How Hard Is Cyber-risk Management in IT/OT Systems? A Theory to Classify and Conquer Hardness of Insuring ICSs

RANJAN PAL, Massachusetts Institute of Technology, USA
PEIHAN LIU, Harvard University, USA
TAOAN LU, Carnegie Mellon University, USA
ED HUA, MITRE Corporation, USA

Third-party residual cyber-risk management (RCRM) services (e.g., insurance, re-insurance) are getting increasingly popular (currently, a multi-billion-dollar annual market) with C-suites managing industrial control systems (ICSs) based upon IoT-driven cyber-physical IT and OT technology. Apart from mitigating and diversifying losses from (major) cyber-threats RCRM services positively contribute to improved cyber-security as an added societal benefit. However, it is also well known that RCRM markets (RCRM for ICSs being a mere subset) are relatively nascent and sparse. There is a huge (approximately 10-fold) supply-demand gap in an environment where (a) annual cyber-losses range in trillions of USD, and (b) CRM markets (residual or otherwise) are annually worth only up to 0.25 trillion USD. The main reason for this wide gap is the age-old information asymmetry (IA) bottleneck between the demand and supply sides of the third-party RCRM market, which is significantly amplified in modern cyber-space settings. This setting primarily comprises interdependent and intra-networked ICSs (and/or traditional IT systems) from diverse application sectors inter-networked with each other in a service supply-chain environment. *In this article, we are the first to prove that optimal cyber-risk diversification (integral to RCRM) under IA is computationally intractable, i.e., NP-hard, for such (systemic) inter-networked societies.* Here, the term "optimal diversification" implies the best way a residual and profit-minded cyber-risk manager can form a portfolio of client coverage contracts. We follow this up with the design and analysis of a computational policy that alleviates this intractability challenge for the social good. Here, the social good can be ensured through denser RCRM markets that in principle improve cyber-security. Our work formally establishes (a) the reason why it has been very difficult in practice (without suitable policy intervention) to densify IA-affected RCRM markets despite their high demand in modern CPS/ICS/IoT societies; and (b) the efficacy of our computational policy to mitigate IA issues between the supply and demand sides of an RCRM market in such societies.

CCS Concepts: • **Applied computing** → *Enterprise computing*; • **Computer systems organization** → *Embedded and cyber-physical systems*; • **Computing methodologies** → *Randomized search;*

Additional Key Words and Phrases: Cyber-risk, security, IoT, ICS, IT, OT, information asymmetry, cyber-risk diversification, residual cyber-risk management, insurance, NP-hard, bipartite graph, expander graph

The research is sponsored partly through internal funds obtained from MIT CAMS.
Authors' addresses: R. Pal (corresponding author), Massachusetts Institute of Technology, USA; email: ranjanpal9@gmail.com; P. Liu, Harvard University, USA; email: peihanliu@fas.harvard.edu; T. Lu, Carnegie Mellon University, USA; email: taoanl@andrew.cmu.edu; E. Hua, MITRE Corporation, USA; email: ehua@mitre.org.

2378-962X/2022/12-ART35 $15.00
https://doi.org/10.1145/3568399

ACM Reference format:
Ranjan Pal, Peihan Liu, Taoan Lu, and Ed Hua. 2022. *How Hard Is Cyber-risk Management in IT/OT Systems?* A Theory to Classify and Conquer Hardness of Insuring ICSs. *ACM Trans. Cyber-Phys. Syst.* 6, 4, Article 35 (December 2022), 31 pages.
https://doi.org/10.1145/3568399

1 INTRODUCTION

CPS-driven smart societies (cities) are examples of service-networked ecosystems that are popularly on the rise around the globe, with major cities such as Singapore, Dubai, Barcelona, and Amsterdam being working examples. According to the **Organization for Economic Cooperation and Development (OECD)** [16], these cities are expected to contribute to a multi-trillion dollar economy by the next decade. The proper functioning of such cities is hugely based on the success of interdependent supply-chain relationships carved out from IoT-driven **industrial control systems (ICSs)** in diverse sectors such as automobiles, electronics, energy, aviation, finance, aerospace, and so on. In the CPS age, these relationships are often realized via large-scale OT and IT systemic network linkages that operate via the interplay of hardware (e.g., sensors, actuators, cameras), application software (e.g., Oracle for DBMS support, cloud service software), and IoT/CPS firmware.

A major flip-side to the significant socio-economic benefits ushered in by modern OT and IT systems-driven smart cities (societies) is the much-desired but non-existent strong cyber-security that should complement technology infrastructures and the services (e.g., energy, manufacturing, water, transportation, retail) such societies provide [14]. Corporate surveys conducted by popular cyber-risk management firms (e.g., Advisen, PartnerRe, Deloitte) keep confirming every year (via their annual reports) that most **industrial control systems (ICSs)** around the globe—small, medium, or big—are successfully breached through malicious events that include cyber-extortion (e.g., ransomware), unintentional data disclosures, lost or stolen data, data breaches, unauthorized data collection and disclosure, identity theft, network/website disruption, business email compromise via social engineering, and denial of service. This, despite the rapid growth in the number of cybersecurity technology solutions over the years. The reason for this strange paradox is multi-fold:

(1) There has been a rising trend in the past few years (thanks to multiple major cyber-catastrophes such as Sunburst APT hacks, Mirai DDoS, WannaCry/Petya, NotPetya ransomware attacks, and Sony/Target data breaches hitting society) among organizational boards acknowledging cyber-risk to be a top-five concern for business continuity, reputation, and profitability [24, 60, 66]. However, the proportional time, resources, and effort have not been put in to design effective board-level policies that aptly incentivize employees to behaviorally improve cybersecurity practices and make best use of installed security products.
(2) Cybersecurity technology solutions is a *market for lemons* (a term coined by economist George Akerlof in 1970 [2]). The root of the technology efficacy problem is primarily economic driven by information asymmetry between the parties that prevent technology buyers (e.g., the CISOs and the enterprise team of organizations) from effectively evaluating technology and incentivizing security vendors to sell sub-optimal solutions in the market, which are not as effective as promised and which reduce trust in cybersecurity technology. More specifically, the technology solutions market is too product-congested for buyers to give in enough effort to evaluate and rank the effectiveness of each product—to the extent that the buyers believe that lack of quality is the reason behind too many market products to exist concurrently.
(3) Cybersecurity products are often an outcome of a high-risk "casino economy." In this economy, a fragmented vendor industry is configured to manufacture products they

think (a) **venture capitalists (VCs)** will invest in, (b) that larger companies might want to integrate and make the smaller companies follow suit, and (c) that customers can be convinced to buy. These products, akin to a gamble, might be innovative enough to occasionally help cybersecurity, or it might not, but frankly nobody has a clue. Now iterate this process over 10–15 years and you end with a lot of complexity, layers of obsolete and often non-performing technology that requires ever more scarce human expertise to maintain and keep running. As Ciaran Martin, former head of the UK **National Security Centre (NCSC)** once said, *"In cybersecurity right now, trust doesn't always sell, and good security doesn't always sell and isn't always easy to buy. That's a real problem."*

(4) The proliferation of IoT-driven OT and IT systems has ushered in a new and difficult challenge to technologically manage cyber-risk in modern CPS societies. Billions of IoT devices (currently, tens of billions, and projected by Cisco to be a whopping 125 billion by 2030) are deployed in most industrial sectors and connected as part of intra- and inter-organization networks. Most of these devices are unattended for long periods of time and have poor security postures. This is due to (a) limited computational capabilities—incapable of running sophisticated security tools even if desired; (b) being loaded with weak default passwords [25]; and (c) suffering from human-in-the-loop issues breaching ICS security despite recommended IT-OT gaps. This has made industrial IoT-driven cyber-physical industry systems considerably easy to be breached, as evident from the (catastrophic) Mirai and Stuxnet attacks a few years ago. The denser the IoT penetration, the greater the likelihood of system and systemic cyber-risk, and consequently greater the negative social and economic impact [14, 67].

1.1 The Steady Rise of Residual Cyber-risk Management Products in the ICS Age

Bolstered by our above viewpoints is the fact that technology products are solely not going to lead us to an ideal cybersecurity state in an IT and/or ICS universe. There will inevitably be residual cyber-risk faced by any IT/IoT controlled organization. This has led to CPS-powered ICS organizations in smart cities round the globe increasingly embracing **cyber-risk management (CRM)** solutions that are a mix of both in-house efforts (e.g., via effectively using security vendor products, raising employee awareness and response to cyber-security, self-insurance) and commercial third-party coverage-based **residual cyber-risk management (RCRM)** products (e.g., cyber-insurance) that eliminate residual risk.

Standard Market RCRM Solutions for the CPS-driven ICS Age - One such commercially popular third-party residual coverage product family consists of cyber (re-)insurance solutions that defend businesses and individuals from the financial losses caused due to cyber risk exposure. *More specifically, cyber (re-)insurance business solutions diversify their clients (in the CPS context, organizations reliant on CPS services) into cyber-risk types* (e.g., high, medium, low). *In addition, they cover their first-party costs* (e.g., cyber extortion, cyber forensics, credit monitoring, civil fines and penalties, and privacy notification), *as well as third-party liability* (e.g., electronic media liability, network security and privacy liability). Apart from providing loss coverage as its salient functionality, *cyber-insurance carries with it the promise of improving cybersecurity*—known through principle, as well as verified in theory through multiple research efforts since the early 2000s [3, 8, 12, 41, 47, 49, 50, 52, 53, 62, 78]. From the viewpoint of organizations deploying CPS systems, cyber-insurance propels the former to voluntarily adopt appropriate security controls (thereby improving cyber-security) and subsequently pay lesser premiums.

A Push for Improved RCRM Market Density in the CPS Age - Today, more **IT and CPS-driven organizations (ITOs)** than ever (nearly 80 percent in the USA), specifically those parts of smart cities (Figure 1) governed by **industrial control systems (ICSs)** [67], carry cyber

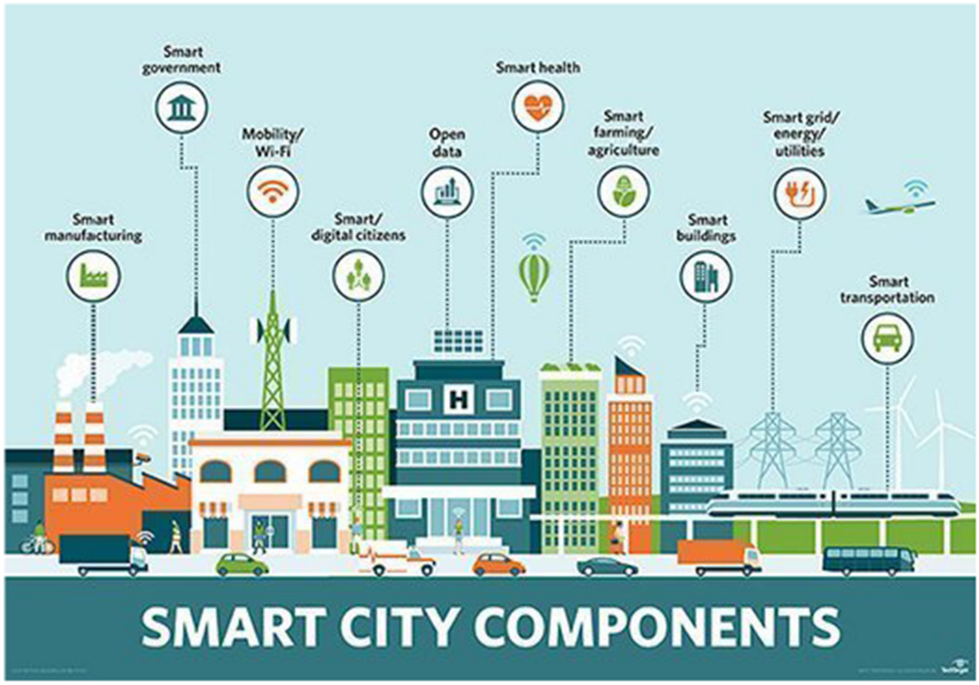

Fig. 1. A collection of diverse application service sectors driving a CPS-reliant smart city (*Source:* Google Images).

insurance-type RCRM solutions. The C-suites of IT organizations and ICSs are increasingly coming to terms with the facts that cyber-risks are inevitable and that cyber-adversaries are always ahead than current cyber-defense practices. Around 55% of IT/ICS firms in North America and Europe buy stand-alone cyber-insurance policies (from at least 200 commercial cyber-insurance vendors just in the USA, as reported by Advisen) in an annual market that is worth approximately 8 billion USD globally (projected to grow to 25 billion USD by 2025). Add to this the current push from the legal and policy front in certain parts of the world to invest in cyber-insurance in CPS societies. For example, in February 2020, the California Assembly introduced a bill to make cyber insurance mandatory to process regulated and protected personal information for all state contractors. The rise in data privacy (a huge concern in the CPS-driven ICS societies) laws, such as the **Personally Identifiable Information (PII)** and the **Health Insurance Portability and Accountability Act (HIPAA)** in the US; the global standard, **Payment Card Industry-Data Security Standard (PCI-DSS)**; and the **European Union's (EU) General Data Protection Regulation (GDPR)** are persuading insurance providers to focus on cyber insurance measures. In February 2020, the **European Insurance and Occupational Pensions Authority (EIOPA)** released cyber underwriting strategies to build a strong cyber insurance market for CPS-driven ICS societies.

1.2 RCRM Market Concerns and Research Motivation

In this section, we first brief the reader on the status quo of existing RCRM markets in CPS-driven ICS societies and the root underlying cause of why such markets are not rocketing (but only growing steadily) despite their high demand from CPS-reliant corporations and organizations. We then showcase our research motivation that is heavily related to the stated underlying cause.

Why Are RCRM Markets Sparse? - Despite a push for RCRM solutions, there is a huge gap between the annual cyber-loss market—worth a few trillion USD—and the CRM (residual or otherwise) market—worth at most a quarter of a trillion USD. This latter market is again subdivided into an ≈200B USD market in vendor products and an ≈10B USD market in cyber-insurance of which the ICS cyber-insurance market is a smaller part). *This gap severely constrains taking leaps in improving cyber-security*, as noted by major security vendors such as Symantec and McAfee in their yearly reports over the years. Parallel to the growth of the security vendor market, it is fairly intuitive to assume that RCRM markets should grow quite fast to reduce the big *loss-CRM* gap and contribute to global cyber-security, alongside being profitable. However, the scale of the Internet (consisting of ≈50 billion connected (IoT) devices), together with the lack of strong regulations/standards related to (a) organizational cyber-breach disclosures and (b) default IoT device security controls, make the RCRM solutions space highly non-transparent between solution buyers (organizations) and the sellers (e.g, cyber (re-)insurance companies). This is the typical ***information asymmetry (IA)*** *problem* characteristic of the traditional insurance industry. In addition, the IA problem is magnified for the ICS ecosystem that is networked and interdependent among various critical diverse sectors. On the one hand, it is externally challenging to robustly estimate intra-ICS cyber-risk emanating from its OT/IT architecture given myriads of ways cyber-vulnerabilities can crop up in a large intra-network. On the other hand, inter-network cyber-risks are equally (if not greater) difficult to estimate in the event of a cyber-breach, given the intricate complexity of supply-chain relationships and C-suite reservations to reveal actual loss details. As a result, in the absence of robust cyber-risk estimation environments, the RCRM industry is growing at a slow and steady pace, instead of a "rocketing" pace as is the need of the hour (see more details in Section 5).

Research Motivation - The basic criteria for designing effective RCRM solution portfolios is *to accurately diversify buyers (i.e., CPS-reliant organizations investing in cyber-insurance) into cyber-risk types (e.g., high, medium, low).* In other words, the most important task of RCRM agencies (e.g., cyber (re-)insurers) is to bundle different organizations with heterogeneous cyber-risk postures into pools in a manner such that only few organizations in the worst case will suffer a cyber-loss at similar times (see Section 2 for more details). This is popularly termed as cyber-risk diversification in the RCRM industry and has become an extremely difficult task to execute for cyber-insurers—especially in an ICS-networked societal setting. The reason for this difficulty is three-fold—namely, the presence of (a) time-space cyber-risk correlations (e.g., zero-day vulnerability in a widely popular OS) [48, 61], (b) imperfect knowledge of organizational security postures/hygiene in the absence of strong cyber-threat information disclosure regulations, and (c) large IT/ICS service supply-chain networks inducing high and uncertain amounts of network security externalities that are costly to internalize for RCRM solution providers, given information and geographical policy constraints [11]. Here, the term "externalities" denote the indirect negative QoS impact (to be covered by a cyber-insurer) on CPS-reliant ICS organizations in a supply chain (each of which are insured by a cyber-insurer) due to a security breach (caused by weak cyber-postures) on some organizations in that supply chain. This has often resulted in *price-QoS-profit* mismatches between the supply and demand sides of the RCRM business—leading to a *market for lemons*, where cyber-insurance market premiums are too high for organizations with a good cyber-posture, and quite low for those with a modest/poor cyber-posture. A thing to observe is that the huge supply-demand gap is steadily decreasing by the year, despite the presence of a lemons RCRM market (thanks to C-suite risk aversion for negative press coverage post cyber-breach events). However, the rate is not fast enough to keep pace with the increase in the annual cyber-loss market. *On a broader note, the above description motivates us to formalize the degree of difficulty of the cyber-risk diversification*

task in ICS societies that is the root cause of the huge supply-demand gap in cyber (re-)insurance markets.

Goal - The primary goal of this research is to *study*, *quantify*, and *mitigate* the extent of computational difficulty of RCRM agencies to effectively diversify CPS-reliant organizations under **information asymmetry (IA)** environments. A proven degree of computational difficulty will act as an official confirmation of the scale of practical challenges (and its subsequent mitigation) that will be faced by RCRM agencies to effectively diversify organizations in CPS societies. In other words, if it is computationally intractable to (a) elicit all possible ways in which a CPS-driven system can be breached and (b) subsequently quantify the resulting cyber-loss impact, then it is impossible for humans to estimate cyber-risk in such systems accurately enough. Here, a "system" is any micro (local)- or macro (societal)-level network carved out of CPS-driven ICSs. This will result in profit-minded, risk-averse, and ambiguity-averse RCRMs (e.g., cyber (re-)insurers) constricting the supply of client-attractive CRM contracts, due to their inability to estimate system cyber-risk accurately. This research will subsequently provide a theoretically verified platform for RCRM solution providers to appropriately under-write cyber-loss coverage contracts for inter-dependent CPS-reliant organizations in ICS-networked societies, and contribute to improved RCRM market density. More specifically, the goal of the proposed theory platform will be to execute the diversification difficulty mitigation task in an approximably optimal fashion—at the same time achieving the mitigation in computationally tractably. *To the best of our knowledge, we are the first to formally characterize the degree of difficulty (in addition to its mitigation) of the cyber-risk diversification process for RCRM agencies.*

1.3 Research Contributions

We make the following research contributions in the article:

- We provide a formal intuition on why the cyber-risk diversification task is computationally hard in general—a unique feature to cyber-settings, when compared to traditional risk settings. The killer relevance of this feature lies in diversifying cyber-risks in ICSs and the interdependent networked societies induced by them. This, due to the myriads of potential cyber-vulnerabilities that open up in ICS systems along with high degrees of associated cyber-vulnerability information asymmetry. In the process, we also explain how the cyber-risk diversification task is different from risk diversification in traditional non-RCRM settings **(see Section 2)**. As a necessary background for the general reader of RCRM, we start Section 2 with a brief background of useful terminologies.
- We model the problem of finding an optimal cyber-risk diversified portfolio of RCRM contracts between the supply and demand sides of a market in the presence of cyber-vulnerability IA as a graph-theoretic exercise. We subsequently prove the problem to be NP-hard through its reduction from the *densest k-subgraph problem* in theoretical computer science **(see Section 3)**. In practice, this implies that the necessary pre-requisite task (dense subgraph detection) for an RCRM solution provider to be part of a commercially profitable CRM business for CPS-driven ICS societies is computationally intractable. To take this point forward, let alone humans, even computers cannot tractably estimate ICS-induced cyber-risks and their impact under IA. This contribution justifies (from the viewpoint of computer science) the current dilemma in the cyber-insurance industry whereby despite its high market demand from organizations in ICS societies, the supply is severely constrained due to IA. This inability of risk managers to estimate ICS cyber-risk accurately enough incurs high opportunity and capital costs for the cyber (re-)insurance companies and does not fulfill the theoretical promise of significant leaps in (ICS) cyber-security (via

cyber-insurance) improvement either. Our contribution in this section has already appeared in a conference version [56] and is re-presented in this article for the purpose of continuity.

- We extend our contribution in Reference [56] to propose the design of a provably optimal *quasi-random* (in a sense approximately random) cyber-risk diversification environment. This environment alleviates the negative impact of the IA problem between the supply and demand sides of an RCRM market in polynomial time. More specifically, the environment is constructed with *bipartite expander graphs* having no dense sub-graphs, unlike in Section 3. Hence, the computationally intractable dense subgraph checking procedure is eliminated **(see Section 4)**. Our proposed construction reflects the notion of a computational policy that we recommend for implementation by regulators (and indeed doable via simple rule adds-ons to the status quo). In doing so, the regulators will contribute to densifying (via computationally enabling diversification portfolio optimization) the market presence of cyber-risk diversifying RCRM solution providers (e.g., cyber-insurers) and significantly improve cyber-security.

1.4 Impact of Our Proposed Research on Cyber-security

In the first place, the outcome of our research directly influences the formation of significantly dense(r) RCRM markets in the presence of cyber-vulnerability IA. In a commercial sense, risk-diversified and denser RCRM markets serving ICS societies will add more capital to the traditional cyber-insurance markets serving the same. As a result, more cyber-insurance companies will be in business having the confidence that in the event of large-enough cyber-risks to cover for, they can rely on cyber re-insurers. In the presence of mitigated IA and loss-averse company C-suites, increasingly more C-suites would embrace in cyber-insurance products. *So, how does this help cyber-security on a technical note?* Our proposed IA-resistant market densifying approach will subsequently promote stronger cyber-security in an ICS society due to existing results in [3, 8, 12, 41, 47, 49, 50, 52, 53, 62, 78] that showcase cyber-insurance improving cyber-security in principle. These results are conditioned on the fact that cyber-insurance is made compulsory—though there will be "proportional" improvement if a high proportion (if not all) of companies invest in cyber-insurance. In practice, the technical improvement in cyber-security is obtained through insured C-suites taking cyber-security more seriously and investing in additional budget to technically and behaviorally protect ICS resources. Examples of such C-suite-driven action items include (a) strengthening the IT/OT security gap, (b) bettering the use of password managers for OT and IT devices, (c) following NIST/MITRE suggestions to improve ICS cyber-posture, (d) training employees to follow "best" networking/security practices while working inside and outside companies, and (e) heavily investing in antivirus and sophisticated firewalls in IT networks. Based on cyber-insurance contract requirements, not investing in such action items will cost the profit-minded C-suite to pay increased premiums. In addition to the social impact of our research, the efforts in this article settle the computational complexity of the general category of cyber-risk diversification problems pertaining to **residual cyber-risk management (RCRM)** for inter-dependent societal CPS networks that involve robust (accurate) cyber-posture characterization. *To the best of knowledge, we are the first to characterize the complexity class of cyber-risk diversification.*

2 BACKGROUND AND INTUITION ON CYBER-RISK DIVERSIFICATION HARDNESS

In this section, we first provide a background for the general reader on economic terminologies related to risk management, along with ways CPS-driven ICSs can be breached. We then provide a formal intuition on why the cyber-risk diversification problem is difficult, even for a computer. *We illustrate our intuition, w.l.o.g., through the example of a cyber re-insurance RCRM product setting.*

We do this for two reasons: (a) re-insurance products have been the de facto choice in the risk-management industry to manage major (catastrophic) risks—increasingly frequent (catastrophic) cyber-risks (e.g., ransomware, aggregate risks) in the IoT age being an apt application scenario, and (b) according to popular insurance giant *SwissRe*, nearly 40% (as of 2019) of global cyber-insurance premiums flow to cyber re-insurers.

2.1 A Basic Background on Cyber-risk Management Terminologies

In this section, we propose basic definitions (for the general reader) of the most import terms related to cyber-risk management.

Cyber-insurance - It is a residual cyber-risk-management tool through which a client (e.g., ICS organizations such as the power grid, water plant) hands over residual cyber-risk to a third-party cyber-risk management agency post a cyber-breach incident (e.g., a zero-day driven ransomware attack). Here, residual cyber-risk is that portion of the post-incident cyber-risk that the client cannot manage through traditional risk-management tools such as vendor products (e.g., anti-virus, firewalls) and self-insurance (keeping aside a monetary organizational budget for post cyber-breach incident response). A popular example of a company selling cyber-insurance products is *Zurich*. In practice, corporate organizations (ICS or otherwise) buy insurance from multiple insurance agencies (e.g., *Colonial Pipeline* buys cyber-insurance from *Lloyd's of London* and *Beazley*).

Cyber-insurance Contract - A cyber-insurance contract (denoted IC in this article—see later) is a cyber-risk management deal signed off between a client and its insurer. The deal fixes the insurance premium and coverage deductible amount (atop the full coverage) for a pre-specified amount of time, along with the conditions under which the contract is applicable. As an example, a popular condition that most insurers around the globe work with is that they would not provide loss coverage in the event of a natural (cyber) catastrophe. Most insurance contracts adhere to the standard *Law of Large Numbers* and the *Central Limit Theorem* in statistics while designing suitable conditions under which to provide client coverage post loss incidents.

Cyber Re-insurance - It is a residual cyber-risk-management tool (denoted RCC in this article—see later) for cyber-insurers. Similar to cyber-insurance, clients pay premiums to a re-insurance company in return for a certain amount coverage drafted in a contract—except that the clients paying premiums to re-insurers are insurance companies themselves. The reason cyber-insurance companies resort to re-insurance firms is to handle large aggregate (first and third party) time and space correlated cyber-risk. An example of such scale of cyber-risk might arise when a zero-day attack affects the root of an IT/IoT-driven supply-chain industry. The cyber-attack first disrupts the business continuity of the root organization (resulting in the organization's first-party losses), and subsequently (and often simultaneously) disrupts business continuity in the supply-chain network (resulting in the organization's third-party losses). The insurance company of the root organization is then considered liable to cover aggregate losses in the entire supply chain. It is in such situations cyber-insurers resort to cyber re-insurers. A popular example of a company selling cyber re-insurance products is *Munich Re*.

Cyber-risk Diversification - It is a process via which cyber (re-)insurers prepare a portfolio of cyber-risk sources (clients) to provide coverage for, such that at any given time only a few sources get activated through cyber-incidents. The process is also termed as *risk aggregation*, and its optimized version is also termed as *portfolio optimization*. Ideally, risk diversification relies on the independence of cyber-risks so the likelihood of multiple cyber-risk sources getting activated is minimized.

Information Asymmetry - The phenomenon of information asymmetry is the central challenge in insurance economics since time immemorial. Simply put, it denoted the non-transparency, i.e., asymmetry, of risk information between the client and its insurance provider. Information

asymmetry usually comprises two sub-challenges: *adverse selection* and *moral hazard*. The adverse selection problem arises when the insurance company does not have perfect knowledge of the risk category (e.g., high/low) of its client. Without a good degree of such knowledge, it is difficult for a profit-minded insurer to design contracts appropriate for each category. Risk-mismatched contracts are termed as *lemons* in the insurance jargon. The problem is exacerbated in the case of cyber-insurance where an absence of strong cyber-vulnerability disclosure regulations makes contract underwriting quite challenging in environments of interdependent and correlated cyber-risk. The moral hazard problem arises when clients behave recklessly post getting insured.
How Can CPS-Driven ICSs Be Breached? - A CPS-driven ICS can be breached at four levels of the system: the perimeter, the network, the workstation, and at the OT devices. At the *perimeter level*, the following un-recommended actions can result in ICS cyber-breaches: (a) not limiting access to an OT network, (b) not placing OT and IT systems behind firewalls and other security protection applications, (c) overly exposing OT devices to the Internet, and (d) sparse ICS monitoring for events indicating attempted unauthorized access. At the *network level*, these actions include (a) not implementing secure access controls, (b) not disabling unused/unneeded communication ports and protocols, (c) not using secure methods for remote access, and (d) not setting up effective IDSs, IPSs, and usage logs to detect cyber-compromises at early stages. Similarly, at the *workstation level*, the actions include (a) not implementing strong password setup and multi-factor authentication mechanisms, (b) not setting up blacklists and whitelists to deny or allow access to communicating entities, and (c) not encouraging and/or inculcating organizational employees to embrace secure workstation habits. Finally, at the *device level*, the following unwanted actions can result in device cyber-breaches: (a) not installing physical controls to help prevent unauthorized device access and (b) not tracking correct operating modes (e.g., PLCs should run in RUN mode) of OT/IoT devices. It is obvious that for all these aforementioned levels, there are innumerable number of ways to execute each action type. An extensive (but not limited to) family of cyber-vulnerabilities applicable to CPS-driven ICS settings can be found in References [1, 5, 18, 35, 43, 64, 68]. Likewise, estimates of organizational cyber-breach impacts due to such cyber-vulnerabilties can be found using IT-centric methodologies as proposed in (but not limited to) References [9, 40, 63, 65, 70, 72].

2.2 The Benefits of Cyber-risk Diversification

The Setting - Consider a single cyber-insurer (an example of A RCRM solution provider) in the CPS age, bearing the responsibility of providing cyber-loss coverage to multiple CPS-reliant organizations and interested in buying a cyber re-insurance policy to cover for catastrophic cyber-events. It wants to cover for aggregate cyber-risks arising from N **individual contracts (ICs)** (a cyber-insurance contract), each sold to organizational clients, via a certain number (say, M) of cyber re-insurance policies. Each of the M policies have in their diversification portfolio, a (overlapping) subset of the N risk sources.

Assume *each* IC carries with it a ***"fairness" quotient* (FQ)** that is w.l.o.g. binary, and is either 0 or 1 with probability $\frac{1}{2}$ independently of all others. More specifically, we assign the FQ to be 0 for an individual contract if the cyber-insurer *significantly* (based on pre-set insurer thresholds that is usually different across insurers) *under* or *over* estimates client cyber-risk during organizational security audit processes (thereby falling under the *adverse selection* trap), and 1 otherwise.[1] To this end, adverse selection in cyber-settings can be of two types: one where a cyber-insurer is not aware of its organization clients' risk types; and one where it is under the influence of **moral hazard (MH)** by its client with the latter being purposefully "careless" about cyber-hygiene. The first type

[1] An individual contract may or may not have deductibles associated with it, but we consider EFQs oblivious of the contract type.

of adverse selection problem is usually an outcome of non-stringent regulations around the globe regarding the necessity of organizations sharing of cyber-threat information. The popular moral hazard problem is usually an outcome of organizational C-Suites not investing enough in CPS/IoT security, with or without them resorting to commercial RCRM services.

In other words, an IC has an FQ of 0 if the premiums associated with that contract are not "fair enough" (outcome of the *adverse selection* effect) from mathematically exact fair premiums that are virtually impossible to derive in practice. Thus, expected fairness for the entire contract bundle is $\frac{N}{2}$. Now suppose that this insurer ex-post over time gathers some "inside" information that an n-sized subset S of the contracts, with probability 1, are MH-driven *lemon contracts* where the policy holders (CPS-reliant organizations), due to loose regulations on cyber-information disclosure, are *purposely* taking advantage of the inevitable adverse selection phenomenon and are paying for an *under-priced* contract. In this case the value of the expected fairness of the entire contract bundle will be $\frac{N-n}{2}$ with a *lemon cost* of $\frac{n}{2}$ incurred by the insurer.

Diversification Reduces Lemon Cost in Principle - In the event of time-correlated cyber-risks, a cyber-insurer might be burdened with a large aggregated cyber-risk post a cyber-breach event. This amount might be beyond the coverage capacity of the cyber-insurer for large client size N. In principle, the cyber-insurance company, to hedge their risk for large N (a specific characteristic in the CPS age), can buy insurance derivatives with a cyber re-insurer to significantly ameliorate the lemon cost [34]. In simple terms, a cyber-insurer will club multiple diverse ICs together into a bundled package and pay for a re-insurance contract (termed as the "derivative" in this case) from another third-party—with the hope that in worst case (catastrophic) scenarios only a "small-enough" number of ICs will demand an aggregate cyber-loss coverage (greater than what individual insurers can cover) at any given time post cyber-breach incidents. Here, a large N is often the reality, as a cyber-insurer could be insuring hundreds of CPS-reliant organizations, if not thousands, in an inter-dependent service network.

In particular, consider the setting where the insurer wants to buy M new **re-insurance claim contracts (RCCs)** from the re-insurer (each akin to a "derivative" in finance jargon [34]), each of them depending on the performance of D bundled underlying **individual contracts (ICs)**, where some of these contracts may overlap in the portfolio of multiple RCCs *(synonymous with claim contracts henceforth). D is the number of "diverse" ICs that a cyber-insurer bundles in a single portfolio, i.e., RCC, before presenting it to a cyber re-insurer.*

A good performance indicates the re-insurer covering less aggregate risk arising from D at any given time instant. *It is usual in practice to have $M \cdot D$ much greater than N to ensure that each IC is packaged in multiple insurance-derivative claim contracts.* Assume each of the M contracts generate an **expected fairness quotient (EFQ)** $\frac{N}{3M}$ to the cyber-insurer as long as the number of ICs that are lemons in an RCC is at most $\frac{D}{2} + t\sqrt{D}$ for some parameter $t = O(\sqrt{\log D})$, and otherwise results in an EFQ of zero. The following result follows:

Proposition 1. *In the best case, if there are no lemon ICs in a single cyber-insurer packaged RCC, then the combined EFQ of M claim contracts, i.e., RCCs, is very close to $\frac{N}{3}$ (ideally, EFQ equals N).*

One can rationalize this result using the **central limit theorem (CLT)** that if the pooling is done randomly (each contract depends on D random ICs), then even if there are n lemon ICs, the value is still $\frac{N}{3} - o(n)$, no matter where these lemon ICs are. More specifically, according to the CLT, the total number of lemon ICs may be assumed to be distributed like a Gaussian. Thus, so long as the fraction of lemon classes is much smaller than the safety margin of t standard deviations, the probability for a single claim contract containing many ICs to generate a significantly low EFQ is tiny. Thus, we clearly see that insurance derivatives do indeed help significantly to reduce the lemon cost from $O(n)$ to $o(n)$.

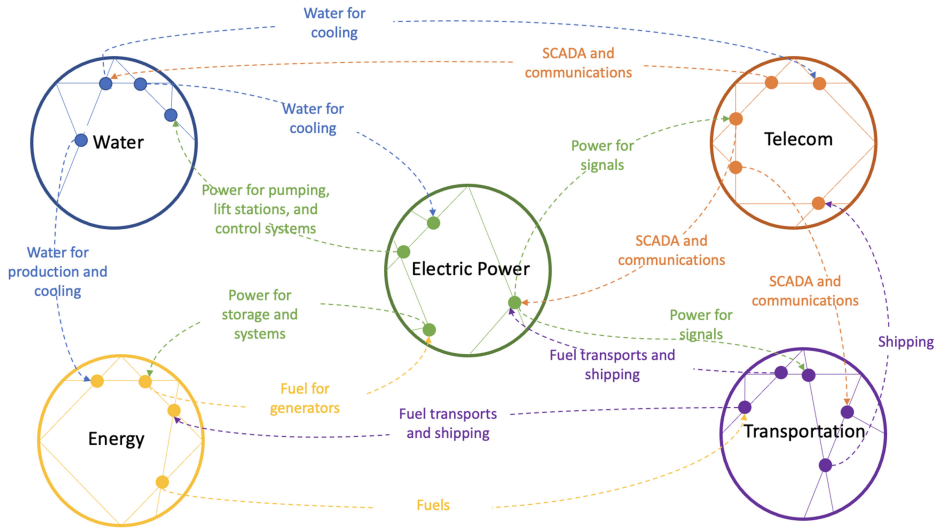

Fig. 2. A representation of example service dependencies in an inter-dependent CPS-reliant service network. Note that each service sector represents multiple CPS-reliant organizational instances, and the security guarantees for each dependency functionality for each organizational instance is a function of multiple hardware/software/behavioral factors characterized over an underlying physical/logical communication network.

Low Lemon Costs Imply Better Cyber-security - Lower lemon costs imply reduced impact of adverse selection on a cyber re-insurer. This effect inductively trickles down to cyber-insurers while selling individual ICs to end-users, which eventually leads to security-improving ICs circulating in a dense end-user market as proven through multiple studies [41, 45, 50, 52, 53, 66].

2.3 Why Is Reducing Lemon Cost Difficult in Inter-dependent CPS Societies?

It must be noted upfront that the risk diversification environment in traditional non-cyber settings is very different than that in cyber-settings. For the latter, in the absence of (a) strong cyber-threat information disclosure regulations circulating around the globe and (b) proper standards on the default security settings preset in manufactured IoT devices used in CPS systems, it is virtually impossible for cyber-insurers (let alone cyber re-insurers) designing ICs to get a clear enough picture of the cyber-posture of all their client organizations. This picture is far clearer in traditional non-cyber environments. Here, information disclosure challenges and its impact include everything "under the sun," including information on (a) insider attacks and (b) voluntary careless attitude towards cyber-security driven by moral hazard. Add to this the added networked-complexity where a cyber-insurer can be an insurer for thousands of organizations (small, medium, big) inter-dependent on each others' (software-run) services. As an example, we can easily infer from Figure 2 that multiple sectors in an inter-dependent CPS-reliant service network might have a large number of organizational instances (e.g., water, telecom), and each instance will further be embedded in an underlying service network with intricate service dependencies. Furthermore, each dependency might involve a suite of software communications over a physical/logical communication network—their security dependent on a large number of hardware/software/behavioral cyber-posture variables. Therefore, it is practically infeasible for cyber-insurance firms to have accurate cyber-posture information even for all its clients embedded in an inter-dependent service network.

Even in the usual case, when cyber-insurance firms (but not the re-insurance firms) might have "inside" information (through properly screened audits) and/or robust estimates of cyber-postures of multiple of its (potentially large number of) clients, the cyber-insurance firm has no incentive to do the pooling activity (see Section 2.1) completely randomly, because it knows S—the observed set of lemon ICs. A randomly pooled RCC takes away the power from these insurers to pack in more lemon ICs into the RCC at its disadvantage. Simple calculations suggest that the optimal strategy for a (adversarial) cyber-insurer is to pick some m of the RCCs and make sure that the lemon ICs are over-represented in them—to an extent about the scale of $\sqrt{D}$ that is just enough to skew the probability that the claim contracts will in total not result in a zero EFQ for the re-insurer. This is rationalized by the fact that, since ICs contained in the same RCC come from different organizations, the yields of non-lemon ICs are uniformly i.i.d., so the expected number of zero-valued FQs among D non-lemon ICs is $\frac{D}{2}$ with variance D. The following result characterizes the lemon cost incurred for a re-insurer post verifying the over-representation of lemon ICs in RCCs.

Proposition 2. *The lemon cost incurred by a perfectly rational, i.e., computationally unbounded (ideal), cyber re-insurer post verifying the over-representation of lemon ICs in RCCs is at most $o(n)$ (indicating a low lemon cost), whereas a boundedly rational (practically realistic) re-insurer verifying the same incurs a lemon cost lower bounded by $O(n)$ (indicating a high lemon cost).*

As a justification of this result, *a fully rational* cyber re-insurer, i.e., the "buyer" (seller) of claim requests (contracts) from the cyber-insurer, can enumerate over all possible n-sized subsets of $[N]$ to verify (as part of its cyber-risk diversification task towards re-insuring ICs) that none of the lemon ICs are over-represented by an adversary,[2] thereby upper bounding a lemon cost of $o(n)$. **An optimal reduction of this lemon cost is equivalent to the optimal cyber-risk diversification problem** and is achieved in the setting of a completely random pooling of ICs. *However, in real-life, the cyber re-insurer is computationally bounded, and hence this enumeration is infeasible for large-sized* $[N]$, as in IoT societies. Put in another way, the cyber-insurer can "plant" a set S of over-represented lemon ICs in claim contracts in a way such that the resulting pooling will be computationally indistinguishable from a random pooling. Consequently, the lemon cost for such cyber re-insurers can be much larger than $O(n)$ in the worst case—*thereby nullifying the vision that introducing insurance derivatives in cyber-settings will mitigate the lemon cost. In practice, contrary to logic, they will instead amplify it.*

2.4 Will Tranching Alleviate Computational Barriers?

In seminal papers, DeMarzo [20] and DeMarzo and Duffie [19] introduced the concept of *tranching* (forming risk layers based on the degree of risk) for non-cyber risk diversification environments that takes advantage of signalling in economic theory, and for our setting, promises to mitigate (moral hazard-driven) IA between a cyber re-insurance buyer and its seller. More specifically, DeMarzo and DeMarzo et al. show (when adapted to our setting) that it is optimal for the cyber-insurer to first bundle ICs and then tranche them in a *single claim contract.* The cyber-insurer can offer the re-insurer the less riskier senior tranche to provide coverage for, and can retain the comparatively more risky junior tranche. The proportion sold versus retained acts as a signalling[3] mechanism to the re-insurer on the quality of ICs to be re-insured; leads to better re-insurance pricing of ICs; and the lemon costs significantly diminish for the re-insurer.

[2]The adversary takes various forms, depending on the type of RCRM contract. It is usually a cyber-insurer for re-insurance type contracts and mother nature for traditional cyber-insurance contracts.

[3]The cyber-insurer knows the identity of lemons, but re-insurance sellers only know the prior distribution of lemon ICs.

However, there is a catch—*in our CPS society setting, the cyber-insurer is offering M (likely large) RCCs instead of a single one.* In contrast, the setting in References [20] and [19] assume that all ICs are packed into *one* RCC with multiple tranches. DeMarzo's analysis has no obvious extension to the multiple RCC case, because the potentially renewed signalling mechanism is far more complex (because N is too large in CPS societies to fit in one RCC) than in the case of a single claim contract, where all ICs have to be bundled into a single pool. In other words, DeMarzo's lemon cost amelioration theory is computationally feasible to be adopted by a boundedly rational cyber re-insurer only if $M = 1$. When $M > 1$, for a perfectly rational re-insurer capable of exponential time computations, lemon costs do get ameliorated by M RCCs. Precisely, the best option for the cyber-insurer (adversarial or otherwise) in this case is to randomly distribute ICs into M equal sized pools and define the less riskier senior tranche identically in all of them. In doing so, a cyber-insurer's signal to the re-insurance policy seller consists of the partition of ICs into pools and the threshold that defines the senior tranche. *Although, for a perfectly rational re-insurer, this signal turns out to contain enough information to design profitable re-insurance contracts to manage aggregate cyber-risk, it is computationally intractable for a practical boundedly rational cyber re-insurer to decipher this signal in general for $M > 1$.*

3 FORMAL ANALYSIS OF CYBER-RISK DIVERSIFICATION HARDNESS

In Section 2, we stated the difference between traditional risk diversification and cyber-risk diversification and the subsequent unique challenges (see Section 2.2) networked CPS-driven ICS societies bring to the latter task. Following the setting in Section 2, we formally verify in this section our intuition that deploying optimal cyber-risk diversifying re-insurance claim contracts (those mitigating the effect of IA between the insurer and the insured) in large-scale cyber-settings such as ICS societies will be computationally costly for a cyber re-insurer. In other words, we prove that the lemon cost for a cyber re-insurer to underwrite optimal re-insurance contracts in CPS societies will be amplified, even in the presence of derivative-centric signaling mechanisms proposed by DeMarzo et al., which have been practically successful in reducing the negative impact of IA in non-traditional non-cyber insurance settings (see Section 2.3). *This section has previously appeared in Reference [56]. We reproduce for the sake of continuity, much of the section in Reference [56] to connect with our contribution (novel to the extended version of our own work in Reference [56]) in Section 4.*

3.1 Proving the NP-hard Nature of Cyber-risk Diversification

We first describe the densest k-sub-graph problem popular in theoretical computer science and state its computational hardness. We then showcase how the optimal cyber-risk diversification application task en route to re-insurance policy formation, and in the presence of IA, is equivalent to finding a k-densest subgraph in a graph.

Simply put, we will argue that detecting the presence/absence of a dense subgraph is *necessary* for a cyber re-insurer to profitably cover diversified cyber-risk portfolios of ICs (put forward by cyber-insurers) in CPS-driven ICS societies. We start with a mathematical formulation of the lemon cost.

Formulating Lemon Cost - Consider a cyber-insurer claim contract (akin an insurance derivative) with performance metric FQ defined on N **individual contract (IC)** inputs. Given the input distribution X over $\{0, 1\}^N$, and $n \leq N$, we denote the lemon cost of FQ comprising n lemon ICs as $\Delta(n) = \Delta_{FQ,X}(n)$ and define it as

$$\Delta(n) = \mathbb{E}[FQ(X)] - \min_{S \subseteq [N], |S|=n} \mathbb{E}[FQ(X)|X_i = 0;\ \forall i \in S],$$

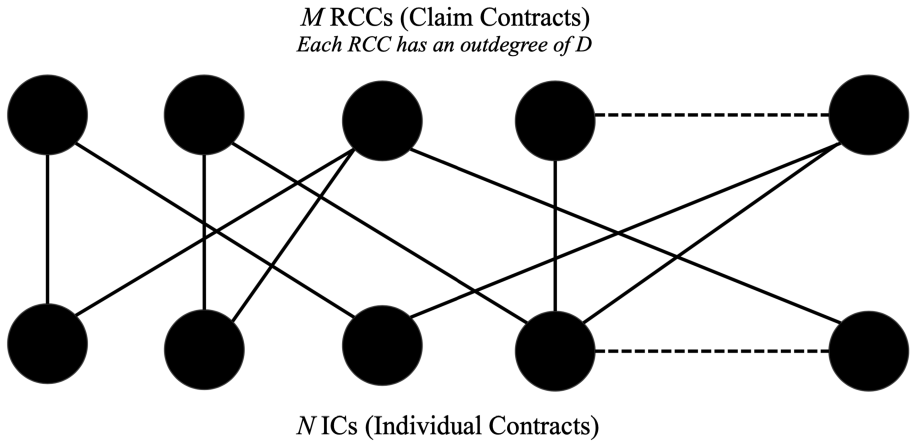

Fig. 3. Illustrating an example (M, N, D) bipartite graph for the Densest Subgraph Problem. Here, each individual IC is a contract between a CPS-reliant organization and its cyber-insurer. Each RCC is a contract of D bundled ICs between a cyber-insurer and its re-insurer. The latter aims to detect a n-dense subgraph of $n < N$ lemon ICs.

where the min operator takes into account all possible ways in which the cyber-insurer could "position" the lemon ICs among the N total ICs while designing its RCC. This lemon cost captures the market inefficiency introduced by lemon ICs.

The Densest Subgraph Problem - Consider an (M, N, D) bipartite graph with M vertices on the "top" side representing claim contracts by the cyber-insurer and N vertices on the "bottom" side representing the ICs owned by the cyber-insurer (see Figure 3). To illustrate Figure 3 more clearly, each cyber-insurance contract, i.e., IC, represented by the bottom nodes of the bipartite graph is written by a cyber-insurer. In a similar fashion each cyber re-insurance contract, i.e., RCC, represented by the top nodes of the bipartite graph is written by a re-insurer. Each RCC can package multiple IC contracts into a single RCC, and each IC can be part of several RCCs. Let each RCC have an outdegree of D indicating the number of ICs that are packaged with a claim contract. We say that such an (M, N, D) graph G contains an (m, n, d) subgraph H, if one can identify m top vertices, i.e., claim contracts, and n bottom vertices, i.e., ICs, of G with the vertices of H in a way that all of the edges of H will be present in G.

Now fix the parameters in the tuple (M, N, D, m, n, d). The *densest subgraph decision problem* [38] for these parameters is to distinguish between the two distributions $\mathcal{R}$ and $\mathcal{P}$ on (M, N, D) graphs, where (a) $\mathcal{R}$ results from choosing for every claim contract, i.e., a single RCC (top vertex), D random individual contract neighbors on the bottom, (b) $\mathcal{P}$ results by first choosing $S \subset [N]$ and $T \subseteq [M]$ such that $|S| = n$, $|T| = m$, and then choosing D random IC neighbors for every vertex outside of T, and $D - d$ random neighbors for every vertex in T, and a choice of an additional d random neighbors in S for every vertex in T.

NP-hardness of the Densest Subgraph Problem - The computational intractability, i.e., NP-hardness, of the densest k-subgraph ($k = n$ in this case) problem was proved (via reduction from k-CLIQUE [69]) in References [4, 7, 22], where it is formally stated that given (M, N, D, m, n, d) such that $N = o(MD)$ and $(\frac{md^2}{n})^2 = o(\frac{MD^2}{N})$, there is no (a) $\epsilon > 0$ and (b) a poly-time algorithm that distinguishes between $\mathcal{R}$ and $\mathcal{P}$ with advantage ϵ. To provide additional relevant details on the NP-hard **densest k-subgraph (DkS)** problem, the DkS problem is known to be NP-hard for graphs whose maximum degree is equal to three [23]. The DkS problem (also know as the heaviest

unweighted subgraph problem [39], k-cluster problem [15], or the k-cardinality subgraph problem [13]) is NP-hard even for very restricted classes of graphs, such as bipartite and chordal graphs [15] or planar graphs [36]. However, it is trivial (not NP-complete) on trees. The DkS problem is also poly-time solvable on graphs whose maximum degree is equal to two, as well as on cographs, split graphs, and k-trees [15]. In other words, even if a graph is bipartite, as long as the maximum degree of this graph is greater than two, the DkS problem is NP-hard.

We will now argue that the setting—incorporating a non-adversarial cyber-insurer who is "blinded" by IA on accurate-enough cyber-posture information of *all* its clients—*cannot result in dense subgraphs.* This setting takes away the power from the cyber-insurer to disproportionately pack in lemons in an RCC, paving the platform for profitable aggregate cyber-risk management services such as cyber re-insurance. However, the latter setting is quite unrealistic in practice and usually cyber-insurers have accurate-enough cyber-posture information on certain fraction of clients. This setting does result in dense subgraphs and provides the cyber-insurers with the power to disproportionately and adversely pack in lemons in an RCC, to the disadvantage of a cyber re-insurer. Thus, the latter wants to necessarily detect the presence/absence of such subgraphs.

Non-adversarial Insurer Implies No Dense Subgraphs - Consider N ICs that are distributed i.i.d. each of which has a probability 1/2 bearing an FQ of zero and probability 1/2 of bearing an FQ of 1. The fact that each IC is either a lemon or non-lemon with equal probability is not considered inside information to be used for adversarial gains. It is usual in practice for cyber-insurers to assume that some ICs will be lemons in view of traditional IA issues between itself and its clients. In our setting the cyber-insurer requests M claim contracts (akin to insurance derivatives) from a cyber re-insurer, where the expected FQ of each claim contract is based on the D ICs forming the portfolio of the contract. Assume a threshold value $b < \frac{B}{2}$, such that each claim contract has EFQ of 0 if more than $\frac{D+b}{2}$ of the ICs contained in it are lemons, and an EFQ value of $V = \frac{D-b}{2D}\frac{N}{M}$, otherwise. *Clearly, the* (M, N, D) *bipartite graph in this setting does not have dense* k*-subgraphs, due to the fact that each IC is a lemon or otherwise independently, with probability* $\frac{1}{2}$. This is similar to a binary valuation setting and represents claim contracts of the simplest type. The case for the non-binary setting is more granular (e.g., more lemon varieties) but equivalent, given assumptions. Given each claim contract depends on D independent ICs, the number of lemon ICs for each RCC closely follows a Gaussian distribution as D gets larger.

Adversary Insurer Implies Existence of Dense Subgraphs - In the case when the cyber-insurer has inside information with probability 1 that there are n lemon ICs, an adversarial cyber-insurer can carefully design the (M, N, D) graph to increase its return from an RCC policy. Note that though each RCC packages D ICs for its re-insurance portfolio, to substantially increase its return on the claim, it suffices for the "adversarial" cyber-insurer to fix about $\sigma \simeq \sqrt{D}$ of the underlying ICs. More precisely, if t of the ICs contained in an RCC are lemons, then the expected number of lemon ICs in the RCC is $D + t$, while the standard deviation becomes $\frac{\sqrt{D}}{2}$. Thus, the probability of this claim contract resulting in an EFQ of 0 is $\frac{\Phi(t-b)}{2\sigma}$, which starts getting larger as t increases. This further implies that the difference in return between an IC pool of zero lemons versus that of t lemons is about $V \cdot \frac{\Phi(t-b)}{2\sigma}$. This gives rise to the following important result:

Theorem 3. *The optimal cyber-risk diversification problem in RCRM for CPS societies under information asymmetry is NP-hard. In other words, it is computationally infeasible (let alone humanly infeasible) for a boundedly rational RCRM solution provider such as a re-insurer selling M RCCs to guarantee that lemon ICs are not over-represented by cyber-insurers in each RCC.*

Proof. Say the cyber-insurer allocates t_i of the lemon ICs to the ith RCC. Given that each of n lemon ICs are contained in $\frac{MD}{N}$ claim contracts, we have $\sum_{i=1}^{M} t_i = \frac{nMD}{N}$, which results in a lemon

cost of $V \cdot \sum_{i=1}^{M} \frac{\Phi(t-b)}{2\sigma}$ to the cyber re-insurer. The function $\frac{\Phi(t-b)}{2\sigma}$ being concave for $t < b$, and convex otherwise, the optimal strategy for the "adversarial" cyber-insurer will have t_i's to be either 0 or $k'\sqrt{D}$, for a small constant k'. Put in other words, the lemon cost for the cyber re-insurer is maximized by choosing some m RCCs, letting each of them have at least $d = k'\sqrt{D}$ edges from the set of lemon ICs. *In* (M, N, D) *bipartite graph, this property corresponds to an* ***(m, n, d) dense subgraph****—representing a set of insurer-manipulated RCCs and a set of lemon ICs that have more edges between them than expected.* However, since the densest subgraph problem is NP-hard, unless P = NP, it is intractable for a computer, forget a boundedly rational cyber re-insurer, to decide whether an (m, n, d) dense subgraph is embedded in an (M, N, D) graph. □

Implication to Cyber-risk Management - The theorem implies that diversification-necessitating optimal cyber-risk management in CPS-driven ICS societies is an NP-hard problem, i.e., a very difficult proposition in practice. Cyber-insurers might in all likelihood have partially complete cyber-posture information about its client organizations embedded in an inter-dependent service network and will use this to their advantage to disproportionately pack in lemon ICs into a diversified RCC portfolio. The cyber re-insurer gauging this fact will not be keen to sell coverage for these RCCs and will result in a sparse (non-dense) RCRM market, as is the status quo in practice.

3.2 Deriving Lemon Costs for a Boundedly Rational Cyber Re-insurer

Thus far, we have established the NP-hardness of the optimal cyber-risk diversification process for a *worst-case* RCC-IC bipartite graph instance. In this section, we *first* show that there indeed exists a set of parameters (m, n, d) for which there is a very small likelihood of a dense subgraph existing in an (M, N, D) graph (or to put in other words, a condition under which there does not exist a dense subgraph). This condition (or its converse) can be verified in poly-time by a boundedly-rational (in terms of computation) cyber (re-)insurer—the main question is at what cost? That brings us to the *second* result on deriving the costs to do such a verification. We now have our first result stating the condition for the *non-existence of a dense subgraph* that a cyber insurer can embed in IC pools—*a necessary condition to prevent lemon costs getting amplified for a cyber re-insurer.*

Theorem 4. *Consider an insured CPS society with a cyber-insurer requesting M re-insurance claim contract policies (acting as insurance derivatives) from a cyber re-insurer, each of them having D ICs (between the insurer and its CPS-reliant client) in its portfolio, where these diversifiable D ICs are sampled over a universe of N* ***individual contracts (ICs)*** *and are potentially overlapping on multiple RCCs. With a high probability, there exists no dense subgraph* (m, n, d) *in a random* (M, N, D) *graph when* $n \ll md$, $\frac{dN}{Dn} > (N + M)^{\epsilon}$ *for some constant* ϵ.

Proof. Let $X_{i,j}$ be the random variable indicating the existence of an edge between IC i and claim contract (insurance derivative) j. For simplicity, we assume the $X_{i,j}$s are independent random variables that with probability $\frac{D}{N}$ take up a value of 1. In the real-world, the $X_{i,j}$s are negatively associated statistically—as a result, our independence-induced results in this article will more likely hold true in realistic settings. Based on proof methods introduced in Reference [6], pick any set A of n ICs and a set B of m claim contracts (a proxy for insurance derivatives) of size m. The probability there exists at least md edges between A and B is the probability of $X = \sum_{i \in A, j \in B} X_{i,j} \geq md$. Given the independence assumption on the $X_{i,j}$s and the fact that sum X has expectation $\mu = \frac{mnD}{N}$, using the Chernoff bound, we have $\Pr[X > (1+\delta)\mu] \leq (\frac{e^{\delta}}{(1+\delta)^{1+\delta}})^{\mu}$, where $\mu = \frac{mnD}{N}$, $1 + \delta = \frac{md}{\mu} = \frac{dN}{Dn}$, the probability is at most $\Pr[X > (1+\delta)\mu] \leq (\frac{e^{\delta}}{(1+\delta)^{1+\delta}})^{\mu} \leq (N+M)^{-\epsilon md}$. The number of such sets

is $\binom{N}{n}\binom{M}{m} \leq (N+M)^{n+m}$. As a result, via the use of the union bound, the probability that there exists a pair of sets that has at least md edges between them is at most $(N+M)^{n+m-md}$, which is much smaller than 1 by the assumption that $n \ll md$. Thus, a random graph will not have embedded dense subgraphs, in high likelihood. □

Implication to Cyber-risk Management - The theorem states that there exists a specific set of parameters m, n, d and an associated relation between them for which the "adversarial" cyber-insurer cannot embed any dense subgraph in an IC pool. In other words, *there is a possibility for a boundedly rational and diversifying cyber re-insurer to verify in the affirmative that lemon ICs are not over-represented in any IC pool (thereby mitigating negative effects of IA), but the question is at what cost? (see Theorem 5).* On an orthogonal note, a thing worth mentioning is that in the real-world, the cyber-insurer tries to put together a diversified portfolio of ICs that are sufficiently independent, and we capture this fact using a random graph—though, the cyber-insurers, in practice, need not construct a random graph in the industry. Having established the possibility of the non-existence of a dense subgraph in a random (M, N, D) graph, the big question then becomes: *how much lemon cost will a boundedly rational cyber re-insurer accrue to verify the non-existence of a dense subgraph?* We have the following theorem in this regard:

Theorem 5. *Consider an insured CPS society with a cyber-insurer requesting M re-insurance claim contract policies (acting as insurance derivatives) from a cyber re-insurer, each of them having D ICs (between the insurer and its CPS-reliant client) in its portfolio, where these diversifiable D ICs are sampled over a universe of N* ***individual contracts (ICs)*** *and are potentially overlapping on multiple RCCs. When $d - b > 3\sqrt{D}$, for a given derivative threshold b, and $n/N \ll d/D$, an (m, n, d) subgraph will generate an extra lemon cost that is at least $(1 - 2p - o(1))mV \approx n\sqrt{N/M}$.*

Proof. Given a cyber-insurer's claim contract (insurance derivative) manipulation activity, for each such contract it manipulates, let Y be the number of lemon ICs. We know that there are d ICs that come from the set of lemon ICs, each of these d ICs will always have an FQ of 0, and the expectation of Y is $\mathrm{E}[Y] = \frac{D+d}{2}$. The Gaussian approximation holds for large enough D, leading to $\Pr[Y \geq \frac{D+b}{2}] = 1 - p$. In line with proof insights introduced in Reference [6], for claim contracts that the cyber-insurer does not manipulate with, we assume the expected number of lemon ICs is x,, which then leads x to satisfy $m \cdot \frac{D+d}{2} + (M-m) \cdot x = \frac{N+n}{2N} \cdot \frac{MD}{N}$. This relation holds true as both the LHS, and the RHS are quantities reflecting the expected number of lemon ICs. Given that $x \geq \frac{D}{2} + \frac{n}{2N}\frac{md}{2M-2m}$, the probability that the number of lemon ICs is more than $\frac{D+b}{2}$ is at least $\Phi(-3 - \frac{md}{2(M-m)D}) = p - \Phi'(-3)\frac{md}{2(M-m)D} = p - O(\frac{md}{2(M-m)D})$. The expected count of non-manipulated claim contracts that gives no return is at least $p(M-m) - O(\frac{md}{2D}) = p(M-m) - o(m) = pM + (1-2p)m - o(m)$. This quantity is $(1-2p-o(1))$ smaller than the expectation without dense subgraphs. Hence, the extra lemon cost incurred by the re-insurer is $(1 - 2p - o(1))mV$. □

Implication to Cyber-risk Management - In the worst adversarial case when (a) $M \ll N \ll M\sqrt{D}$ (the number of IC portfolios per RCC are excessively large), (b) $m = \Theta(n\sqrt{\frac{M}{N}})$, in which case a random (M, N, D) graph with an embedded (m, n, d) subgraph remains indistinguishable from a random graph under the densest subgraph assumption, and (c) $b = 2\sigma\sqrt{\log\frac{MD}{N}}$, the theorem implies that a boundedly rational cyber re-insurer will incur a lemon cost of $n\frac{N}{M} = \omega(n)$ to verify the non-existence of a dense subgraph in a random (M, N, D) graph, whereas a perfectly rational cyber-insurer will only incur a lemon cost of $n\frac{N}{2M\sqrt{D}} = o(n)$. *It is evident that the lemon cost for a boundedly rational cyber-insurer is orders of magnitude higher when compared to a perfectly*

rational one. Note that these orders of magnitude increases are usually in worst-case scenarios. However, the risk-averse mindset of profit-minded cyber (re-)insurers psychologically weighs on them, and the worst-case settings get perceived as average case settings. **This (perceived) high cost prevents densification of RCRM businesses in CPS societies.**

4 POLICY MITIGATING RISK DIVERSIFICATION HARDNESS IN CPS SOCIETIES

Cyber-risk diversification hardness is dampening to the success and scale of futuristic much-needed RCRM markets serving CPS societies. In this section, we design a computational policy that mitigates the diversification hardness in CPS societies encountered in Section 3. We start with providing an intuition and working rationale of our policy methodology. We then describe the (graph-theoretic) modeling framework for our policy solution. Finally, we analyze our model to prove our hardness mitigation claims.

4.1 Intuition Behind Policy Design

The pivotal challenge to effective cyber-risk diversification for a cyber re-insurer in Section 3 was to verify in a reasonable amount of time the existence of a dense k-subgraph in a bipartite graph that reflected strategic adversarial placement of lemon ICs by cyber-insurers taking advantage of **information asymmetry (IA)**. Without undertaking this step, a cyber re-insurer cannot be confident of covering aggregate cyber-losses of a portfolio of cyber-risks in an RCC put forward by a cyber-insurer. However, this task was proved to be computationally intractable, i.e., NP-hard, for the former. Ideally, any RCRM team designing a cyber-risk diversification portfolio would want to deal with a bipartite graph with no dense subgraphs—thereby relieving them of the computationally hard task of finding them. *The main solution principle behind our proposed hardness-mitigating policy is the construction of such a bipartite graph.* If such a graph can often be constructed in practice, then (a) the regulators could mandate cyber-insurance sellers to allocate lemon ICs on such a graph, decreasing the IA between the sellers and cyber re-insurers, and (b) the increase in seller-buyer informational transparency will boost the confidence of RCRM solution providers to densify the security-improving RCRM business.

Fortunately, there exists graph objects such as *bipartite expander graphs* [33] (in contrast to traditional bipartite graphs in Section 3) that are both sparse and highly connected, for which no dense k-subgraphs exist. More specifically, no matter how lemon ICs are placed by a strategic adversarial cyber-insurer on such a graph, the lemon costs are similar to that when the same cyber-insurer arbitrarily randomly places the lemon ICs on the graph. *Our challenge in this section is to show through theory that such a graph can be practically realized to promote optimal cyber-risk diversification portfolio design by RCRM entities such as re-insurers, under information asymmetry.*

4.2 A Framework to Model Policy Solution

As a representative example of an RCRM product, we consider cyber re-insurance for the same reasons as mentioned in Section 3. As usual, we consider two types of ICs handled by cyber-insurers for their individual organizational clients: good ICs and lemon ICs. We assume that the return on these ICs (both to the seller and the buyer sides) are dependent on "natural" phenomena occurring in cyber-space, which are captured through a "global" random variable Z. Z in practice represents the impact of events such as (a) zero-day attacks exploiting an OS vulnerability affecting organizations buying ICs, (b) DDoS attacks on targeted firms, (c) or simply, a state of cyber-space where cyber-attacks are non-correlated. For each instance z of the random variable (r.v.) Z, we have two probability distributions, $D_g = D_g(z)$ and $D_\ell = D_\ell(z)$, for good and lemon ICs, respectively. More

specifically, conditioned on $Z = z$, we assume (w.l.o.g.) all ICs are independent,[4] with good ICs chosen according to D_g and lemon ICs chosen according to D_ℓ. Moreover, we also assume that good ICs first-order stochastically dominate lemon ICs, i.e., for any z, a,

$$\Pr_{X\sim D_g(z)}[X \geq a] \geq \Pr_{Y\sim D_\ell(z)}[Y \geq a].$$

We normalize the returns of ICs, each to take a maximum value of 1, with μ and λ being the expected values of good and lemon ICs, respectively. First-order stochastic dominance implies $\mu \geq \lambda$, with $\delta = \mu - \lambda$ being the bonus expected value of a good IC, over a lemon IC.

Consider the most general case of a tranched **re-insurance claim contract (RCC)** being a derivative (in the general financial sense [34]) on an underlying portfolio of individual ICs sold by cyber-insurers to their clients. For $0 = a_0 < a_1 < \cdots < a_s$ (we call them attachment points), the ith tranche of the RCC is given by the interval $[a_{i-1}, a_i]$. Usually, there are three "riskiness" tranches: low, medium, and high. In practice, the cyber-insurer might cover the riskiest tranche of ICs (though not necessary for our model), thereby signalling to the re-insurer of its honesty and goodwill to claim coverage only for good and moderate ICs. Riskier tranches carry with them higher premiums. Suppose x is the payoff/return to a re-insurer of its underlying diversified portfolio of cyber-risks derived from the ICs. The value gained from the ith tranche is $\texttt{value}[a_{i-1}, a_i]\,(x) = \min\,(x, a_i)$ - $\min\,(x, a_{i-1})$. Given IC returns are normalized to 1, for a tranched RCC that depends on d ICs, the last attachment point is $a_s = d$. We denote by (n, m, d) - RCC family as one consisting of a set of m RCCs on n ICs identified with the set $[n] = \{1, 2, \ldots, n\}$, where each RCC depends on d ICs. Similar to that in Section 3, $n \ll md$, indicating each IC being in the portfolio of several RCCs. The cyber-insurer knows that some ℓ ICs are IA-driven lemons and may identify the lemons with any subset $L \subseteq [n]$ of size ℓ. For $L \subseteq [n]$, let $\text{tval}_{[a,b]}(L)$ be the total expected return to the re-insurer from all $[a, b]$ tranches in the RCC family, from the set L of lemon ICs. We define $\texttt{tval}(L) = (\text{tval}_{[a_0,a_1]}(L), \ldots, \text{tval}_{[a_{s-1},a_a]}(L))$.

A cyber-insurer will try to choose the subset L to minimize $\texttt{tval}(L)$ for the re-insurer. In other words, it will try to take advantage of information asymmetry to pay lesser premiums than what is fair for the lemon ICs it wants to re-insure against. Through the following definition, *we introduce the concept of a quasi-random RCC family, for which a cyber-insurer cannot gain significantly by adversarial placement of lemon ICs, over a random placement.* Our goal is to construct such a family.

Definition 1. Given any $[a, b]$ tranche, an (n, m, d)-RCC family is quasi-random for ℓ lemon ICs for that tranche with (adversarial vs. random IC placement) error ϵ, if for any two subsets $L, L' \subseteq [n]$ of size ℓ, $|t\ \text{val}_{[a,b]}\,(L') - \text{tval}_{[a,b]}(L)\ | \leq \epsilon \cdot m(b-a)$. Moreover, the entire (n, m, d)-RCC family is quasi-random for ℓ lemons with (adversarial vs. random IC placement) error ϵ, if for any two subsets $L, L' \subseteq [n]$ sized ℓ

$$\|\text{tval}\,(L') - \text{tval}(L)\|_1 \leq \epsilon \cdot md.$$

The maximum possible return on any tranche $[a, b]$ is $m(b-a)$—the situation when the tranche has no lemon ICs. Thus, for any RCC family, the error ϵ is at most 1. On a similar note, the maximum possible return on the entire RCC family is md, with ϵ for the entire family being bounded by the maximum error on any given tranche, i.e., 1. Note that adverse selection-induced lemon costs have two components: the unavoidable cost (a result of random lemon IC placement) and the cost of adversarial information asymmetry-driven lemon IC placement. Definition 1 says that the normalized cost of adversarial placement of lemon ICs is upper bounded by the quasi-random

[4]Our analysis also holds for the case when the conditional distribution on ICs is d-wise independent, i.e., any d of them are independent, as usual when re-insurers design coverage portfolios. (Note that d-wise independence does not imply mutual independence.)

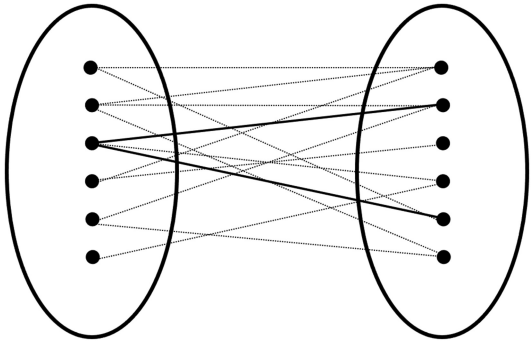

Fig. 4. Illustrating an example of bipartite expander graph exhibiting the edge expander and the unique neighbor properties. The set of vertices on the left denote ICs sold by a cyber-insurer to its client organizations. The set of vertices on the right denote IC-bundled RCCs put forward by cyber-insurers to its re-insurers.

error. We now introduce the concept of *bipartite expander graphs* [33] that will be used to relate and construct (see Section 4.3) the quasi-random RCC family stated above.

Definition 2. A bipartite graph on $[n] \cup [m]$, where $[n]$ is the set of ICs, and $[m]$ is the set of RCCs, is an $(\ell_{\max}, \gamma)$-expander graph if for every subset $S \subseteq [n]$ of size at most $\ell_{\max}$, $|\Gamma(S)| \geq \gamma|S|$, where $\Gamma(S) = \{v \mid (\exists w \in S)\{v, w\} \in E \subseteq [n] \times [m]\}$ denotes the set of S's neighbors. Moreover, let the set $\Gamma_i(S \subseteq [n])$ be that of vertices $v \in \Gamma(S)$ with $|\Gamma(v) \cap S| = i$, and $\Gamma_1(S)$ being called the unique neighbors of S. Then, the same bipartite graph on $[n] \cup [m]$ becomes an $(\ell_{\max}, \gamma)$-unique-neighbor expander if for every subset $S \subseteq [n]$ of size at most $\ell_{\max}$, $|\Gamma_1(S)| \geq \gamma|S|$.

Role of Expander Graphs in Cyber-risk Diversification Portfolio - The relevance of using bipartite expander graphs to RCRM policy-making is two-fold: (i) to ensure ICs from the left vertex set can be part of multiple cyber-risk diversifying RCCs on the right vertex set (the edge *expander* property), and (ii) each IC on the left vertex set is part of only a few RCC risk-diversification portfolios on the right vertex set (the *unique neighbor* property) *to avoid the existence of dense subgraphs* (see Figure 4). The avoidance of dense subgraphs mitigates IA-driven adverse selection issues for the benefit of RCRM solution providers.

4.3 Policy Analysis and towards its Computational Construction in Practice

Analysis - Thus far, we have independently described the concept and diversification-centric benefits of our proposed quasi-random RCC portfolio and the seminal unique-neighbor **bipartite expander graph (BEG)**. *What is the relation between the two of them?* It turns out that any $(n, m, d) -$ RCC family is quasi-random against ℓ lemon ICs for any $[a, b]$ RCC tranche, with error at most $\frac{\Delta \ell \delta}{m(b-a)}$, where $\delta = \mu - \lambda$ is the difference between the expected values of a good and lemon ICs, respectively, and Δ is a pre-defined constant. This is an outcome of the fact that a conversion of ℓ good ICs to lemon ICs reduces the re-insurer returns of the entire RCC family by $\Delta \ell \delta$. *We will now prove (based on theory in Reference [79]) that a maximum error of* $\frac{\Delta \ell \delta}{m(b-a)}$ *is achieved by a RCC family built from a* (t, d)*-biregular* $(\ell, t - \Delta)$*-unique neighbor BEG*, thereby establishing the above-mentioned relation. Here, a BEG is (t, d)-biregular if it is t-left-regular and d-right-regular; with a BEG being t-left(right)-regular if all left(right) vertices have degree t.

Theorem 6. *A RCC cyber-risk diversification portfolio built from a* (t, d)*-biregular* $(\ell, t - \Delta)$*-unique neighbor expander BEG is quasi-random for* ℓ *lemon ICs. For such a portfolio, the (adversarial*

vs. random IC placement) RCC tranche $[a, b]$ *error is at most* $\frac{\Delta \ell \delta}{m(b-a)}$*, and for the entire RCC portfolio the error is upper bounded by* $\frac{2\Delta \ell \delta}{md}$.

Proof. The proof starts by showing that in a portfolio of d ICs where g of them are good, val is a non-decreasing function of g due to the fact that good ICs first-order stochastically dominate lemon ICs. Here, $\text{val}_{[a,b]}(g) = \mathbb{E}[\text{value}_{[a,b]}(X)]$, where X is a random variable denoting the returns (to the re-insurer) of the portfolio of d ICs. The next part of the proof requires us to show (via **algebraic manipulation (AM)**) that for RCC tranche $[a, b]$, and for any $L, L' \subseteq [n]$, with $|L| = |L'| = \ell$, we have $\text{tval}(L') - \text{tval}(L) = \sum_{i=1}^{d}(t_i(L) - t_i(L'))(\text{val}(d) - \text{val}(d-i))$. Finally, for individual tranche analysis, we show (via AM) that for $u_i \in [-\beta, \beta]$, $v_i \in [0, \delta]$, we have (for all i from 1 to d) $|\sum_{i=1}^{d} u_i v_i| \leq \beta\delta$ if $\sum_{i=1}^{d} u_i = 0$, and $\sum_{i=2}^{d} |u_i| \leq \beta$. These three steps ensure that the (adversarial vs. random IC placement) RCC tranche $[a, b]$ error is at most $\frac{\Delta \ell \delta}{m(b-a)}$. Similarly, for the entire RCC portfolio analysis, we show that for $u_i \in [-\beta, \beta]$, $v_{ij} \in [0, \delta]$, we have (for all i from 1 to d) $\sum_{j=1}^{s} |\sum_{i=1}^{d} u_i v_{ij}| \leq 2\beta\delta$ if $\sum_{i=1}^{d} u_i = 0$, and $\sum_{i=2}^{d} |u_i| \leq \beta$, and $(\forall i)\ \sum_{j=1}^{s} v_{ij} \leq \delta$. This fourth step together with the three above ensures that the entire RCC portfolio error is upper bounded by $\frac{2\Delta \ell \delta}{md}$. □

Implication to Cyber-risk Management - In the context of cyber-insurance economies, the theorem states that designing quasi-random RCC cyber-risk diversification portfolios induced upon **bipartite expander graphs (BEGs)** *does not incentivize adversarial cyber-insurers to take advantage of information asymmetry* and *strategically* pack lemon ICs in RCCs at the disadvantage of the cyber re-insurer. The benefit for the former in doing so is minuscule.

BEG Construction - Thus far, we have shown the power of BEG-induced quasi-random cyber-risk diversification portfolios to alleviate the negative effects of information asymmetry in the RCRM business for CPS societies. However, an important question that follows is: *how can we engineer/construct such graphs in practice?* To this end, we borrow the following seminal result from References [26, 27, 79] to explicitly construct BEGs with an associated parameter set.

Theorem 7 ([26, 27, 79]). *For any* $\alpha \in (0, 1]$ *and positive integers* n, m*, and* t*, there is an explicit construction of an* $(\ell_{\max}, t - \Delta)$ *BEG on* $[n] \cup [m]$ *with left degree* t *for* $\ell_{\max} = (m/(4t^2))^{\alpha}$ *and* $\Delta = (2t)^{\alpha}(\log_t n)\log_t m$*. Moreover, suppose we are given a do-left-regular* $(\ell_{\max}, \gamma)$ *BEG on* $[n] \cup [m_0]$*, and parameters* m, t, d *such that* $nt = md$, $t_0 < t \leq m_0$*, and* $m \geq m_0 t/(t - t_0)$*. We can efficiently construct a* (t, d) *-biregular* $(\ell_{\max}, \gamma)$ *BEG on* $[n] \cup [m]$.

Implication to Cyber-risk Management - The theorem computationally realizes an exact parametric construction of bipartite expander graphs in theory that directly maps to a quasi-random RCC of n ICs and m RCC cyber-risk diversification portfolios. This construction can be done by the cyber re-insurer prior to deciding to opt in or opt out of the aggregate cyber-risk coverage business in CPS societies. In practice, given m, t, and d, it is straightforward to extend the parametric construction to real-world realization pivoted upon probabilistic approaches such as the ones in References [58, 71] (used in optical and data-center networks, respectively). Such engineering approaches could generate non-unique BEGs (unless uniqueness property from Definition 2 is satisfied). The regulatory good thing, though, is each BEG will equally contribute to IA mitigation, in line with Theorem 6.

5 RELATED WORK

In this section, we provide a brief overview of research related to our efforts. *To the best of our knowledge, there is no existing work that formally establishes the tractability/intractability of the cyber-risk diversification problem*, be it in the domain of finance/actuarial sciences or in the area of

cyber-security—let alone efforts to mitigate intractability challenges. This problem is synonymous with characterizing the computational hardness of mitigating **information asymmetry (IA)**. *Hence, our efforts in this article are completely new with respect to finding how hard it is to mitigate IA in the security-improving RCRM business for CPS societies.* However, we provide a concise description of existing research on RCRM.

5.1 Success of Cyber-insurance-type RCRM Markets

Introductory foundational research works on cyber-insurance [32, 41, 47] have mathematically shown the existence of economically inefficient insurance markets. Intuitively, an efficient market is one where all stakeholders (market elements) mutually satisfy their interests. These works state that cyber-insurance markets satisfy every stakeholder apart from the regulatory agency (e.g., government), and sometimes the profit-minded cyber-insurer itself. In Reference [49], the authors proposed a Coasian bargaining approach among cyber-insured network entities to achieve an efficient insurance market—however, costless bargaining under which the Coase theorem holds is idealistic in nature and might not be feasible to implement in practice. Lelarge et al. in Reference [41] recommended the use of fines and rebates on cyber-insurance contracts to make each user invest optimally in self-defense and make the network optimally robust. However, their work neither mathematically proves the effectiveness of premiums and rebates in making network users invest optimally, nor does it guarantee the strict positiveness of insurer profits at all times. In relative more recent works [37, 50, 52, 53], the authors overcome the drawbacks of the mentioned existing works and propose ways to form provably efficient monopolistic cyber-insurance markets by satisfying market stakeholders, including a risk-averse cyber-insurer, in environments of interdependent risk with partial IA. The authors in References [45, 46, 50] further state the importance of compulsory insurance for optimizing social welfare for primary cyber-insurance markets.

Differences - In the first place, these works are orthogonal in their goals, when compared to our effort. Our main focus is to computationally justify the fact that optimal cyber-risk diversification task enroute to underwriting of RCRM contracts for CPS societies under IA issues is hard even for a computer (which directly affects the success of RCRM markets)—let alone profit-minded human-driven RCRM organizations. In contrast, the above-mentioned works primarily focus on investigating the market efficiency (or lack of) of (voluntary) cyber-insurance-type RCRM markets in the presence and (partial) absence of IA, oblivious of the hardness of the cyber-risk diversification task, i.e., mitigating IA. Another major difference with the existing works is their way in concretely modeling IA. While they use economic contract-theoretic constructs to capture adverse selection, we encapsulate the latter uniquely via graph theory.

5.2 On Re-insurance-type RCRM that Aggregates Cyber-risk

Re-insurance type RCRM product businesses to manage catastrophic cyber-incidents in CPS societies are in a similar boat as their traditional cyber-insurance counterparts when it comes to facing the brunt of IA between the supply and the demand sides. More so, because *the systemic nature of cyber and the potential for losses that transcend geography, industry, and class, is leading to the rapid demand for aggregate cyber-risk coverage on multiple service supply-chain lines of service-networked businesses in CPS societies.* There are quite a few instances in practice where individual cyber-risks (a candidate cyber-risk distribution on each of these lines) have shown heavy-tailed impact [21, 42, 74]. There are also studies establishing the dependence among cyber risks. Notable among them are References [10, 29, 44, 51, 59, 73, 75–77]. In all these statistics, IA has a crucial role to play in the sense that IA-mitigating policies and subsequent organizational liabilities in operation could in the first place prevent the cyber-risks becoming heavy-tailed. In a recent set of works [54, 55, 57], the authors develop formal analysis frameworks to decide the feasibility and

sustainability of aggregating a set of (correlated) cyber-risks with heterogeneous tail properties for a cyber-risk aggregating RCRM business. As a new result (and in contrast to previous empirical ones), they formally prove that even i.i.d., heavy-tailed cyber-risks of certain types are not commercially suitable for aggregation by the RCRM firms—let alone correlated ones.

Differences - Our efforts in this article are orthogonal to the above-mentioned works. We focus on investigating the hardness of mitigating IA issues in the design of re-insurance-type RCRM contracts for CPS societies. We address a computational dimension to assess the feasibility of cyber-risk diversification when compared to a statistical dimension in prior works.

6 DISCUSSION

In this section, we first discuss about the inherent market density landscape from the supply and demand sides in the current RCRM (e.g., cyber (re-)insurance) business-serving CPS societies and associate it with existing inevitable information asymmetry challenges. We comment, where applicable, on how these challenges might impact cyber-insurance solution markets for service-networked smart ICS societies. We then realize the NP-hardness flavor of the optimal dense subgraph problem by studying simulation trends on the time it takes in practice to compute dense subgraphs with varying bipartite graph sizes.

6.1 The Current RCRM Market Density Landscape

Companies looking to buy cyber insurance protection today face a fairly volatile environment shaped by low prices for protection and high levels of risk sustained by insurers. This has led to a big gap in allocated insurance policy capacity (when compared to the demand) we are unsurprisingly seeing right now. According to a personal communication with Marsh, between 2016 and 2020 cyber insurance for CPS societies fell into a soft market, with broadening coverage and low prices. Since 2021, post COVID 19 and in the age of work-from-home, the cost of capital for cyber insurance has risen dramatically, and most insurers and re-insurers have reacted with higher prices, limited coverage, and stringent control conditions that cyber-insurance policy buyers need to satisfy to get meaningful coverage under environments of increasingly high information asymmetry. Information asymmetry between the supply and demand sides of the cyber-insurance market that is fueled by lack of strong vulnerability disclosure regulations is one of the major causes for the aforementioned big gap in practice and the subsequent recent rise in insurance premiums. But the underlying problem is not going away: Cyber risks will persist and evolve and will scale up significantly in CPS-run ICS societies, and companies will need to manage the increasing risk, including securing insurance protection. Because of the lack of historical experience as an industry in addition to existing information asymmetry challenges, there is no easy way to fix the market. Perhaps strong regulations on cyber-breach information sharing initiatives along with standardized IoT device controls would greatly help the cause in reducing the supply-demand gap in the near future. In addition, the NP-hardness of basic cyber-risk diversification constructs (as shown in the article) will not help either in promoting dense markets if information disclosure and the IoT controls standards (e.g., default security set in IoT devices) are not upgraded.

According to a recent Harvard Business Review article,[5] one of the most difficult barriers to addressing the structural challenges that the cyber insurance sector faces is that insurers have disproportionately relied on cyber re-insurance. Reinsurance—again, casually thought of as insurance for insurance companies—allows insurers to lay off risk to another capital source. Much as you turn to your insurer when you have a claim, insurers may look to re-insurers for support—and in the case of cyber, insurers cede an estimated 50% of the premium they collect to the re-insurance

[5]https://hbr.org/2021/01/cybersecurity-insurance-has-a-big-problem.

market. So, they do not retain as much of the risk as one might think. As a result, the concentration of capital among re-insurers is simply striking. This is only going to increase as the number and size of CPS smart societies grow around the globe, and along with this the size of aggregated cyber-risks liable upon any service sector. Four re-insurers in the USA today account for more than 60% of premium—and that cohort's concentration could grow as a result of market volatility in the coming years, as smaller RCRM solution providing players reassess their commitment to cyber. In fact, more than 75% of the re-insurers writing cyber re-insurance have less than $100 million in premium, and most of them less than $50 million. With the largest re-insurer in the market likely seeing more than $500 million in premium, it is roughly the same size as the collection of companies writing less than $100 million, based on known data. The advent and growth of service-networked smart cities might most likely increase the demand for RCRM products. However, uncertainty of loss outcomes, existing and increasing information asymmetries regarding the cyber-postures of ICS-reliant organizations, the subsequent cyber-risk diversification challenges in a scaled-up organizational network environment, large-scale negative network externalities, and the possibility of new cyber-risk types could make insurers cautious about rising to meet the increases in demand, even if there exists a pricing structure (and externally invested capital) that supports it.

Thus, on one hand, it might be a while until which organizations might have to significantly invest in self-insurance, when compared to commercial cyber-insurance. However, on the other hand, it is also true that the organizational demand (pushed through rising C-suite concerns) for cyber-insurance solutions will steadily increase (as we see from market statistics) over time and with the rise in the number of smart cities around the globe. Given that information asymmetry challenges will be on both sides (the supply and demand) of the RCRM solutions market,[6] it would be difficult for the demand side, i.e., organizations in ICS societies buying cyber-insurance policies, to feel "taken-advantaged-of" through theoretically unfair premiums that may necessarily have to be induced on them by the supply side to scale the RCRM business (hence its slow and steady rise). Consequently, as mentioned above, organizations will continue to allocate budget in both self-insurance and RCRM (e.g., cyber-insurance) solutions. The collective need looking forward is a greater co-operation between leading cyber (re-)insurers, policy-makers, risk model vendors, researchers, and major technology companies, to "sail" through the challenges posed by information asymmetry issues in the complex inter-dependent service network for improved cyber-risk diversification in CPS-reliant societies. One recent example of such a cooperation being the pioneering *Munich Re* partnership[7] with *Google Cloud* and *Allianz.*

6.2 How Does the NP-hard Nature of Dense Subgraph Detection Reflect in Practice?

In theory, proving the dense subgraph problem to be NP-hard just showcases the fact that for certain worst-case bipartite graph configurations, it could take a lot of time to detect dense subgraphs. However, in practice, most scenarios that a cyber-risk manager encounters are usually "average" cases. Nonetheless, conveying a message of NP-hardness of a basic, critical, and necessary cyber-risk management component such as diversification can un-necessarily amplify the risk-averse mindset of RCRM solution providers. We conjecture that the "average" cases with respect to dense subgraph detection for bipartite graphs in the practical world will not be relatively compute-demanding. To this end, we study simulation trends on the time it takes in practice to compute

[6]The demand side might in most cases not have complete knowledge of its exposure to cyber-risk. This is due to the fact that C-suite executives are often uncertain about the cyber-risk landscape their organization is exposed to, and hence unsure of appropriate amounts of investments in cyber-security controls.

[7]https://cloud.google.com/press-releases/2021/0302/allianz-munichre-risk-management.

dense subgraphs with varying bipartite graph sizes, hoping to provide confidence to scale RCRM markets in CPS-driven ICS societies, despite the inherent NP-hard nature of the cyber-risk diversification problem. However, this does not reduce the importance of designing techniques in practice to mitigate worst-case IA. This simply because it just takes one worst-case cyber-event (no matter how rare it is) to take an insurance company (not serious about mitigating IA) out of business.

Monte Carlo Simulation Experiment - We conduct Monte Carlo simulations to computationally solving the k-densest subgraph problem and study the time it takes to find such a subgraph. The detailed simulation process is illustrated as follows through the stated steps:

(1) First, we fix the number of nodes of a randomly generated bipartite graphs to lie in the following set $N \in \{50, 100, 500, 1{,}000, 5{,}000, 10{,}000\}$. This set covers bipartite graph sizes ranging from small to big. Note that for any commercial cyber-insurer, 10,000 client organizations is big enough in CPS society settings.
(2) To sample k for the densest k-subgraph problem, we choose five k's (in our work $k = n$) between $0.01 * N$ and $0.1 * N$ uniformly. Note that $k = n$ represents the number of lemon ICs.
(3) In each random trial instance (out of a total of 100,000 instances) of a Monte Carlo simulation for any fixed N, we randomly choose the number of nodes in one set, X between 1 and N, and let the number of nodes in the other set to be $Y = N - X$, as it is a bipartite graph; we also randomly choose the number of edges, Z, of a random instance of a bipartite graph; finally given X, Y, and Z, we construct a random instance, B, of a bipartite graph.
(4) We iterate through all possible subgraphs for any instance of $B|N$ and find, for each instance, the number of edges E of the densest k-subgraph, given $k(=n)$.
(5) We re-iterate the above step for all possible k values and record the time until we find optimal densest k-subgraphs for varying N and $k|N$.
(6) Finally, we plot the simulation statistics.

Simulation Results - We observe from our simulation results that the mean and median time (stated in normalized machine time units) to detect the existence of dense subgraphs over all simulated instances for a specific $(k = n, N)$ pair is far lower (approximately $\frac{1}{8}$-fold) than the mean and median time to detect the same in outlier (the top five percentile) instances (see Figure 5). These outlier instances are the ones that contribute to NP-hardness of the dense subgraph detection problem in bipartite graphs. Evidently, the mean and median gap between the simulation running time for general and outlier graph instances is monotonically increasing with k and N (see Figures 5 to 8). The results in Figures 5 to 8 show that for "average" cases, the cyber-risk diversification problem may not turn out that computationally challenging for a cyber re-insurer as may be perceived from the fact that the general problem is NP-hard. We would expect, from practical experience, the occurrence probability of outlier cases to be quite low—though we cannot guarantee this belief for fast-evolving CPS-driven ICS societies and risk terrains. In other words, one (i.e., a cyber-risk manager) should always be aware of rare events arising from a fat-tailed distribution reflecting the occurrence of dense subgraphs that may take an exponential time to detect. It is such events for which we have designed our graph-theoretic IA-mitigation mechanism, potentially preventing cyber-risk managers going out of business due to mis-estimating risk estimates. Hence, we can conservatively recommend a safe scaling of the much-needed RCRM business for current CPS-driven ICS societies without worrying about the general cyber-risk diversification intractability.

7 DISCUSSION AND SUMMARY

In this section, we first discuss two directions that can improve the density of current cyber (re-)insurance markets, apart from our proposed methodology in the article. The densification of cyber-insurance markets will necessarily happen only when organizations better their

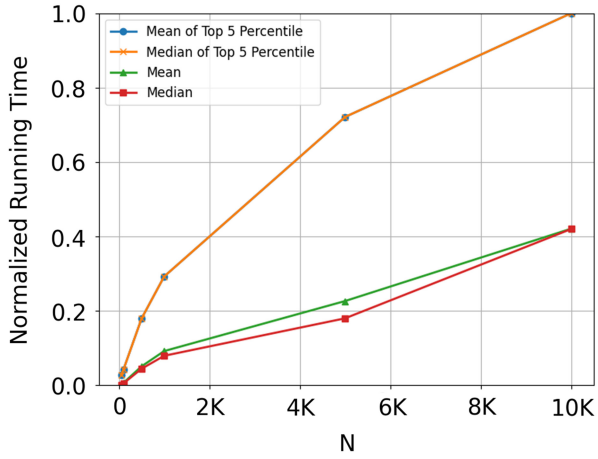

Fig. 5. Illustration of the normalized running time for detecting a densest k-subgraph as a function of the bipartite graph size (N), and $k(=n) \in [.01N, .1N]$. We simulate 100,000 random instances of bipartite graphs for each N.

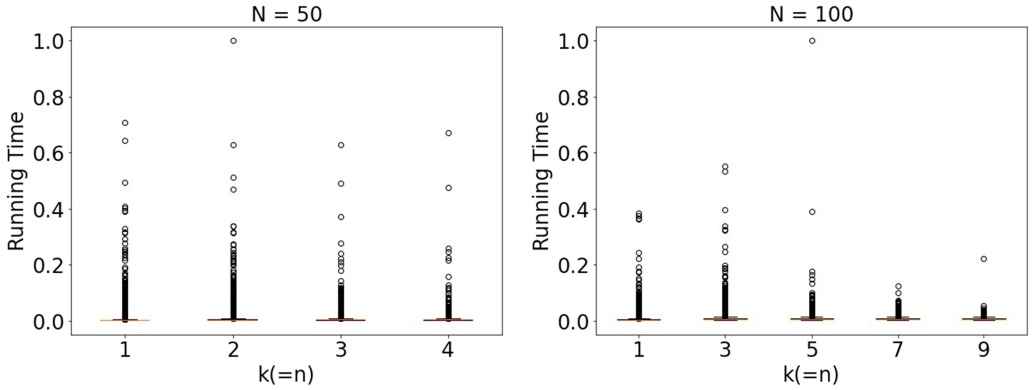

Fig. 6. Illustration of normalized running time statistics for detecting a densest k-subgraph as a function of the bipartite graph size (a) $N = 50$ (left) and (b) $N = 100$ (right). We simulate 100,000 random instances for each k.

cyber-postures through "best practice" compliance and governance. This is a non-trivial exercise but necessary to provide confidence to cyber-insurers in IA environments. To this end, we also discuss about related organizational compliance and governance issues that might arise. We finally summarize our proposed research.

A Discussion on Alternative Approaches - Despite the fact that our proposed computational policy to mitigate IA will promote cyber (re-)insurance market densification, it is always better to improve (ICS) cyber-vulnerability information sharing "standards." As an alternative (and complementary) approach to improve the density of cyber (re-)insurance markets, regulators should frame laws to boost cyber-vulnerability information sharing in society—this, more in the wake of modern nation wars (e.g., the Russia-Ukraine war) that are likely to target ICSs to cripple the economy and society. As an example of such laws, on March 15, 2022, President Joe Biden signed into law the Cyber Incident Reporting for Critical Infrastructure Act. According to this

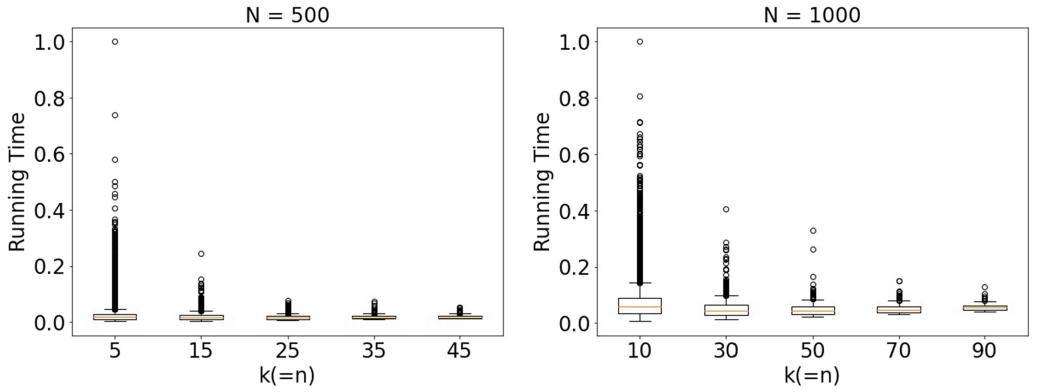

Fig. 7. Illustration of normalized running time statistics for detecting a densest k-subgraph as a function of the bipartite graph size (a) $N = 500$ (left) and (b) $N = 1{,}000$ (right). We simulate 100,000 random instances for each k.

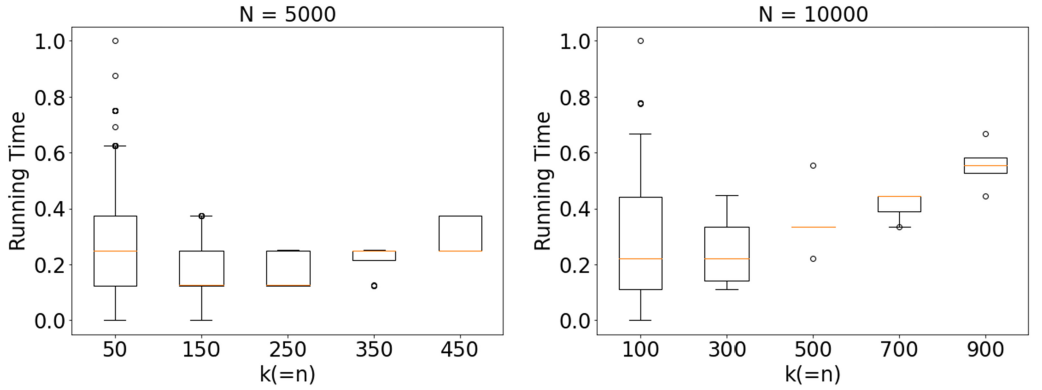

Fig. 8. Illustration of normalized running time statistics for detecting a densest k-subgraph as a function of the bipartite graph size (a) $N = 5$K (left) and (b) $N = 10$K (right). We simulate 100,000 random instances for each k.

act, IT/OT organizations must report certain cyber incidents to the **Cybersecurity and Infrastructure Security Agency (CISA)** of the U.S. DHS within 72 hours and an obligation to report ransomware payments within 24 hours. In addition, the C-suite of each ICS should share best risk governance practices among business partners and encourage its ICS management to participate in public cyber-vulnerability sharing platforms. Another alternative approach to densify security-improving cyber (re-)insurance markets is the introduction of insurance-linked securities market via the sale of financial catastrophic bonds. The crux behind the potential effectiveness of ILS solutions is that catastrophic bond markets rely on trading securities in non-correlated multi-trillion financial markets. Since financial markets can draw on larger, more liquid, and increasingly diversified pool of capital than the equity of cyber (re-)insurance markets, diversifying (at most) a few hundred billion dollars of cyber-risk by cyber (re-)insurers in a 30-odd trillion dollar market is akin to insuring a *drop in an ocean*. In the context of insuring ICS, cyber-insurers can get enough capital injected into their business (through re-insurers) using ILSs to be able to rapidly proliferate markets and alleviate the *market of lemons problem*—typical of the cyber (re-)insurance industry.

Organizational Compliance and Governance Challenges - Cyber-security policies established by organization (ICS or otherwise) C-suites will not be effective if employees (including the C-suite) and organizational end-users are not keen or are unwilling to follow security policies. Subsequently, cyber-insurance firms will not be confident of selling solutions to such organizations. The authors in [31] study using literature on IS adoption, protection-motivation theory, deterrence theory, and organisational behavior—the compliance mindset of organizational employees. The results in Reference [31] suggest that (a) threat perceptions about the severity of breaches and response perceptions of response efficacy, self-efficacy, and response costs are likely to affect policy attitudes; (b) organizational commitment and social influence significantly impact employee compliance intentions; and (c) cyber-security resource availability significantly enhances self-efficacy, which eventually increases employee intentions to comply with C-suite policies. With respect to the latter point, the authors in Reference [30] propose a statistical auditing method to evaluate the value of organizational IT/OT resources for the C-suite to allocate an appropriate budget for incident prevention, management, and response. In Reference [17], the authors use coping theory to study an underlying relationship between employee stress and deliberate violations in information security policy. This stress in question is caused by burdensome, complex, and ambiguous information security requirements. These requirements lead to a moral and cognitively rationalized disengagement from security policy violations by employees. Given a recent research idea of making cyber-security an organizational strategic advantage over competitors [28], it is imperative that well-designed security policies merge with good employee compliance.

Article Summary - Effective cyber-risk diversification is a necessary pre-requisite for successful and scalable residual cyber-risk management solutions to proliferate in a market. In this article, we proved that optimal cyber-risk diversification (an integral step to RCRM contract design) under **information asymmetry (IA)** is computationally intractable, i.e., NP-hard, for the cyber-space spanned by IT/IoT driven ICS societies. In other words, there are innumerable number of ways in which an ICS organization can get cyber-breached, and in the presence of IA, even a computer cannot figure those out in a reasonable amount of time in the worst case. As a result, ICS cyber-risk might not be effectively estimated. This leads to a lack of re-insurer confidence in diversifying/aggregating cyber-risks specific to intra ICS-networked (supply chain) settings affected by correlated and inter-dependent cyber-risks. We proved the NP-hardness of the optimal cyber-risk diversification problem under IA by viewing it from the lens of a NP-hard graph-theoretic problem of finding the densest k-subgraph in a graph. Subsequently, as part of computational policy design, we followed this up with the design of a provably optimal *quasi-random* cyber-risk diversification environment induced upon *expander graphs* that significantly alleviates the IA problem between the supply and demand sides of the RCRM business in polynomial time. In other words, our proposed approximately random computational policy optimally mitigates (but does not eliminate) negative cyber-vulnerability IA effects in favor of the optimal cyber-risk diversification task. Our work formally established the reason why it has been very difficult to date to significantly densify steadily rising (and much-wanted) RCRM markets under information asymmetry in CPS-driven ICS societies despite their high demand, and proposed an environment to turn things around for the benefit of the cyber-society. More specifically, we proposed a computational policy that we recommend for implementation by regulators (and indeed doable via simple rule adds-ons to the status quo) to densify (via computationally enabling diversification portfolio optimization) the market presence of cyber-risk diversifying RCRM solution providers and significantly improve cyber-security.

ACKNOWLEDGMENT

This is not a student article. The list of authors is in decreasing order of contributions. R. Pal is the lead and corresponding author. P. Liu and T. Lu contributed equally to the article. Qualitative opinions expressed by the authors in the article are completely their own and do not necessarily reflect the opinions of the organizations they represent.

REFERENCES

[1] Ali Ahmed, Amit Deokar, and Ho Cheung Brian Lee. 2021. Vulnerability disclosure mechanisms: A synthesis and framework for market-based and non-market-based disclosures. *Decis. Supp. Syst.* 148 (2021), 113586.
[2] George A. Akerlof. 1978. The market for "lemons": Quality uncertainty and the market mechanism. In *Uncertainty in Economics*. Elsevier, 235–251.
[3] Ross Anderson and Tyler Moore. 2009. Information security: Where computer science, economics and psychology meet. *Philos. Trans. Roy. Societ. A: Math., Phys. Eng. Sci.* 367, 1898 (2009), 2717–2727.
[4] Benny Applebaum, Boaz Barak, and Avi Wigderson. 2010. Public-key cryptography from different assumptions. In *42nd ACM Symposium on Theory of Computing*. 171–180.
[5] Ashish Arora, Rahul Telang, and Hao Xu. 2008. Optimal policy for software vulnerability disclosure. *Manag. Sci.* 54, 4 (2008), 642–656.
[6] Sanjeev Arora, Boaz Barak, Markus Brunnermeier, and Rong Ge. 2011. Computational complexity and information asymmetry in financial products. *Commun. ACM* 54, 5 (2011), 101–107.
[7] Aditya Bhaskara, Moses Charikar, Eden Chlamtac, Uriel Feige, and Aravindan Vijayaraghavan. 2010. Detecting high log-densities: An O (n 1/4) approximation for densest k-subgraph. In *42nd ACM Symposium on Theory of Computing*. 201–210.
[8] Christian Biener, Martin Eling, and Jan Hendrik Wirfs. 2015. Insurability of cyber risk: An empirical analysis. *Geneva Papers Risk Insur.—Iss. Pract.* 40, 1 (2015), 131–158.
[9] Baidyanath Biswas, Arunabha Mukhopadhyay, Sudip Bhattacharjee, Ajay Kumar, and Dursun Delen. 2022. A text-mining based cyber-risk assessment and mitigation framework for critical analysis of online hacker forums. *Decis. Supp. Syst.* 152 (2022), 113651.
[10] Rainer Böhme and Gaurav Kataria. 2006. Models and measures for correlation in cyber-insurance. In *Workshop on the Economics of Information Security*.
[11] Rainer Böhme, Galina Schwartz, et al. 2010. Modeling cyber-insurance: Towards a unifying framework. In *Workshop on the Economics of Information Security*.
[12] Jean-Chrysostome Bolot and Marc Lelarge. 2008. A new perspective on internet security using insurance. In *IEEE 27th Conference on Computer Communications*. IEEE, 1948–1956.
[13] Maurizio Bruglieri, Matthias Ehrgott, Horst W. Hamacher, and Francesco Maffioli. 2006. An annotated bibliography of combinatorial optimization problems with fixed cardinality constraints. *Discr. Appl. Math.* 154, 9 (2006), 1344–1357.
[14] Andrew Coburn, Eireann Leverett, and Gordon Woo. 2018. *Solving Cyber Risk: Protecting your Company and Society*. John Wiley & Sons.
[15] Derek G. Corneil and Yehoshua Perl. 1984. Clustering and domination in perfect graphs. *Discr. Appl. Math.* 9, 1 (1984), 27–39.
[16] Renata Paola Dameri, Clara Benevolo, Eleonora Veglianti, and Yaya Li. 2019. Understanding smart cities as a glocal strategy: A comparison between Italy and China. *Technol. Forecast. Social Change* 142 (2019), 26–41.
[17] John D'Arcy, Tejaswini Herath, and Mindy K. Shoss. 2014. Understanding employee responses to stressful information security requirements: A coping perspective. *J. Manag. Inf. Syst.* 31, 2 (2014), 285–318.
[18] Saini Das, Arunabha Mukhopadhyay, Debashis Saha, and Samir Sadhukhan. 2019. A Markov-based model for information security risk assessment in healthcare MANETs. *Inf. Syst. Front.* 21, 5 (2019), 959–977.
[19] Peter DeMarzo and Darrell Duffie. 1999. A liquidity-based model of security design. *Econometrica* 67, 1 (1999), 65–99.
[20] Peter M. DeMarzo. 2005. The pooling and tranching of securities: A model of informed intermediation. *Rev. Finan. Stud.* 18, 1 (2005), 1–35.
[21] Benjamin Edwards, Steven Hofmeyr, and Stephanie Forrest. 2016. Hype and heavy tails: A closer look at data breaches. *J. Cybersecur.* 2, 1 (2016), 3–14.
[22] Uriel Feige, David Peleg, and Guy Kortsarz. 2001. The dense k-subgraph problem. *Algorithmica* 29, 3 (2001), 410–421.
[23] Uriel Feige, Michael Seltser, et al. 1997. *On the Densest K-subgraph Problem*. Citeseer.
[24] Kevin M. Gatzlaff and Kathleen A. McCullough. 2010. The effect of data breaches on shareholder wealth. *Risk Manag. Insur. Rev.* 13, 1 (2010), 61–83.
[25] Alasdair Gilchrist. 2017. *IoT Security Issues*. Walter de Gruyter GmbH & Co KG.

[26] Venkatesan Guruswami, James R. Lee, and Alexander Razborov. 2010. Almost euclidean subspaces of 1 N VIA expander codes. *Combinatorica* 30, 1 (2010), 47–68.
[27] Venkatesan Guruswami, Christopher Umans, and Salil Vadhan. 2009. Unbalanced expanders and randomness extractors from Parvaresh–Vardy codes. *J. ACM* 56, 4 (2009), 1–34.
[28] Manuel Hepfer and Thomas C. Powell. 2020. Make cybersecurity a strategic asset. *MIT Sloan Manag. Rev.* 62, 1 (2020), 40–45.
[29] Hemantha Herath and Tejaswini Herath. 2011. Copula-based actuarial model for pricing cyber-insurance policies. *Insur. Mark. Compan.: Anal. Actuar. Computat.* 2, 1 (2011), 7–20.
[30] Hemantha S. B. Herath and Tejaswini C. Herath. 2008. Investments in information security: A real options perspective with Bayesian postaudit. *J. Manag. Inf. Syst.* 25, 3 (2008), 337–375.
[31] Tejaswini Herath and H. Raghav Rao. 2009. Protection motivation and deterrence: A framework for security policy compliance in organisations. *Eur. J. Inf. Syst.* 18, 2 (2009), 106–125.
[32] A. Hoffman. 2007. Internalizing externalities of loss prevention through insurance monopoly. *Geneva Risk Insur. Rev.* 32 (2007).
[33] Shlomo Hoory, Nathan Linial, and Avi Wigderson. 2006. Expander graphs and their applications. *Bull. Amer. Math. Soc.* 43, 4 (2006), 439–561.
[34] John C. Hull. 2003. *Options Futures and other Derivatives*. Pearson Education India.
[35] Karthik Kannan and Rahul Telang. 2005. Market for software vulnerabilities? Think again. *Manag. Sci.* 51, 5 (2005), 726–740.
[36] J. Mark Keil and Timothy B. Brecht. 1991. The complexity of clustering in planar graphs. *J. Combinat. Math. Combinat. Comput.* 9 (1991), 155–159.
[37] Mohammad Mahdi Khalili, Parinaz Naghizadeh, and Mingyan Liu. 2018. Designing cyber insurance policies: The role of pre-screening and security interdependence. *IEEE Trans. Inf. Forens. Secur.* 13, 9 (2018), 2226–2239.
[38] Subhash Khot. 2006. Ruling out PTAS for graph min-bisection, dense k-subgraph, and bipartite clique. *SIAM J. Comput.* 36, 4 (2006), 1025–1071.
[39] Guy Kortsarz and David Peleg. 1993. *On Choosing a Dense Subgraph*. IEEE.
[40] Juhee Kwon and M. Eric Johnson. 2018. Meaningful healthcare security: Does meaningful-use attestation improve information security performance? *MIS Quart.* 42, 4 (2018), 1043–1068.
[41] Marc Lelarge and Jean Bolot. 2009. Economic incentives to increase security in the internet: The case for insurance. In *IEEE International Conference on Computer Communications*. IEEE, 1494–1502.
[42] Thomas Maillart and Didier Sornette. 2010. Heavy-tailed distribution of cyber-risks. *Eur. Phys. J. B* 75, 3 (2010), 357–364.
[43] Arunabha Mukhopadhyay, Samir Chatterjee, Kallol K. Bagchi, Peteer J. Kirs, and Girja K. Shukla. 2019. Cyber risk assessment and mitigation (CRAM) framework using logit and probit models for cyber insurance. *Inf. Syst. Front.* 21, 5 (2019), 997–1018.
[44] Arunabha Mukhopadhyay, Samir Chatterjee, Debashis Saha, Ambuj Mahanti, and Samir K. Sadhukhan. 2013. Cyber-risk decision models: To insure IT or not? *Decis. Supp. Syst.* 56 (2013), 11–26.
[45] Parinaz Naghizadeh and Mingyan Liu. 2014. Voluntary participation in cyber-insurance markets. In *Workshop on the Economics of Information Security (WEIS)*.
[46] Parinaz Naghizadeh and Mingyan Liu. 2016. Exit equilibrium: Towards understanding voluntary participation in security games. In *35th IEEE International Conference on Computer Communications*. IEEE, 1–9.
[47] N. Shetty, G. Schwarz, M. Feleghyazi, and J. Walrand. 2009. Competitive cyber-insurance and internet security. In *Workshop on the Economics of Information Security*.
[48] Hulisi Öğüt, Srinivasan Raghunathan, and Nirup Menon. 2011. Cyber security risk management: Public policy implications of correlated risk, imperfect ability to prove loss, and observability of self-protection. *Risk Anal.* 31, 3 (2011), 497–512.
[49] Ranjan Pal and Leana Golubchik. 2010. Analyzing self-defense investments in internet security under cyber-insurance coverage. In *IEEE 30th International Conference on Distributed Computing Systems*. IEEE, 339–347.
[50] Ranjan Pal, Leana Golubchik, and Konstantinos Psounis. 2011. Aegis, a novel cyber-insurance model. In *International Conference on Decision and Game Theory for Security*. Springer, 131–150.
[51] Ranjan Pal, Leana Golubchik, Konstantions Psounis, and Tathagata Bandyopadhyay. 2019. On robust estimates of correlated risk in cyber-insured IT firms: A first look at optimal AI-based estimates under "Small" data. *ACM Trans. Manag. Inf. Syst.* 10, 3 (2019), 1–18.
[52] Ranjan Pal, Leana Golubchik, Konstantinos Psounis, and Pan Hui. 2014. Will cyber-insurance improve network security? A market analysis. In *IEEE Conference on Computer Communications*. IEEE, 235–243.
[53] Ranjan Pal, Leana Golubchik, Konstantinos Psounis, and Pan Hui. 2018. Improving cyber-security via profitable insurance markets. *ACM SIGMETRICS Perform. Eval. Rev.* 45, 4 (2018), 7–15.

[54] Ranjan Pal, Ziyuan Huang, Sergey Lototsky, Xinlong Yin, Mingyan Liu, Jon Crowcroft, Nishanth Sastry, Swades De, and Bodhibrata Nag. 2021. Will catastrophic cyber-risk aggregation thrive in the IoT age? A cautionary economics tale for (re-) insurers and likes. *ACM Trans. Manag. Inf. Syst.* 12, 2 (2021), 1–36.

[55] Ranjan Pal, Ziyuan Huang, Xinlong Yin, Sergey Lototsky, Swades De, Sasu Tarkoma, Mingyan Liu, Jon Crowcroft, and Nishanth Sastry. 2020. Aggregate cyber-risk management in the IoT age: Cautionary statistics for (re) insurers and likes. *IEEE InternetThings J.* (2020).

[56] Ranjan Pal, Taoan Lu, Peihan Liu, and Xinlong Yin. 2021. Cyber (re-) insurance policy writing is NP-hard in IoT societies. In *Winter Simulation Conference (WSC)*. IEEE, 1–12.

[57] Ranjan Pal, Konstantinos Psounis, Jon Crowcroft, Frank Kelly, Pan Hui, Sasu Tarkoma, Abhishek Kumar, John Kelly, Aritra Chatterjee, Leana Golubchik, et al. 2020. When are cyber blackouts in modern service networks likely? A network oblivious theory on cyber (re) insurance feasibility. *ACM Trans. Manag. Inf. Syst.* 11, 2 (2020), 1–38.

[58] Ramamohan Paturi, Dau-Tsuong Lu, Joseph E. Ford, Sadik C. Esener, and Sing H. Lee. 1991. Parallel algorithms based on expander graphs for optical computing. *Appl. Optics* 30, 8 (1991), 917–927.

[59] Chen Peng, Maochao Xu, Shouhuai Xu, and Taizhong Hu. 2018. Modeling multivariate cybersecurity risks. *J. Appl. Statist.* 45, 15 (2018), 2718–2740.

[60] David M. Pooser, Mark J. Browne, and Oleksandra Arkhangelska. 2018. Growth in the perception of cyber risk: Evidence from US P&C insurers. *Geneva Papers Risk Insur.—Iss. Pract.* 43, 2 (2018), 208–223.

[61] Trivellore E. Raghunathan, Jerome P. Reiter, and Donald B. Rubin. 2003. Multiple imputation for statistical disclosure limitation. *J. Offic. Statist.* 19, 1 (2003), 1.

[62] Sasha Romanosky, Lillian Ablon, Andreas Kuehn, and Therese Jones. 2019. Content analysis of cyber insurance policies: How do carriers price cyber risk? *J. Cybersecur.* 5, 1 (2019), tyz002.

[63] Spyridon Samonas, Gurpreet Dhillon, and Ahlam Almusharraf. 2020. Stakeholder perceptions of information security policy: Analyzing personal constructs. *Int. J. Inf. Manag.* 50 (2020), 144–154.

[64] Kalpit Sharma and Arunabha Mukhopadhyay. 2021. Kernel naïve Bayes classifier-based cyber-risk assessment and mitigation framework for online gaming platforms. *J. Organiz. Comput. Electron. Commerce* 31, 4 (2021), 343–363.

[65] Kalpit Sharma and Arunabha Mukhopadhyay. 2022. Sarima-based cyber-risk assessment and mitigation model for a smart city's traffic management systems (SCRAM). *J. Organiz. Comput. Electron. Commerce* (2022), 1–20.

[66] Sachin Shetty, Michael McShane, Linfeng Zhang, Jay P. Kesan, Charles A. Kamhoua, Kevin Kwiat, and Laurent L. Njilla. 2018. Reducing informational disadvantages to improve cyber risk management. *Geneva Papers Risk Insur.—Iss. Pract.* 43, 2 (2018), 224–238.

[67] Leonie Tanczer, Ine Steenmans, Irina Brass, and M. M. Carr. 2018. Networked world: Risks and opportunities in the Internet of Things. (2018).

[68] Orcun Temizkan, Ram L. Kumar, Sungjune Park, and Chandrasekar Subramaniam. 2012. Patch release behaviors of software vendors in response to vulnerabilities: An empirical analysis. *J. Manag. Inf. Syst.* 28, 4 (2012), 305–338.

[69] T. H. Cormen, C. L. Leiserson, R. Rivest, and C. Stein. 2001. *An Introduction to Algorithms*. MIT Press.

[70] Manas Tripathi and Arunabha Mukhopadhyay. 2020. Financial loss due to a data privacy breach: An empirical analysis. *J. Organiz. Comput. Electron. Commerce* 30, 4 (2020), 381–400.

[71] Asaf Valadarsky, Gal Shahaf, Michael Dinitz, and Michael Schapira. 2016. Xpander: Towards optimal-performance datacenters. In *12th International on Conference on Emerging Networking EXperiments and Technologies*. 205–219.

[72] Ali Vedadi, Merrill Warkentin, and Alan Dennis. 2021. Herd behavior in information security decision-making. *Inf. Manag.* 58, 8 (2021), 103526.

[73] Spencer Wheatley, Annette Hofmann, and Didier Sornette. 2021. Addressing insurance of data breach cyber risks in the catastrophe framework. *Geneva Papers Risk Insur.—Iss. Pract.* 46, 1 (2021), 53–78.

[74] Spencer Wheatley, Thomas Maillart, and Didier Sornette. 2016. The extreme risk of personal data breaches and the erosion of privacy. *Eur. Phys. J. B* 89, 1 (2016), 1–12.

[75] Maochao Xu and Lei Hua. 2019. Cybersecurity insurance: Modeling and pricing. *North Amer. Actuar. J.* 23, 2 (2019), 220–249.

[76] Maochao Xu, Lei Hua, and Shouhuai Xu. 2017. A vine copula model for predicting the effectiveness of cyber defense early-warning. *Technometrics* 59, 4 (2017), 508–520.

[77] Maochao Xu, Kristin M. Schweitzer, Raymond M. Bateman, and Shouhuai Xu. 2018. Modeling and predicting cyber hacking breaches. *IEEE Trans. Inf. Forens. Secur.* 13, 11 (2018), 2856–2871.

[78] Zichao Yang and John C. S. Lui. 2014. Security adoption and influence of cyber-insurance markets in heterogeneous networks. *Perform. Eval.* 74 (2014), 1–17.

[79] David Zuckerman. 2019. Certifiably pseudorandom financial derivatives. *SIAM J. Comput.* 48, 6 (2019), 1711–1726.

Received 16 January 2022; revised 24 June 2022; accepted 10 October 2022