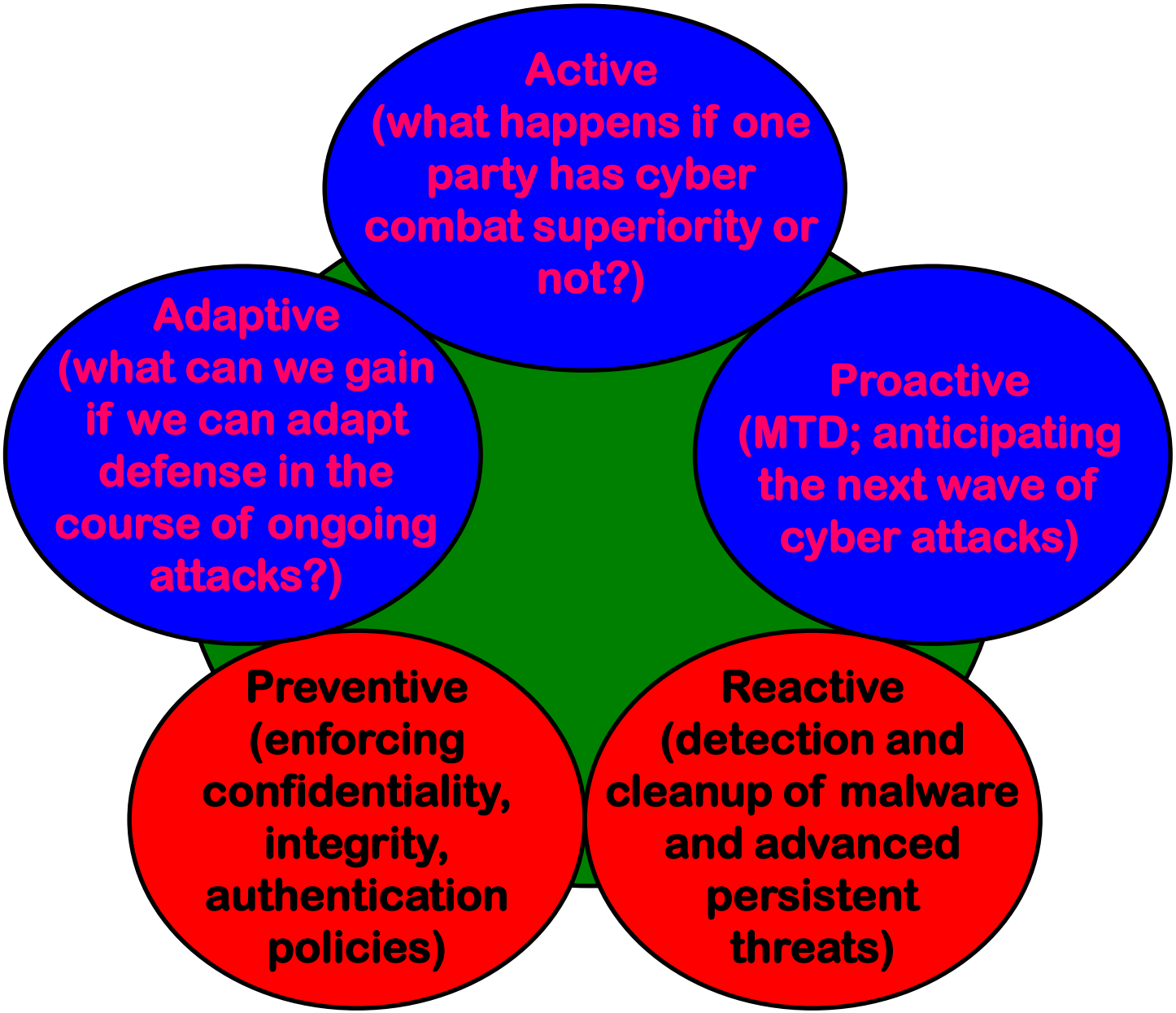
Inspiration: A subject (or discipline) often studies a single, most fundamental concept.

- ❑ Physics: Interaction
- ❑ Cryptography: Indistinguishability
- ❑ Cybersecurity: ??????????
  - ❖ How can we even get some clues?

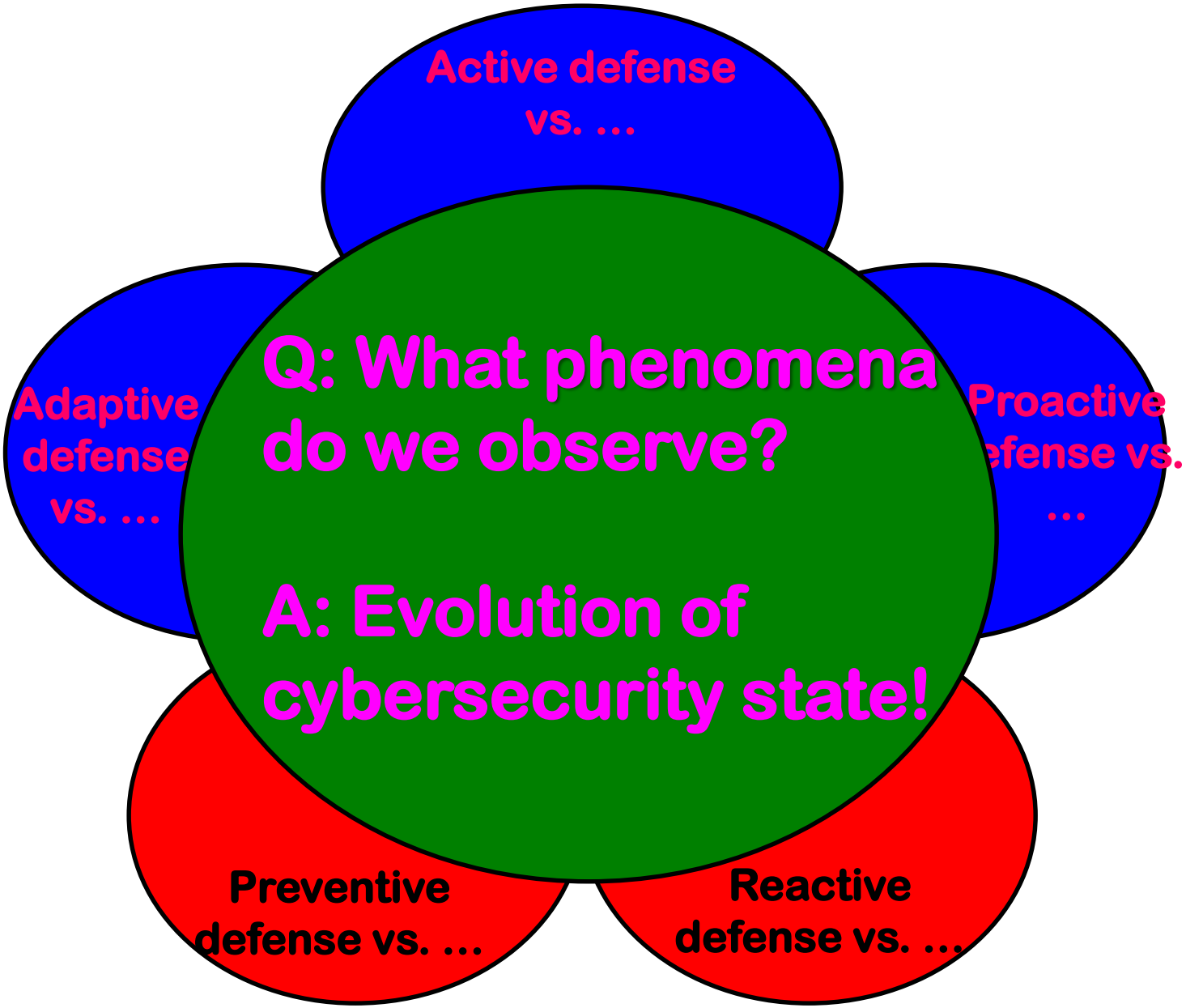
Observation: Kinds of Attacks



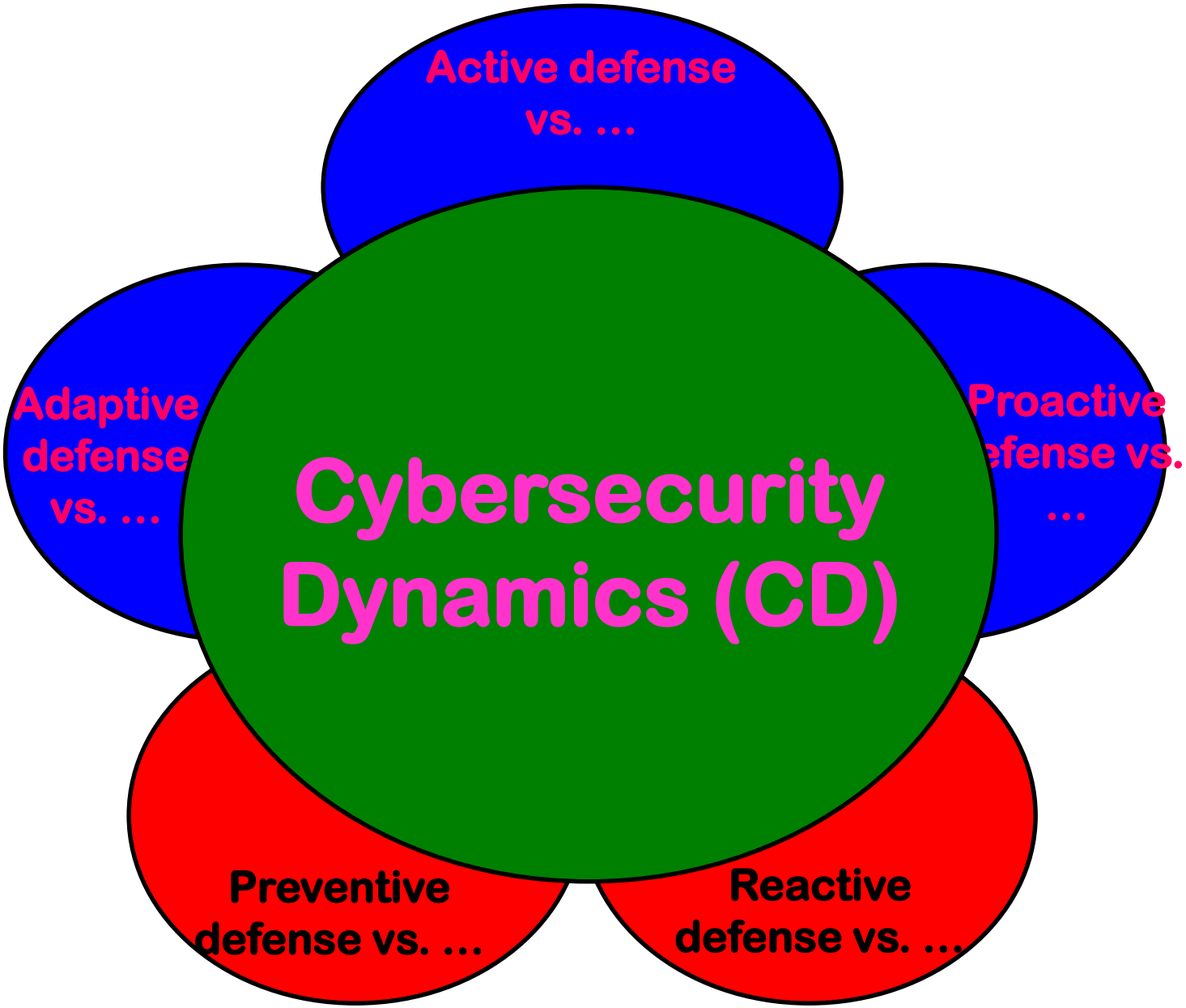
Observation: Kinds of Defenses



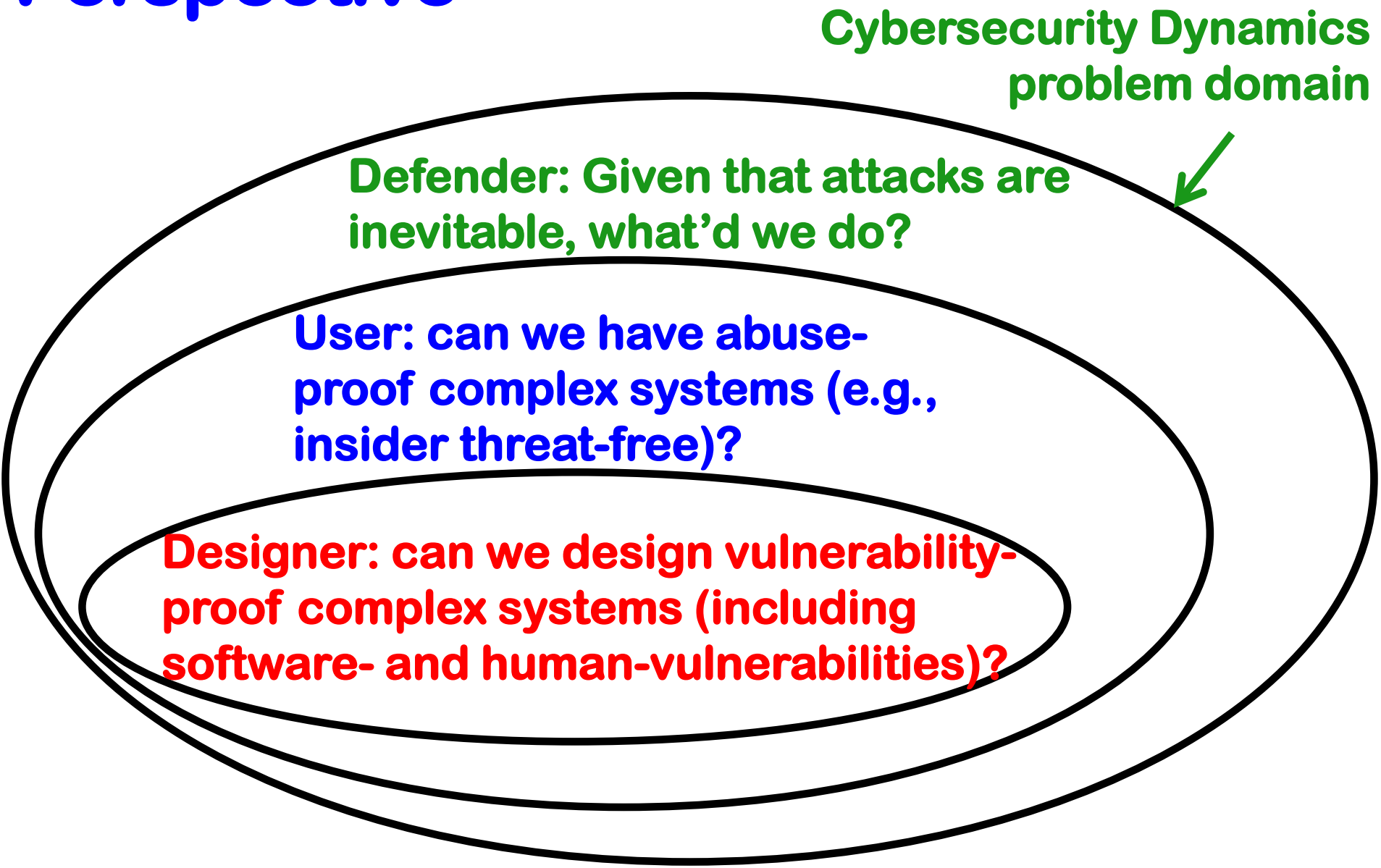
Fundamental Phenomena in Cyberspace



The Evolution Phenomena Hint Dynamics



Putting Cybersecurity Dynamics in Perspective



Mission Statement of Cybersecurity Dynamics

A systematic framework for modeling, quantifying, managing cybersecurity from a holistic perspective

- Centered at modeling the evolution of the global cybersecurity state caused by attack-defense interactions --- largely influenced by Physics.
- Cybersecurity Dynamics = Mathematicalization of "knowing yourself and knowing your enemy"
- Holistic vs. building-blocks perspectives
- Connecting the many dots

Research Methodology Towards Achieving the Ultimate Goal

Strong assumptions

Weak enough assumptions

Weakening assumptions

One abstract world → Another abstract world $\cong$ The real world

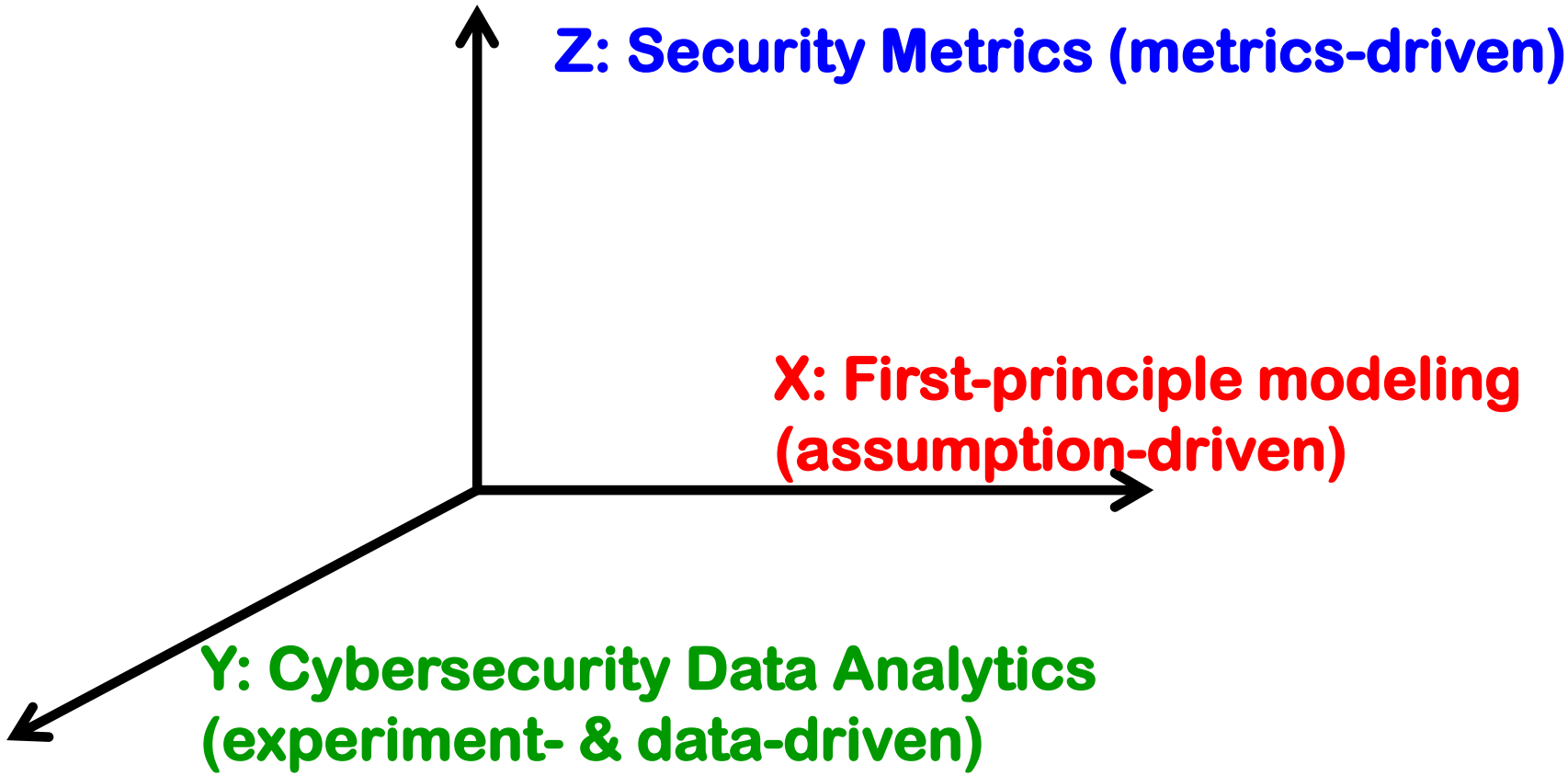
Technical Approach: An X-Y-Z-t "Coordinate System"

Metaphor:

- Human kind used the X-Y-Z-t coordinate system to explore the universe.
- Cybersecurity Dynamics naturally leads to the following X-Y-Z-t coordinate system that can be used to explore cyberspace.
- The t-axis means everything is dynamic (time-dependent).

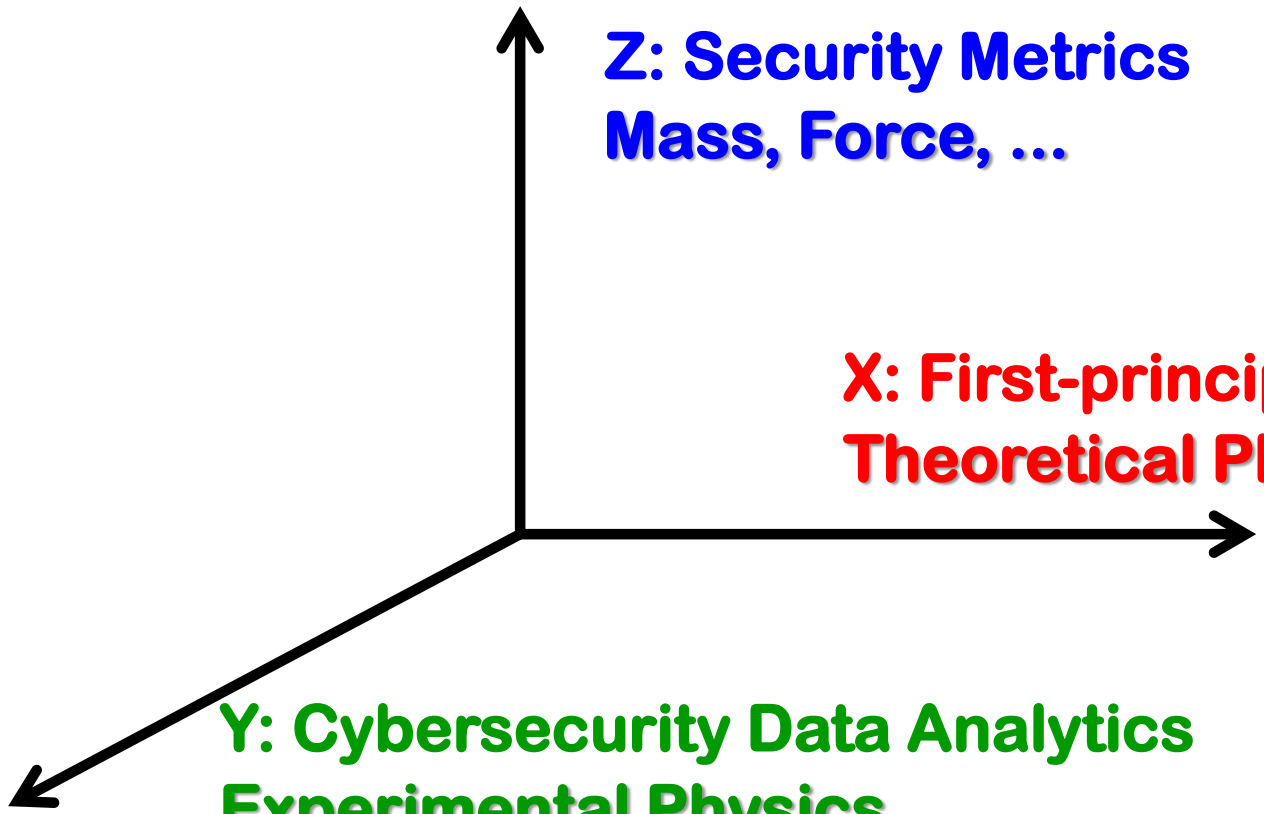
The X-Y-Z-t "Coordinate Systems"

T: 4th dimension (time)



One Analogy: Physics

T: 4th dimension (time)



Z: Security Metrics
Mass, Force, …

X: First-principle modeling
Theoretical Physics

Y: Cybersecurity Data Analytics
Experimental Physics

Example Research Topics

T: 4th dimension (time)

Z: Security Metrics

Defining security/resilience metrics

Axiomatic properties

Measurements of metrics

X: First-principle modeling

Dynamical System models

Dynamical Systems: A mathematical tool that amazingly has been playing critical roles in both macroscopic cybersecurity models (i.e., Cybersecurity Dynamics) and Microscopic Ciphers mechanisms!

Vulnerability analysis & detection & prediction

"Grey-box" predictive models

Adversarial Machine Learning

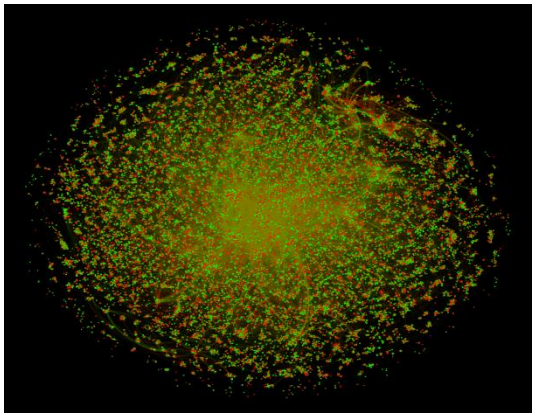
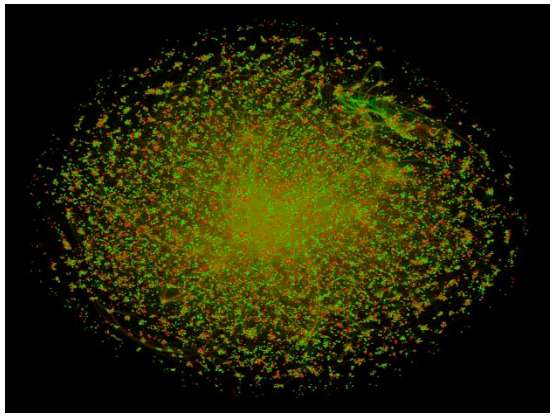
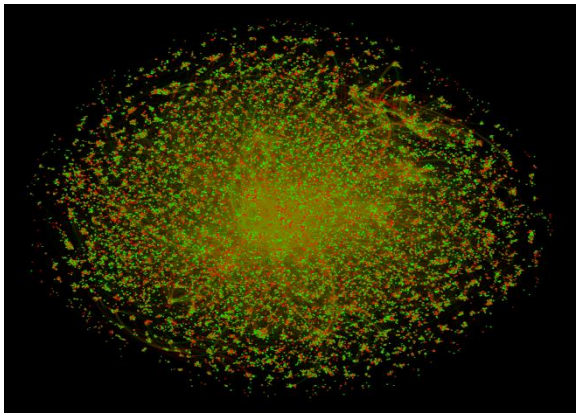
Outline

- ❑ The Cybersecurity Dynamics framework: Concept
- ❑ **The x-axis: First-principle cybersecurity modeling**
- ❑ The y-axis: Cybersecurity data analytics
- ❑ The z-axis: Security metrics
- ❑ Conceptual clarifications
- ❑ Takeaway messages

X-axis: First-Principle Modeling

- What is first-principle?
- It might mean different things to different people.
- To me, it means:
  - Starting from the original problem (not necessarily the current understanding): What is the most fundamental concept/phenomenon to study?
  - Building models as realistic as possible, while using as few parameters as possible and making as weak assumptions as possible.

How to Get Started Technically?

Computer state evolves: green -- secure; red -- compromised

Three kinds of outcomes of evolution of global security state

Natural question: what are the governing laws?

Complex Network (Network Science) based representation:

- Nodes abstract computers
  - Node state: green -- secure; red -- compromised
- Edges abstract who can attack whom directly

Results in First-Principle Modeling

Methodology: Divide → Conquer → Unify

- Preventive and reactive cyber defense dynamics [IEEE TDSC 2011, ACM TAAS 2012, IEEE TDSC 2012a, IEEE TDSC 2012b, Internet Math 2014, HotSoS'14a, Internet Math 2015a, IEEE TNSE 2018, Manuscript 2018]
- Adaptive cyber defense dynamics [ACM TAAS 2014]
- Active cyber defense dynamics [Internet Math 2015b, GameSec'13, HotSoS'15]
- Proactive cyber defense dynamics [HotSoS'14b]

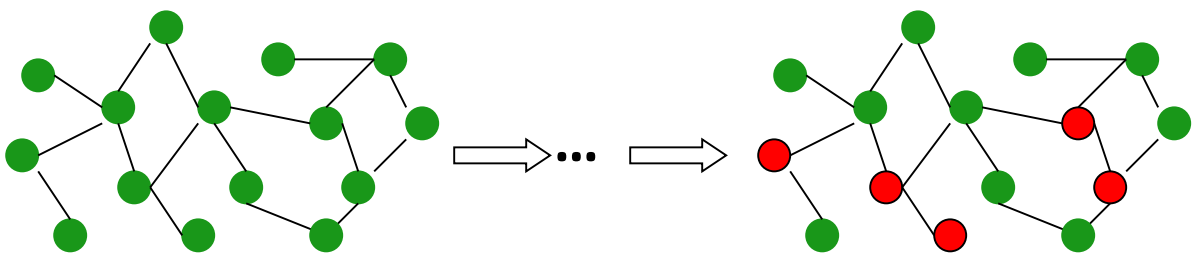
Preventive and Reactive Defense Dynamics [ACM TAAS 2012, IEEE TNSE 2018]

Characterizing the dynamical evolution of the global cybersecurity state caused by interactions between

- Preventive defense: attack prevention
- Reactive defense: anti-malware

and

- Push-based attacks: malware spreading
- Pull-based attacks: drive-by download



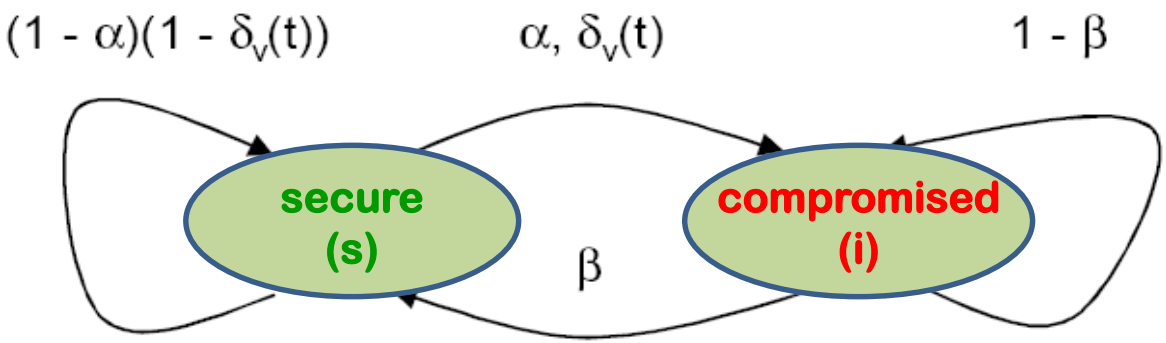
(● secure node ● compromised node)

Preventive and Reactive Defense Dynamics

High-level idea of modeling the dynamics:

- **G=(V, E): attack-defense interaction structure, where (u,v)$\in$E means computer u can directly attack v**
- **α: capability of pull-based attacks against preventive defenses**
- **γ: capability of push-based attacks against preventive defenses**
- **β: capability of reactive defense**

Mathematical Model



State

The mast is

Need to analyze: How does the probability that node v is compromised evolve over time?

Difficulty: There can be 10^9 nodes ($|V| = 10^9$)!

For $v \in V$

$$i_v(t+1) = [1-(1-\alpha)(1-\delta_v(t))]\, s_v(t) + (1-\beta) i_v(t).$$

where

$$\delta_v(t) = 1 - \prod_{(u,v)\in E} [1-\gamma i_u(t)]\,.$$

Partial Understanding

[w/ W. Lu and L. Xu; ACM TAAS 2012]

An example theorem:

Cybersecurity meaning of the theorem:

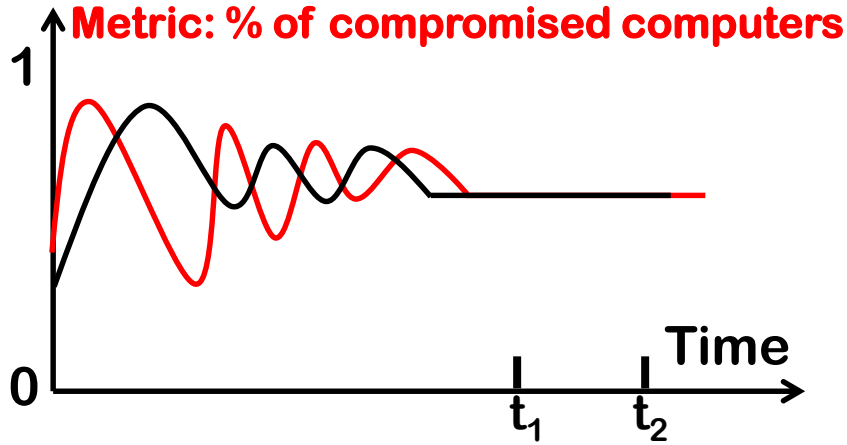
Under a certain circumstance, the dynamics converges to an equilibrium.

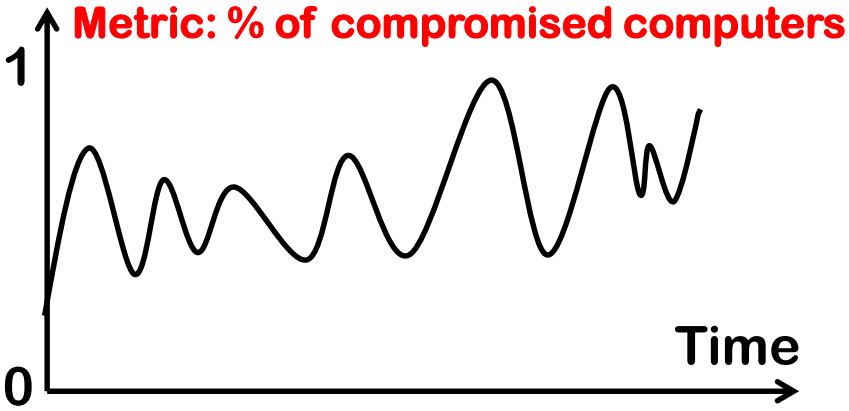
The "circumstance": a special parameter regime (representing certain push-based and pull-based attacks against certain preventive and reactive defenses).

Q: Why is converging to equilibrium important?

Why Is Equilibrium Important?

A: Equilibrium is necessary for measuring the global cybersecurity state or situational awareness (e.g., the % of compromised computers via sampling when the model parameters are not known) and meaningful real-time cyber defense decision-making.





When the dynamics enters equilibrium, the defender is given sufficient time to measure the global cybersecurity state.

The "unmanageable" situation of the global state (e.g., Chaotic): it is infeasible to measure the global state.

Exciting New Result (of Knowledge)

Parameter Universe

Complete understanding of preventive and reactive cyber defense dynamics!

The dynamics is globally stable in the entire parameter universe. The convergence speed is also characterized.

Even newer result (in progress): Global stability in the parameter universe while unifying two classes of models.

New Result:
paper to be submitted (with Z. Lin and W. Lu)

- A more general class of preventive and reactive cyber defense dynamics
- Unifying and going beyond two models *that have been widely investigated in the literature*
  - The Π-model mentioned above (discussed above)
  - The Σ-model in the literature
- Result: The generalized preventive and reactive cyber defense dynamics is still globally stable.

The Π-Model [ACM TAAS 2012, IEEE TNSE 2018]

Model Review

Preventive and Reactive Cyber Defense Dynamics

$$\frac{di_v(t)}{dt} = -\beta_v i_v(t) + \left[1 - (1-\alpha_v) \prod_{u \in \mathcal{N}_v} (1 - \gamma_{vu} i_u(t))\right] (1 - i_v(t)) \quad (3)$$

Result:

- Dynamic sytem (3) is globally asymptotically stable in the entire parameter universe for arbitrary $\mathcal{G}$.

The Σ-Model by Other Researchers

Model Review

SIS model

$$\frac{\mathrm{d}i_v(t)}{\mathrm{d}t} = -\beta_v i_v(t) + \sum_{u \in \mathcal{N}_v} \gamma_{vu} i_u(t)(1 - i_v(t)) \tag{1}$$

ϵ-SIS model

$$\frac{\mathrm{d}i_v(t)}{\mathrm{d}t} = -\beta_v i_v(t) + \left(\alpha_v + \sum_{u \in \mathcal{N}_v} \gamma_{vu} i_u(t)\right)(1 - i_v(t)). \tag{2}$$

Result for SIS model:

- If $\lambda_{max}(-B + \Gamma) \leq 0$, the equilibrium $\mathbb{0}$ is globally asymptotically stable.
- If $\lambda_{max}(-B + \Gamma) > 0$, there exists an $i^* \in (0, 1)^n$ such that i^* is globally asymptotically stable

where B=$diag(\beta_1, \ldots, \beta_n)$ and $\Gamma = [\gamma_{vu}], v, u \in \{1, \ldots, n\}$.

The Unified Model

Generalized Preventive and Reactive Cyber Defense Dynamics

$$f_v(i(t)) \stackrel{def}{=} \frac{di_v(t)}{dt} = -h_v(i(t)) \times i_v(t) + g_v(i(t)) \times (1 - i_v(t)) \quad (4)$$

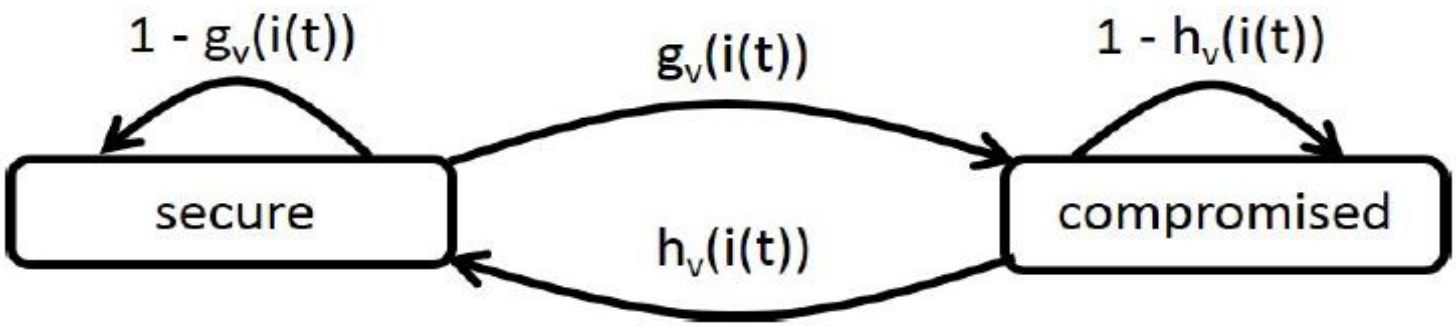


Figure: State-transition diagram.

Natural setting: $\frac{\partial h_v(i)}{\partial i_u} \leq$ and $\frac{\partial g_v(i)}{\partial i_u} \geq 0 \rightarrow$ Cooperative system

New Result for the Unified Model

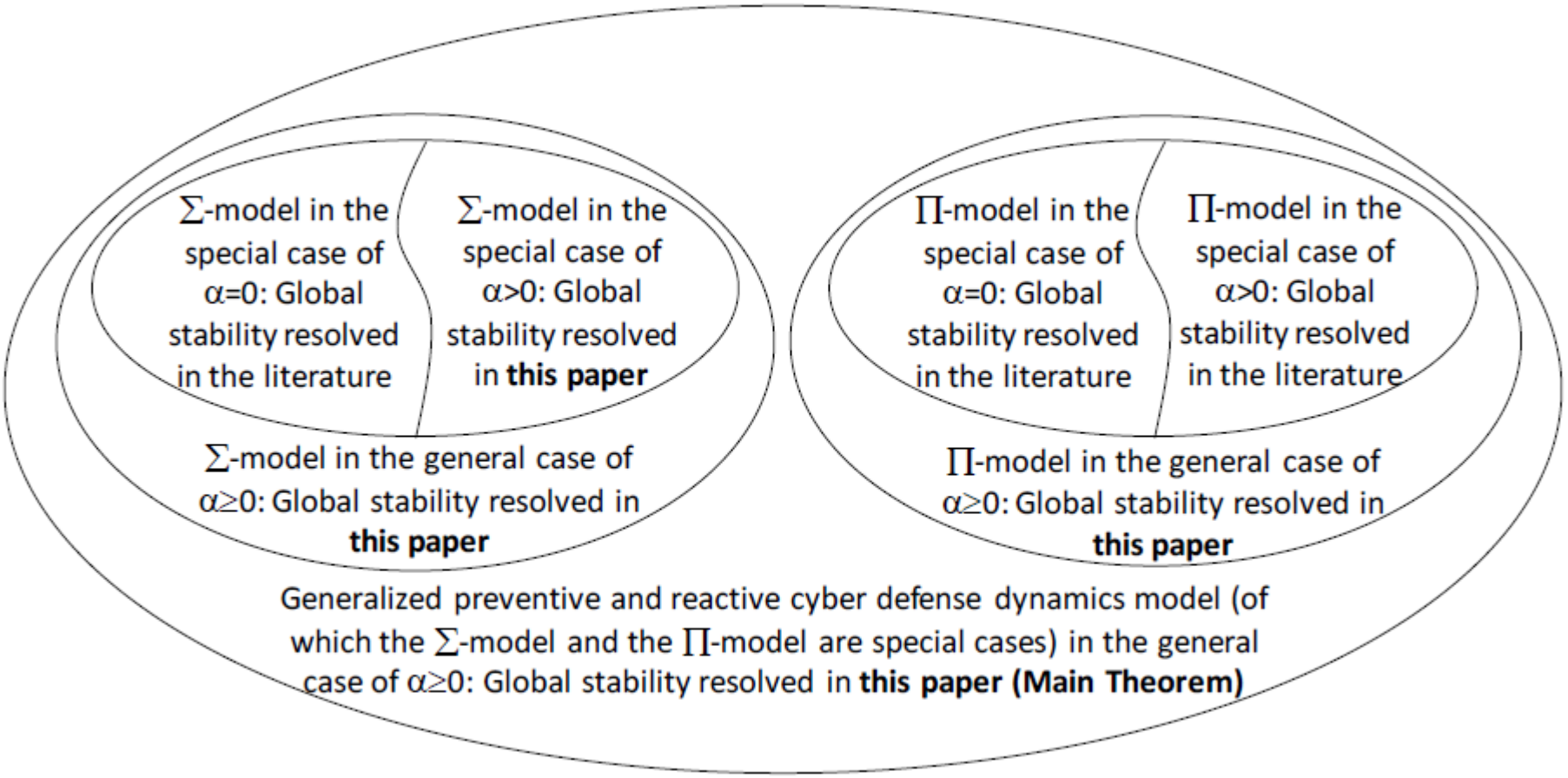


Figure: Overview of relation of exists model and Our generalized model.

Active Cyber Defense Dynamics

[w/ W. Lu, R. Zheng, X. Yi, and H. Li]

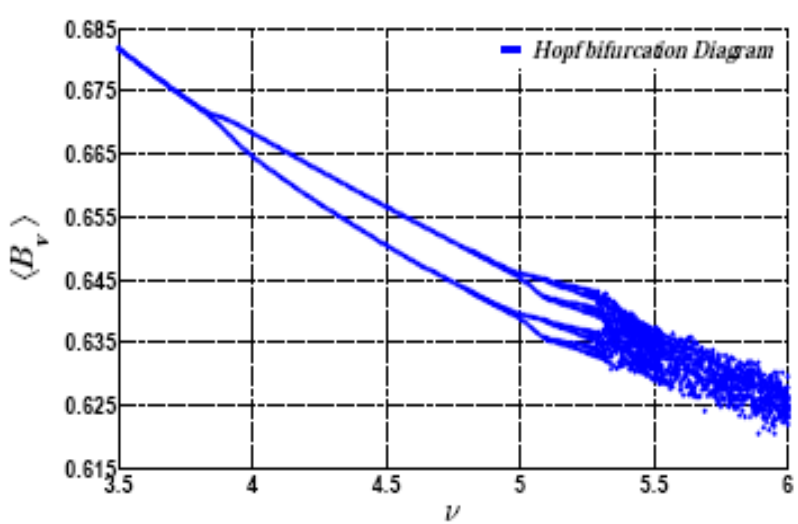
Idea: Defender uses "defense-ware" or "white-worms" to defend push-based attacks

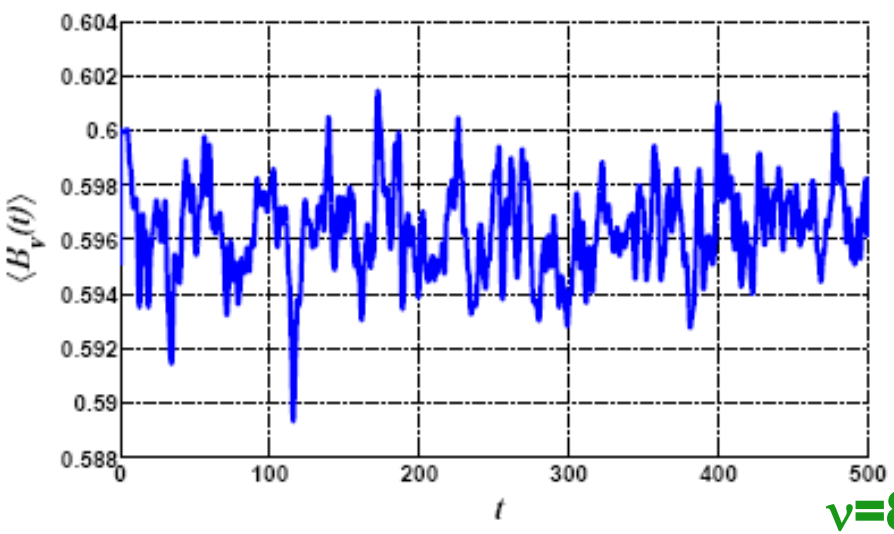
- [Internet Math 2015] characterizing effectiveness of Active Cyber Defense
  - Eliminating an asymmetry that benefits attackers
  - Still, not universally powerful.
- [GameSec'13] Orchestrating active defense

Active Cyber Defense Dynamics

[Internet Math 2015, GameSec'13, HotSoS'15]

- The dynamics exhibits rich phenomena [HotSoS'15]: Bifurcation and Chaos (i.e., unmanageable situation mentioned above) are relevant in cybersecurity.
- Implication: Need to use Control Theory to prevent them!





ν=8

Open Problems

- Deep understanding and characterization of each kind of cybersecurity dynamics
  - After spending 10+ years on preventive and reactive cyber defense dynamics, there are still many open problems that have yet to be tackled
- Should we have a unique Cybersecurity Dynamics Theory for each kind of systems? Or: Should we have a universal Theory and adapt it to specific settings?
  - Cyber vs. CPS vs. IoT vs. (future buzzword)

Open Problems (cont.)

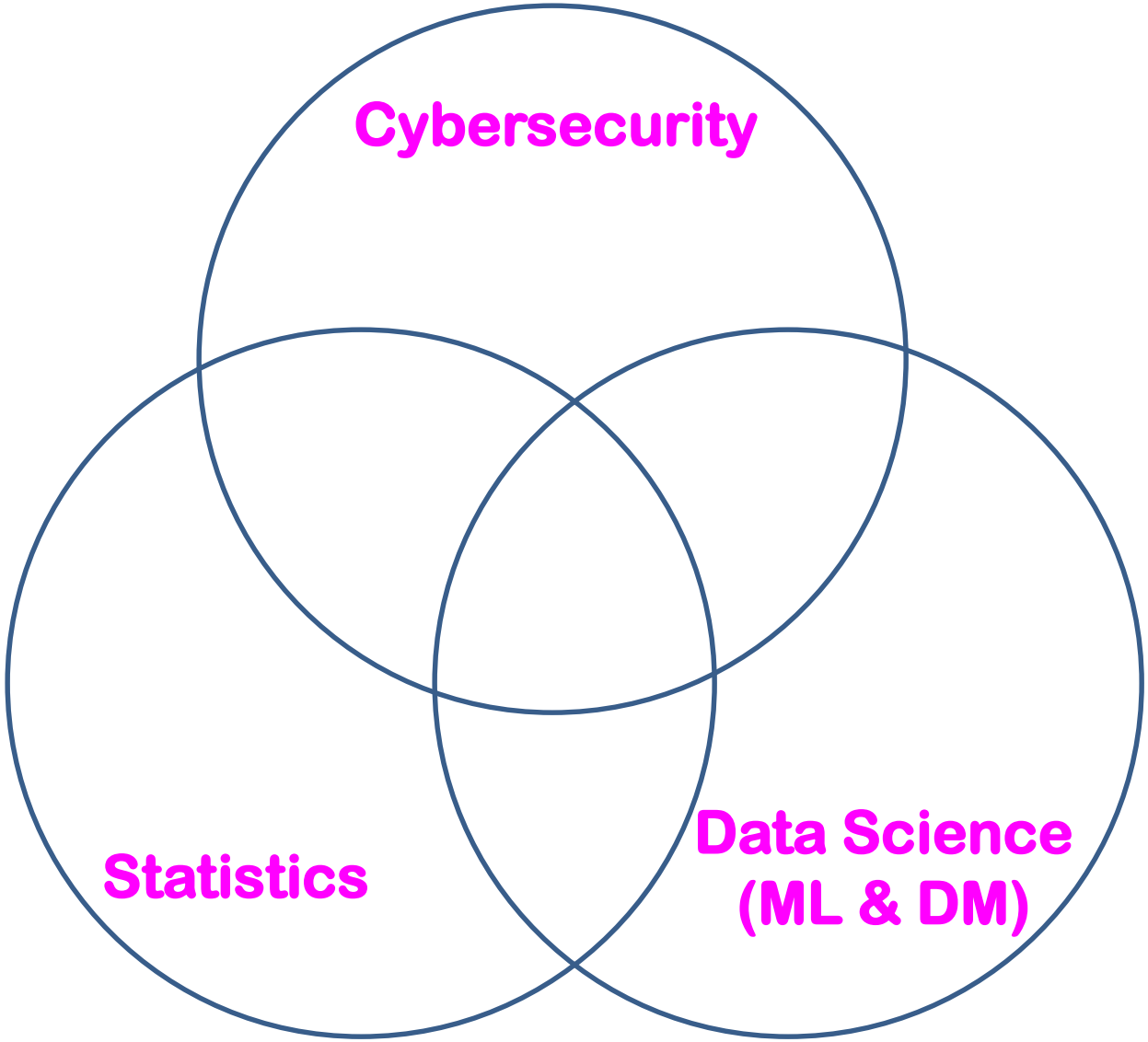
Technical Barriers That Cutting Across All Kinds of Dynamics

- Scalability barrier
- Dependence barrier
- Nonlinearity barrier
- Parameter/structure dynamics barrier
- Transient behavior barrier
- Uncertainty barrier
- Human factor barrier

Outline

- The Cybersecurity Dynamics framework: Concept
- The x-axis: First-principle cybersecurity modeling
- **The y-axis: Cybersecurity data analytics**
- The z-axis: Security metrics
- Conceptual clarifications
- Takeaway message

Cybersecurity Data Analytics Scope



Y-axis: Cybersecurity Data Analytics

Motivating questions:

1. **How can we predict the evolution of *, and to what extent can we predict (i.e., predictability)?**
2. **How can we obtain the model parameters used in the first-principle models?**
3. **How can we validate/invalidate the assumptions made in the first-principle models? (To Do)**

In what follows we will look into some results in regards to the 1st and 2nd questions mentioned above.

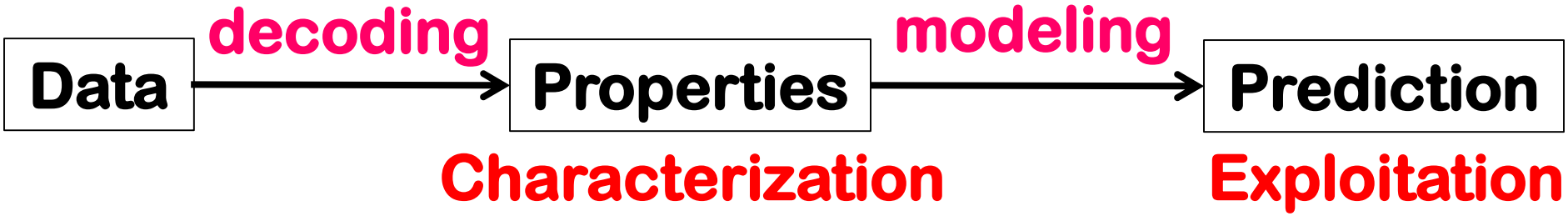
Y-axis: Part 1

Towards Answering Motivating Question 1

Example: How to predict/forecast in cybersecurity domain?

- Black-box cybersecurity data analytics
  - "Blindly" use Machine Learning models
- Grey-box cybersecurity data analytics
  - How to take full advantage of the available data?
- White-box cybersecurity data analytics
  - Knowing the processes that induce the data
  - First-principle modeling and analysis

Grey-Box Prediction Methodology



- Characterizing the properties of data is "cool"
  - Deepening our understanding & knowledge
- Being able to predict is even "cooler"
  - Enabling cybersecurity situation awareness and proactive allocation of defense resources

Grey-Box Prediction Methodology

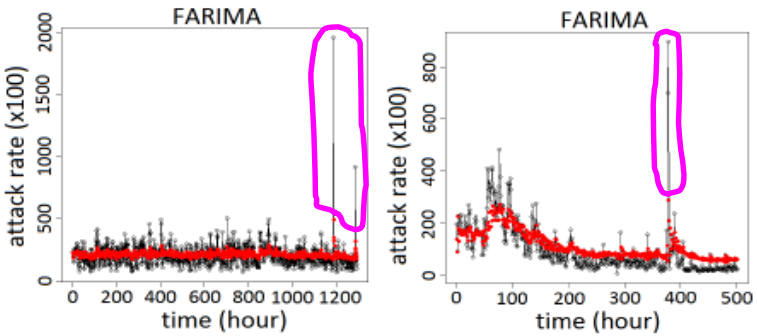
Data $\xrightarrow{\text{decoding}}$ Properties $\xrightarrow{\text{modeling}}$ Prediction

- Long Range Dependence
- Extreme value
- Dependence

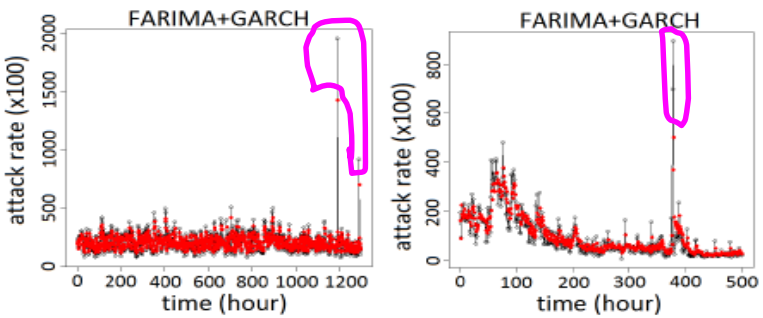
- Point
- Distribution

- Grey-box prediction methodology is exciting
- because it paves a way for cyber "weather forecasting"

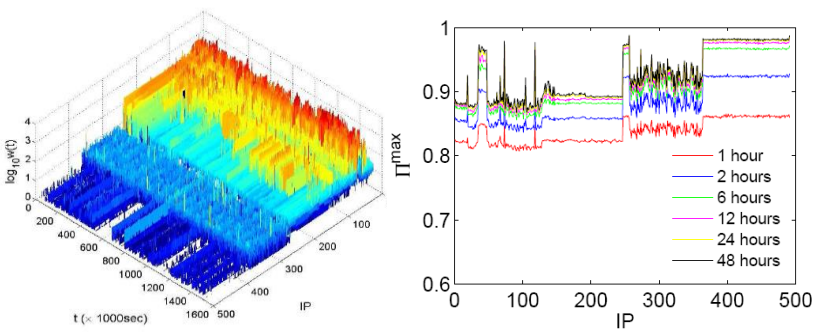
Grey-Box Prediction of Attack Rates



(h) Period III (i) Period IV



(c) Period III (d) Period IV



Grey-box prediction: using statistical properties exhibited by the data to guide the design of prediction models

FARIMA can predict attack rates 1-hour ahead of time with ~80% accuracy [IEEE TIFS 2013]: missed extreme values

FARIMA+GARCH can predict attack rates 1-hour ahead of time with ~88% accuracy, by additionally coping with extreme values (i.e., spikes or outliers) [IEEE TIFS 2015]: accommodated extreme values

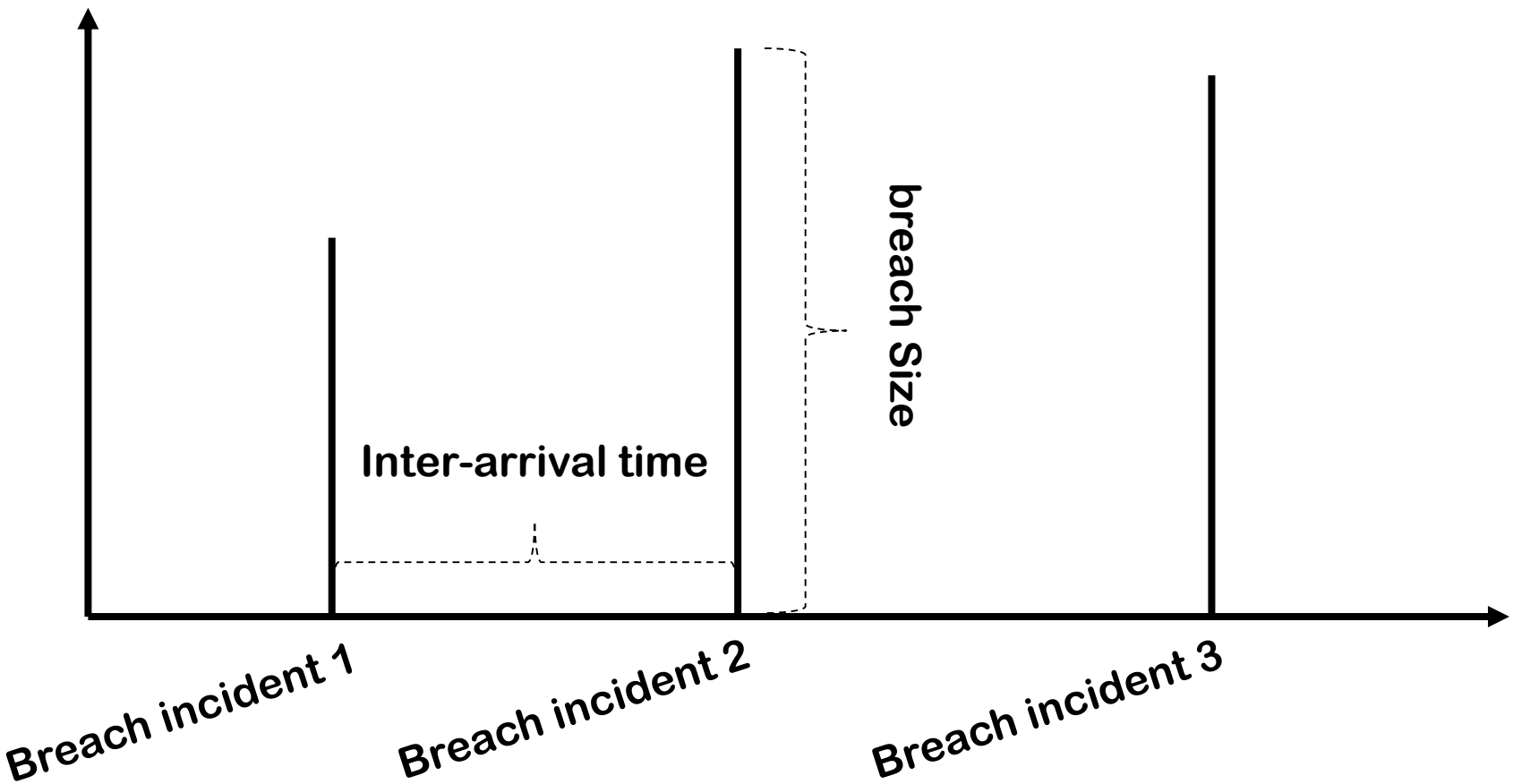
Predictability upper bound: ~93% [PLoS ONE 2015]

Ongoing investigation: how to achieve 93% upper bound prediction accuracy?

More Recent Result I

"Modeling and Predicting Cyber Hacking Breaches"

[w/ M. Xu, K. Schweitzer, R. Bateman; IEEE T-IFS, 2018]



More Recent Result I (cont.)

Some Findings:

- Hacking breach incidents inter-arrival times should be modeled as stochastic process, not distribution.
  - A type-1 log-ACD (autoregressive conditional duration) model proposed in our paper
- Hacking breach sizes should be modeled as stochastic process, not distribution.
  - Log-transformed hacking breach sizes can be modeled by ARMA(1,1)-GARCH(1,1) with innovations following a mixed extreme value distribution

More Recent Result II

"Modeling Multivariate Cybersecurity Risks"

[w/ P. Chen, M. Xu, T. Hu; Journal of Applied Statistics, to appear]

- **Use copulas to model the dependence contained in multivariate time series (69 dimensions)**
- **Accurate prediction of such multivariate time series**
- **Dependence is one of the fundamental problems that must be tackled adequately!**

Open Problems

Building a systematic body of knowledge in cybersecurity-oriented grey-box statistics, especially in coping with:

- ❑ Univariate time series: each observation is a number
- ❑ Multivariate time series: each observation is a vector
- ❑ Functional time series: each observation is a function
- ❑ Networked time series: each observation is a network

Y-axis: Part 2
Towards Answering Motivating Question 2

Example: Towards deriving parameters values representing system/software susceptibility/vulnerability, attack power, defense power, etc.

- Software vulnerability/susceptibility
  - Vulnerability detection
- Adversarial malware detection
  - HashTran-DNN

Vulnerability Detection

- Q: How can we represent programs as *vectors* that accommodate syntax and semantic information suitable for vulnerability detection?
- A: *Syntax-based, Semantics-based, and Vector Representations* (SySeVR)
  - The first systematic framework for using deep learning to detect vulnerabilities.
  - Supersede our VulDeePecker [NDSS'18]
- With Zhen Li et al. (HUST)

SySeVR Overview

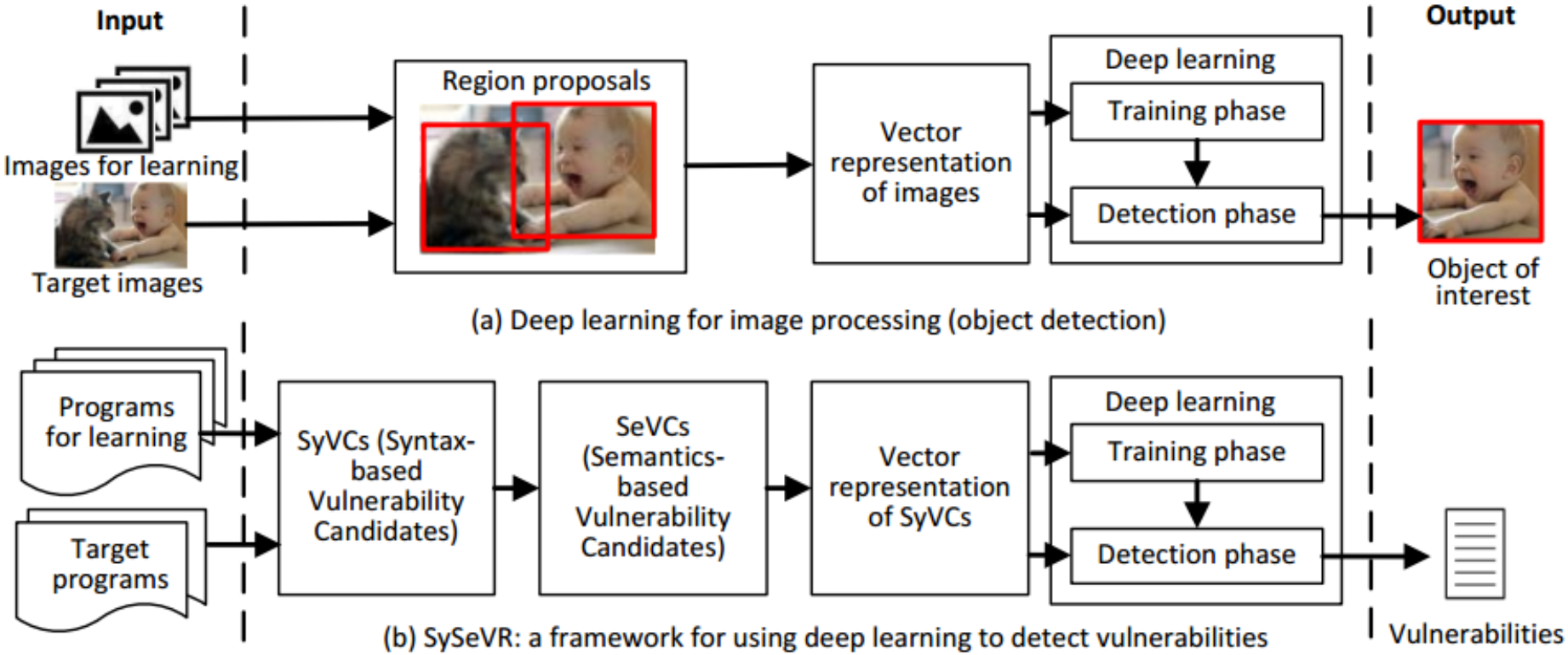


Figure 1: (a) The notion of region proposal in image processing. (b) The SySeVR framework inspired by the notion of region proposal and centered at obtaining SyVC, SeVC, and vector representations of programs.

Using SySeVR in Practice

Applying BGRU to detect vulnerabilities in software products

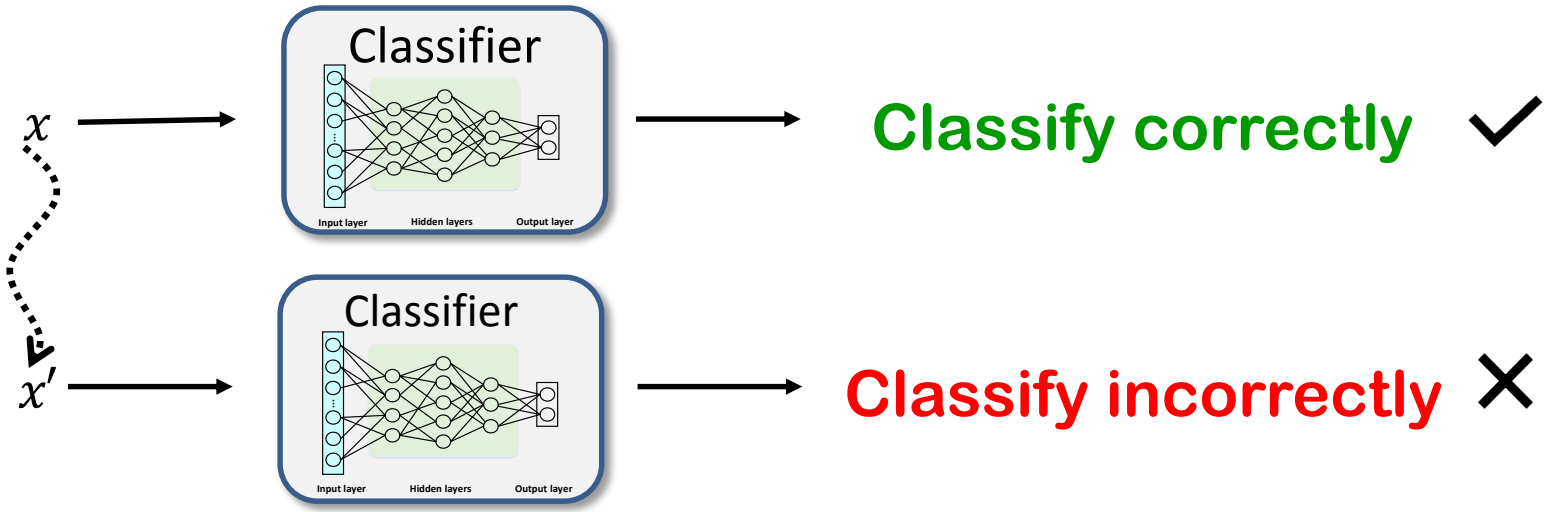
- 4 products: Libav, Seamonkey, Thunderbird, and Xen
- SySeVR detected 15 vulnerabilities that were not reported in the NVD.
- 7 are unknown (i.e., 0-day) and 8 have been “silently” patched by the vendors.

Result on Adversarial Malware Detection

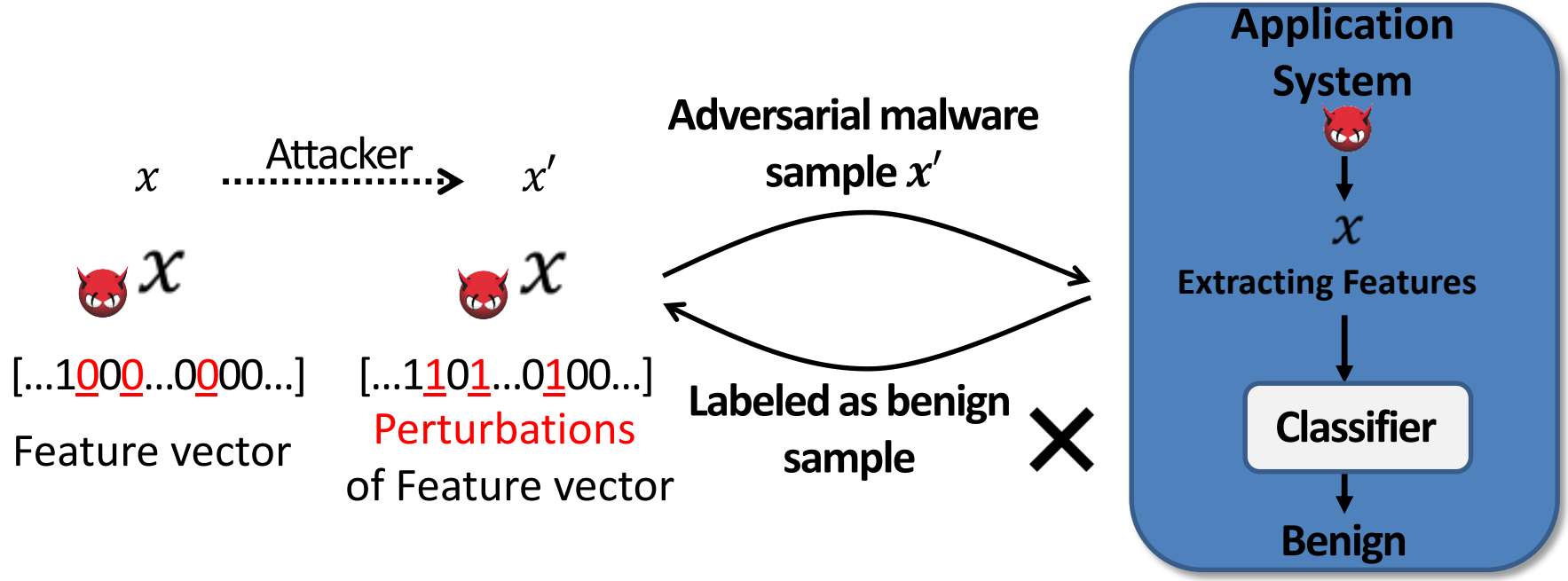
- HashTran-DNN: A Framework for Enhancing Robustness of Deep Neural Networks against Adversarial Malware Samples. Under review
- With Deqiang Li et al. (NJUST and FIU)

Adversarial examples fool DNNs

- Testing: Input a sample x to a DNN classifier which classifies x correctly.
- Attack: Attacker perturbs x into adversarial example x', such that
  - x' is misclassified by DNN classifier
  - x' preserves the functionality and the perturbations are small, e.g. $\|x' - x\|_0 \leq \epsilon$

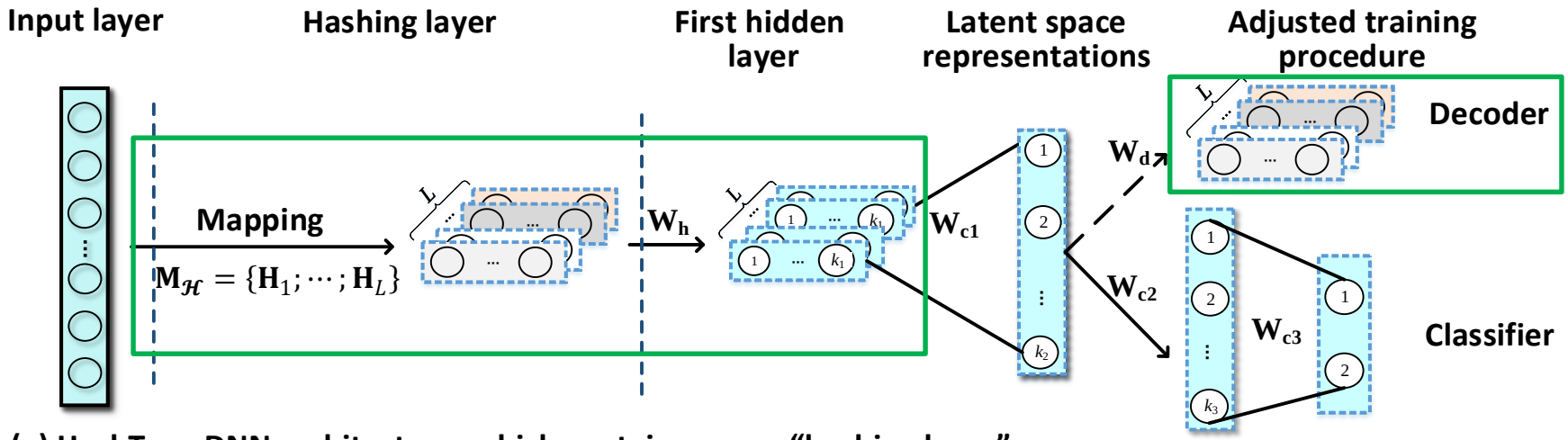


Example: Malware detection



- Attacker evades the classifier in the application system by exploiting the adversarial example x'.
- Attacker knows some knowledge (e.g., features) leveraged by application system.
- Attacker adds components into x with $\|x' - x\|_0 \leq \epsilon$, e.g. $\epsilon = 10$.

Hash-Tran DNN framework

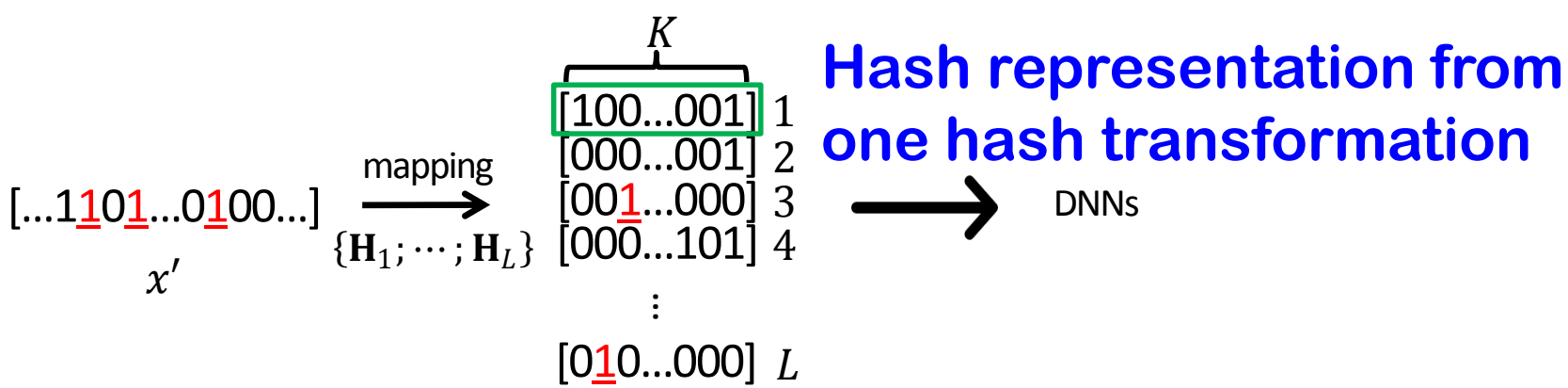


(a) HashTran-DNN architecture, which contains a new "hashing layer"

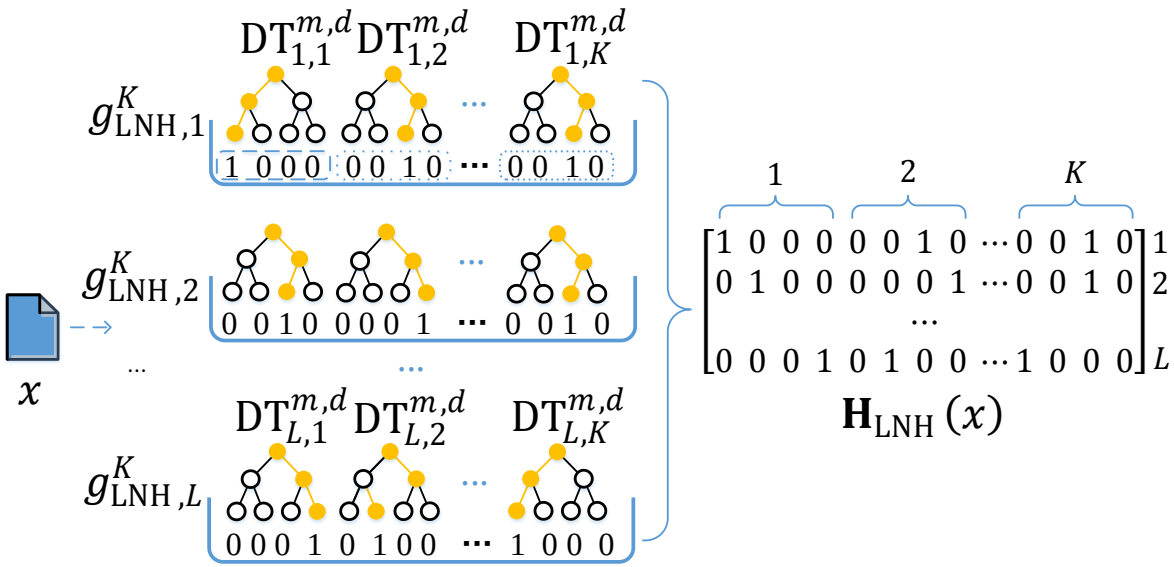
- Hashing layer contains hash transformations with locality-preserving property to map the input vector to multiple representations learned by DNNs separately.
- Denoising autoencoder further reduces the impact of the perturbation in the hash representations.

Instantiation: Locality-Sensitive Hashing (LSH)

- Bit sampling: Given a binary vector, randomly select a bit from the binary vector.
- Hash transformations
  - Step 1: Index K times bit sampling function uniformly at random (with replacement) to construct one hash transformation $\mathbf{H}_1$;
  - Step 2: repeat step 1 L times.



Instantiation: Locality-Nonlinear Hashing



- A Decision Tree (DT) is learned from m times bit-sampling
- Steps of hash transformations are analogous to LSH.

Experimental Result

Performance of applying HashTran-DNN on Drebin dataset:

Defense	The Drebin dataset			Acc (%) under the FGS attack			Acc (%) under the CW attack		
	Acc	FNR	FPR	$\epsilon = 10$	$\epsilon = 20$	$\epsilon = 30$	$\epsilon = 10$	$\epsilon = 20$	$\epsilon = 30$
No defense (standard DNN)	98.66	3.18	1.25	69.60	44.91	21.70	3.63	0.00	0.00
RFN	97.02	5.57	2.86	85.11	69.48	48.64	11.85	0.16	0.00
Adversarial Training	98.85	10.1	0.74	96.41	91.72	80.00	1.09	0.00	0.00
HashTran-DNN w/ LSH-DAE	98.17	3.03	1.77	86.88	69.02	45.54	83.71	65.69	44.40
HashTran-DNN w/ LNH-DAE	98.11	5.23	1.74	85.47	74.38	66.72	97.03	97.03	97.03

Experimental Result

Performance of HashTran-DNN on Private dataset:

Defense	The Private dataset			Acc (%) under the FGS attack			Acc (%) under the CW attack (%)		
	Acc	FNR	FPR	$\epsilon = 10$	$\epsilon = 20$	$\epsilon = 30$	$\epsilon = 10$	$\epsilon = 20$	$\epsilon = 30$
No defense (standard DNN)	92.42	13.16	1.84	91.61	83.19	73.81	55.49	38.03	17.55
RFN	91.89	12.90	3.34	95.43	92.57	88.53	66.82	57.40	54.75
Adversarial Training	92.38	13.07	2.02	100.0	100.0	100.0	56.07	41.59	36.96
HashTran-DNN w/ LSH-DAE	92.16	12.79	2.76	96.79	94.24	91.27	93.82	88.25	84.44
HashTran-DNN w/ LNH-DAE	92.05	13.09	2.69	95.23	92.01	86.98	95.23	91.44	85.30

Insights:

- Standard DNNs can be ruined by adversarial samples, especially the CW attack.
- Adversarial Training is effective against FGS, but can be ruined by CW.
- RFN has a limited success against CW. HashTran-DNN is effective against both FGS and CW attacks

Open Problems

- What are more effective, and the best possible, vulnerability detection algorithms?
- How can we measure the residue vulnerability in a software (i.e., susceptibility)?
- Adversarial malware detection is largely open, affecting the measurement of measuring attack/defense capabilities.

Outline

- ❑ The Cybersecurity Dynamics framework: Concept
- ❑ The x-axis: First-principle cybersecurity modeling
- ❑ The y-axis: Cybersecurity data analytics
- ❑ **The z-axis: Security metrics**
- ❑ Conceptual clarifications
- ❑ Takeaway message

Cybersecurity Metrics

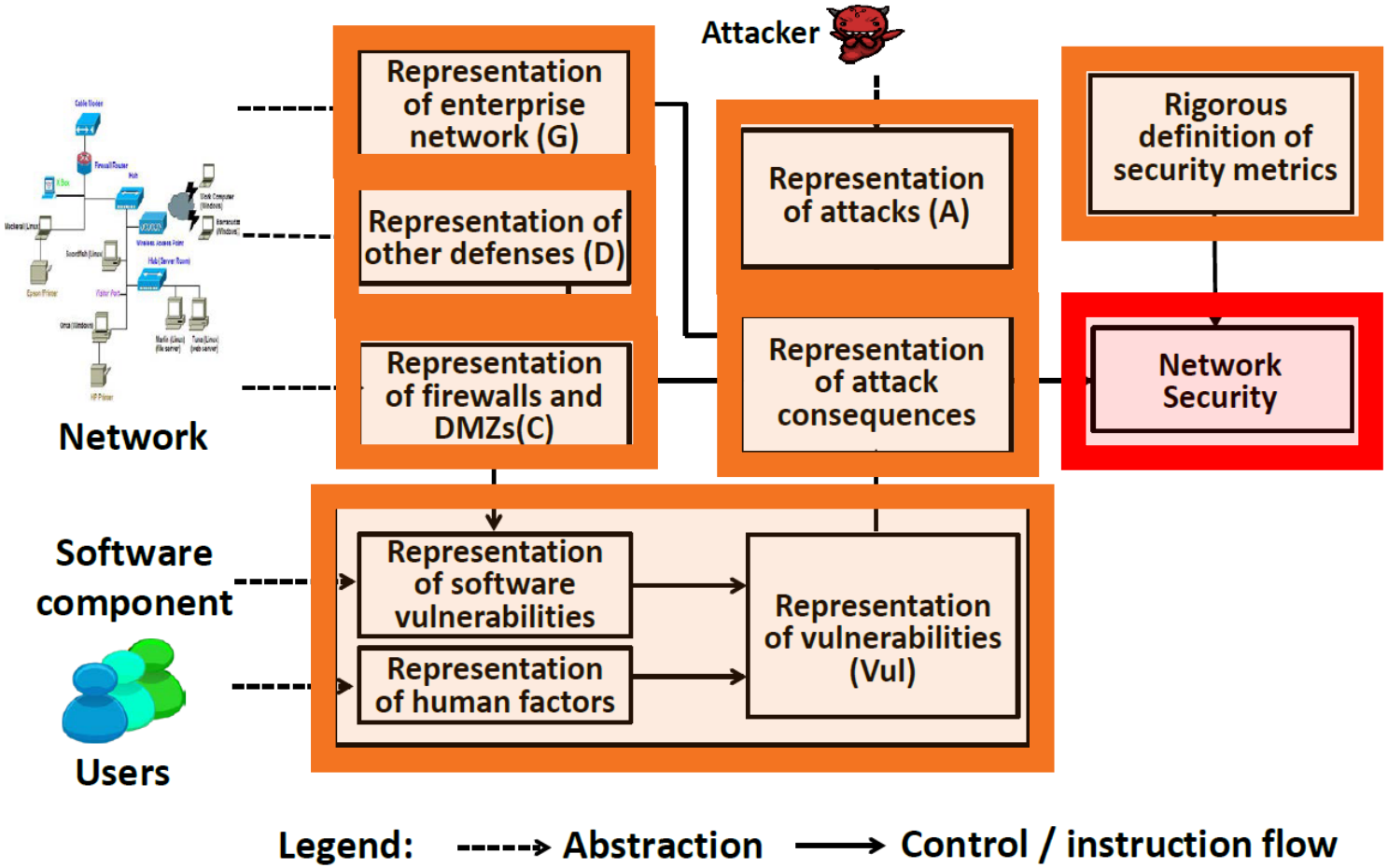
- Metrics is a notoriously difficult problem that has "tortured" the security community from day 1!
- We are making good progress

Some of Our Results

- First systematization on security metrics
  - ACM Computing Survey, Jan. 2017
- Quantifying the effectiveness of firewall, DMZ, etc
  - HotSoS'18
- Statistical Estimation of Malware Detection Metrics in the Absence of Ground Truth
  - IEEE T-IFS, 2018
- First systematization on system trustworthiness metrics
  - STRAM framework: Security, Trust, Resilience, Agility
  - Paper under review

A Framework for Quantifying the Security Effectiveness of Firewall and DMZ

[w/ H. Chen and J. Cho, HotSoS'2018]



Security Metrics

- Percentage of compromised applications (pca) at time t

$$\text{pca}(t) = |\{v \in V_{(app)} : \text{state}(v, t) = 1\}| / |V_{(app)}|$$

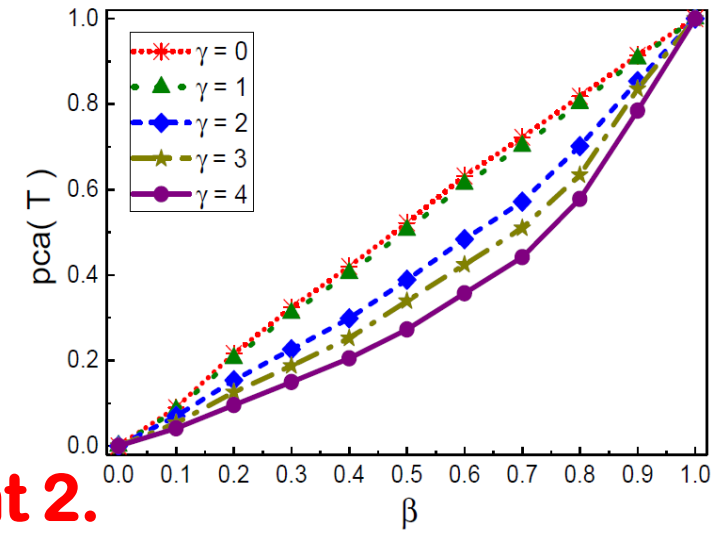
- Percentage of compromised server applications (pcsa) at time t

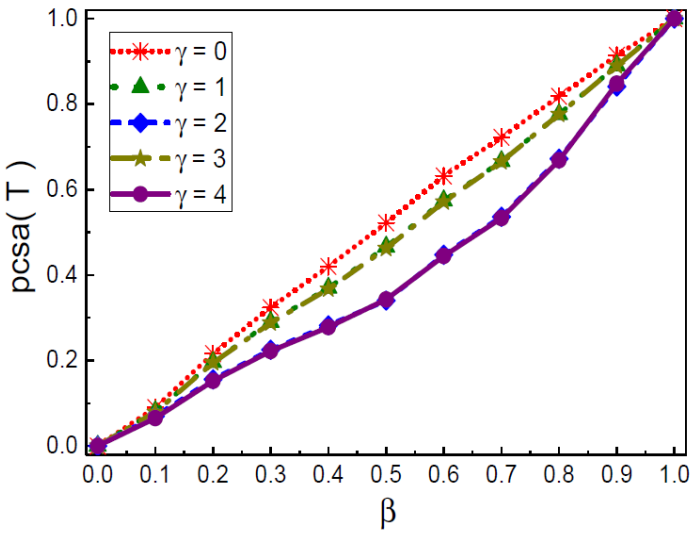
$$\text{pcsa}(t) = \frac{|\{v \in V_{(app)} \wedge \eta(v) \neq 0 : \text{state}(v, t) = 1\}|}{|\{v \in V_{(app)} \wedge \eta(v) \neq 0\}|}$$

- Percentage of compromised OSes (pcos) at time t

$$\text{pcos}(t) = |\{v \in V_{(os)} : \text{state}(v, t) = 1\}| / |V_{(os)}|$$

Sec. Effectiveness of Firewalls & DMZ (1)





Insight 2.

- When OSes are not vulnerable, security effectiveness of a fixed combination of firewalls and DMZ decreases as fraction of vulnerable applications increases.
- Firewalls and DMZ are not effective when few or most computers are vulnerable.
  - Caveat: Under the assumption that HIPS and NIPS are not effective (full version to be available soon)

Statistical Estimation of Malware Detection Metrics in the Absence of Ground Truth

[w/ P. Du, Z. Sun, H. Chen, and J. Cho. IEEE T-IFS, 2018]

Q: What can we do in the absence of ground truth?

A: We can design statistical estimators to infer useful information, under certain assumptions.

Open Problems: Bridging The Gaps

What we can do now

- Quantify building-block properties
- What can be measured
- No metrics curriculum
- "1 + 1 + 1 = ?" in the current partnership?
- Most security papers offer no metrics
- Ad hoc definitions of metrics
- Uncertainty largely ignored
- No research community

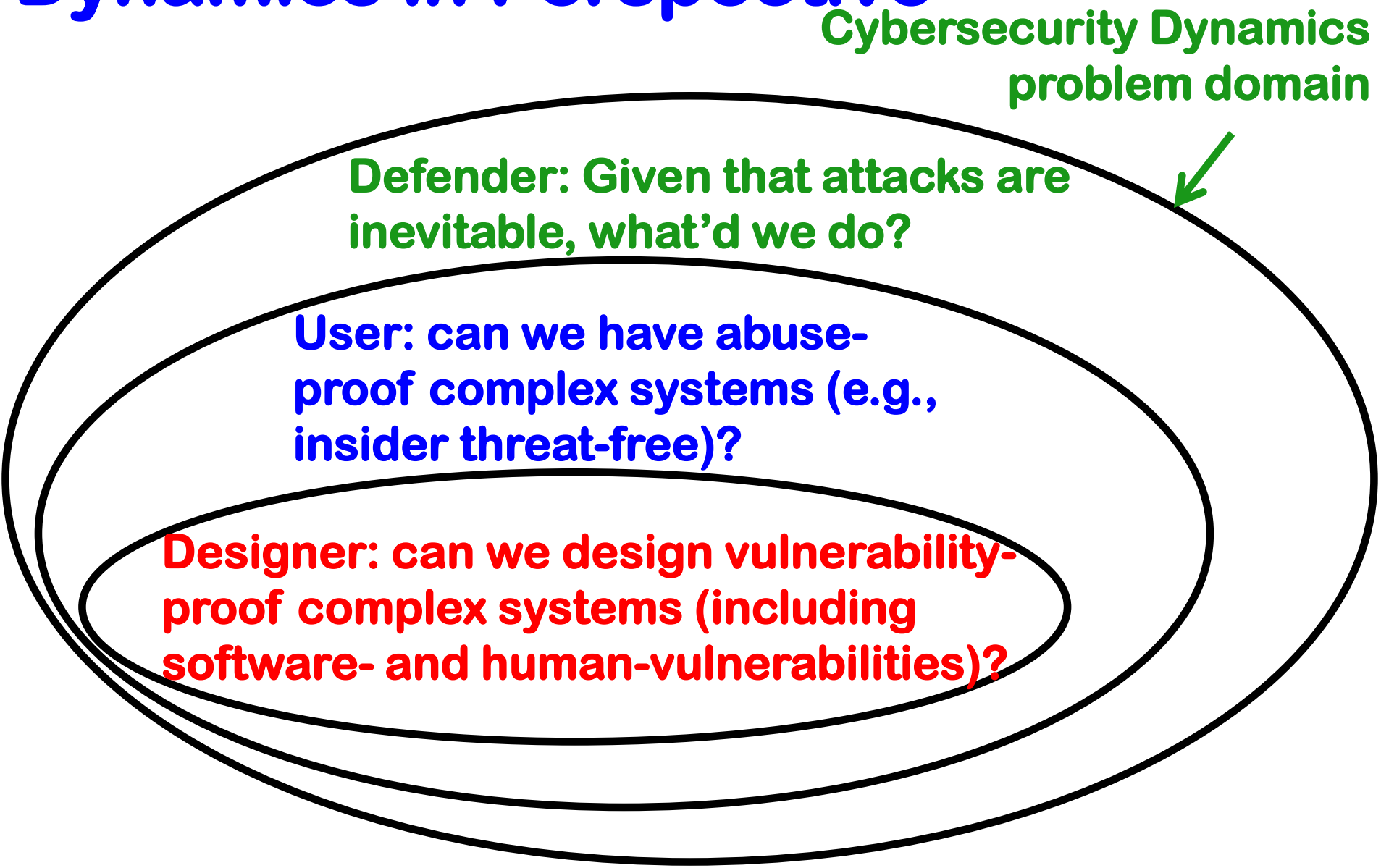
What need to be done

- Quantify holistic system properties
- What must be measured
- Metrics curriculum
- Government & industry & academia: 1+1+1>3
- Each security paper has clearly defined metrics
- Clear understanding of metrics (e.g., additivity?)
- Theory of uncertainty quantification
- A research community

Outline

- The Cybersecurity Dynamics framework: Concept
- The x-axis: First-principle cybersecurity modeling
- The y-axis: Cybersecurity data analytics
- The z-axis: Security metrics
- **Conceptual clarifications**
- Takeaway message

Clarification 0: Cybersecurity Dynamics in Perspective



Clarification 1: Science of Cybersecurity Is Not Applied "X"

The elementary entities of science X obey the laws of science Y, but science X is not "just applied Y."

X	Y
Solid state or many-body physics	Elementary particle physics
Chemistry	Many-body physics
Molecular biology	Chemistry
Cell biology	Molecular biology
...	...
Psychology	Physiology
Social science	Psychology

[P. Anderson. More is different. Science, Vol. 177, No. 4047, August 4, 1972, pp 393-6.]

Clarification 2: Inspirations from Other Disciplines

Cybersecurity Dynamics stands on the shoulders of:

- **Cryptography: A science rooted in a single concept!**
- **<u>Complexity Science</u>**
- **Network Science**
- **Biological Epidemiology (security tries to mimic biological systems, such as Artificial Immune System)**
- **Interacting Particle Systems**
- **Statistical Physics**
- **Macrofoundation in Economics**

But goes far beyond them because of the unique technical barriers discussed above (and recapped below).

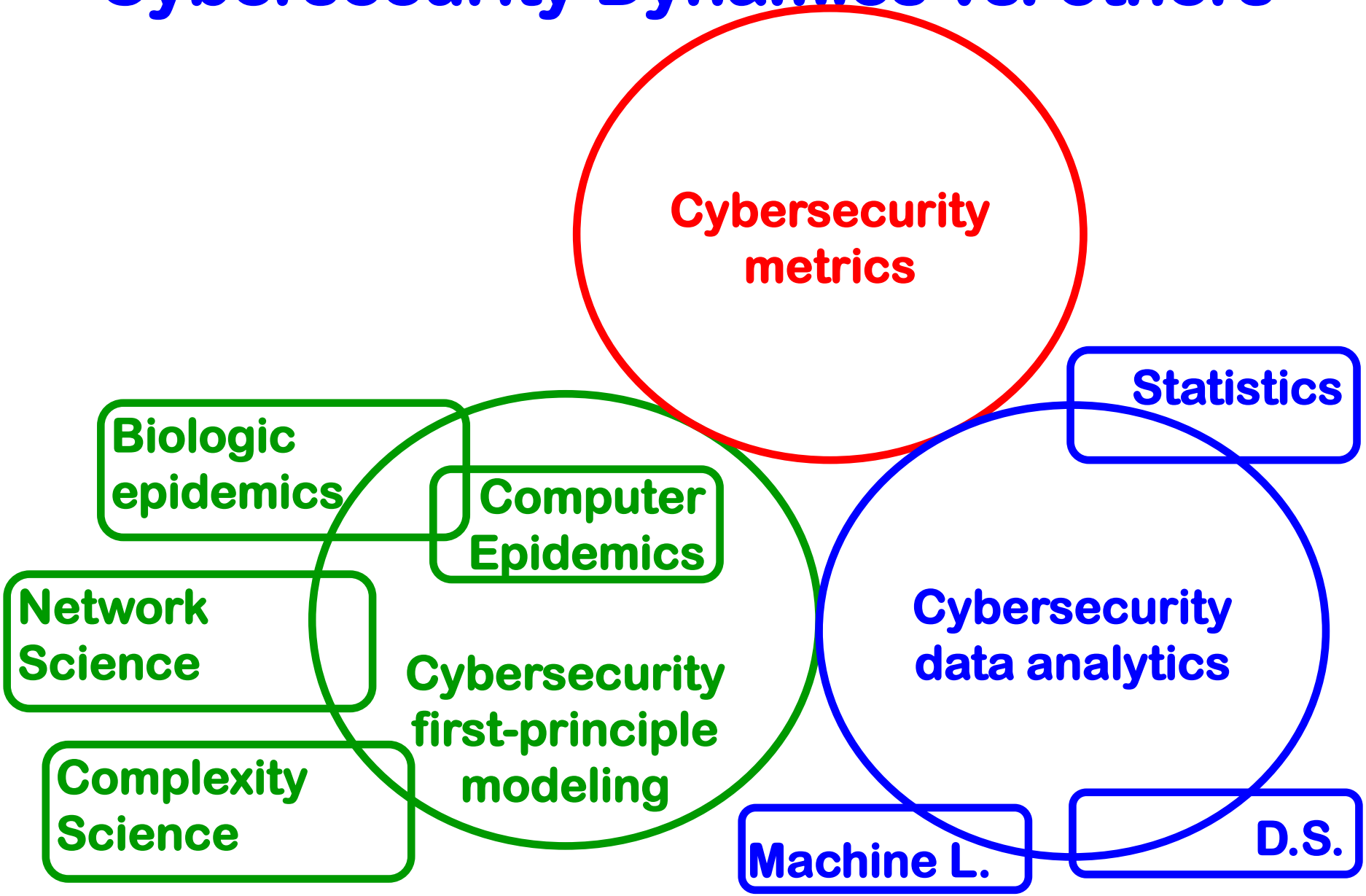
Going Beyond the Inspiration Disciplines

A unique set of inherent technical barriers going beyond the inspiration disciplines mentioned above:

- **Scalability barrier: exponentially many states**
- **Dependence barrier: dependent random variables**
- **Nonlinearity barrier: highly nonlinear systems**
- **Parameter/structural dynamics barrier: dynamic parameters and structures**
- **Transient behavior barrier: knowing equilibrium behavior is not sufficient.**
- **Uncertainty barrier: uncertain/deceptive information**

<u>These barriers are inherent (i.e., cannot be bypassed!)</u>

Clarification 3: Illustration of Cybersecurity Dynamics vs. others



Clarification 4: Degree Distribution-based vs. Matrix-based First-Principle Modeling

- The approach of "Degree Distribution"-based modeling of dynamics over complex networks may be too coarse grained for cybersecurity purposes.
- Instead, we might need matrix-based modeling as described in our first-principle modeling

Outline

- The Cybersecurity Dynamics framework: Concept
- The x-axis: First-principle cybersecurity modeling
- The y-axis: Cybersecurity data analytics
- The z-axis: Security metrics
- How did we get here, and where are we heading for?
- **Takeaway message**

See more publications at www.cs.utsa.edu/~shxu/socs

- Internet Mathematics
- IEEE Trans. on Dependable and Secure Computing
- IEEE Transactions on Network Science and Engineering
- Physical Review E
- IEEE Transactions on Information Forensics & Security
- ACM Trans. On Autonomous and Adaptive Systems
- Technometrics
- Journal of Applied Statistics
- ACM Computing Survey
- HotSoS, NDSS, ACSAC, GameSec, …

A Conjecture: Complexity Science Comes to Rescue Again?

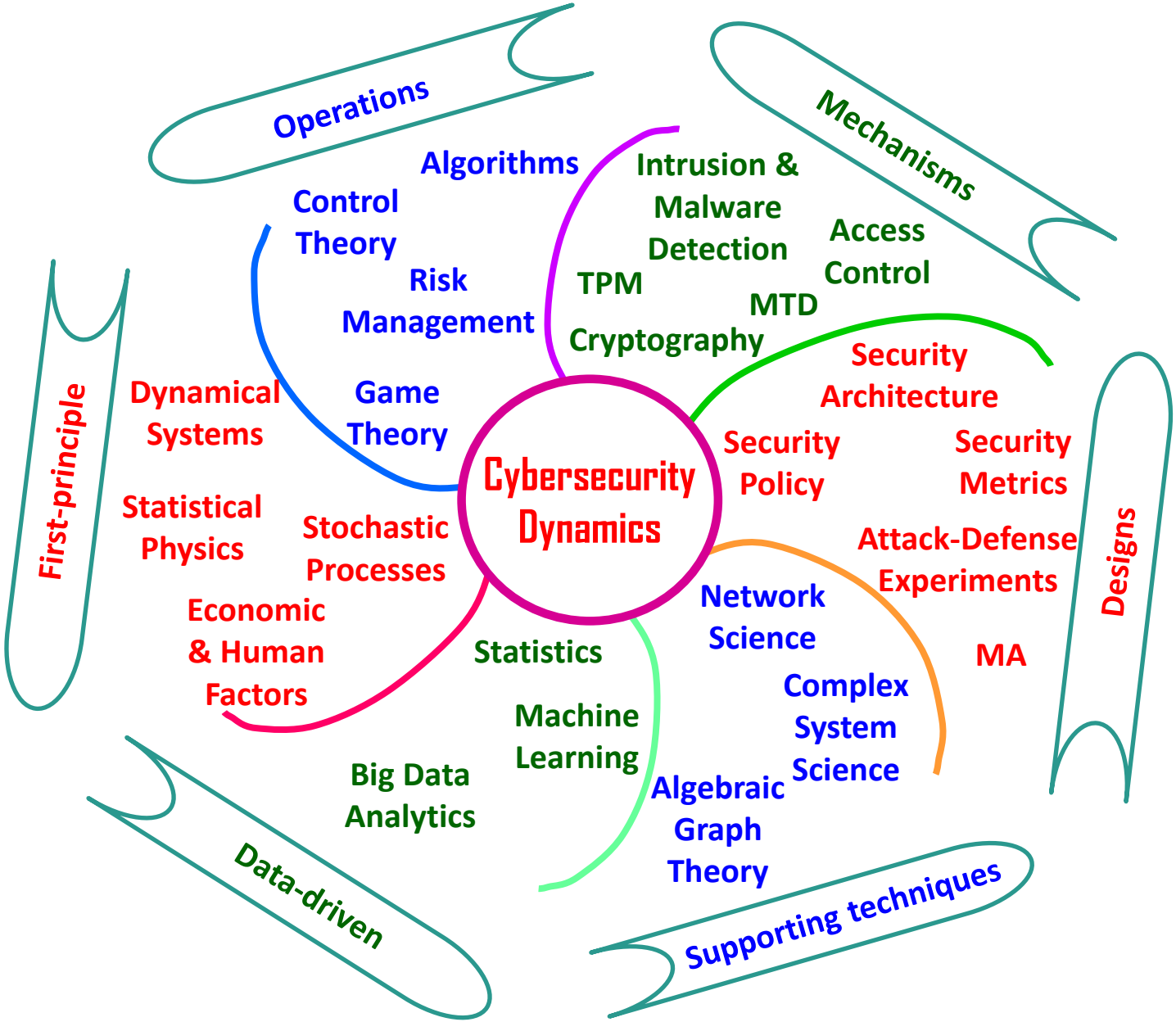
The Science of Cryptography:

- Soul: Security (semantics/concepts)
- Brain: Comp. Complexity Theory (kind of Complexity Science)
- Muscle & Blood: Probability Theory, Number Theory, Abstract Algebra, etc.

Science of Cybersecurity:

- Soul: Security (semantics/concepts)
- Brain: Cybersecurity Dynamics (kind of Complexity Science)
- Muscle & Blood: Security (policies, architectures, mechanisms), Computer Science, Complex Systems, Applied Math, Statistical Physics, Control Theory, Game Theory, Statistics, Algebraic Graph Theory, etc.

A New Body of Knowledge in the Making



Need Community Effort

- The way ahead is difficult, but exciting!
- A new conference: SciSec

  International Conference of Science of Cyber Security

  - 8/12-8/14 2018, Beijing, China
  - Website: http://www.sci-cs.net/
- SciSec'2019 in Nanjing