MIT
POLITICAL SCIENCE

Massachusetts Institute of Technology
Political Science Department

Resilient Mechanism Design
Foundations for Governance of Cyberspace:
Exploration in Theory, Strategy, and Policy

Silvio Micali, MIT
Nazli Choucri, MIT
Jing Chen, MIT
Cindy Williams, MIT

Electronic copy available at: http://ssrn.com/abstract=2317524

Explorations in Cyber International Relations

Massachusetts Institute of Technology   Harvard University

Resilient Mechanism Design
Foundations for Governance of Cyberspace
Exploration in Theory, Strategy, and Policy

Silvio Micali

CSAIL
MIT
silvio@csail.mit.edu

Nazli Choucri

Political Science
MIT
nchoucri@mit.edu

Jing Chen

Computer Science
Stony Brook University
jingchen@cs.stonybrook.edu

Cindy Williams

Center for International
Studies, MIT
cindywil@mit.edu

September 2013

Contents

1. Introduction

Three related trends in world politics – shifting in power relations, increased diversity of actors and entities, and the growing mobilization and politicization of global constituencies are contributing to a global "tussle" which threatens to erupt in a full-fledged international confrontation. Such contests may well reinforce the potentially powerful cleavages, such as those that became evident before, during, and after the World Conference on Information Technology, WCIT-2012. If present trends continue, it is unlikely that WCIT-2013 will reduce the cleavages and resolve the contentions.

Background

The purpose of the 2012 conference was to renegotiate the terms of the 1988 international treaty establishing worldwide order in the domain of information and communication broadly defined. The treaty, known as the International Telecommunication Regulations (ITR), was put in place prior to worldwide diffusion of the Internet and the attendant forging and expansion of cyberspace. In many ways, it reflected the "old" order. Today cyberspace has become an inherent feature of the world in which we live – for both industrial and developing societies – permeating the most remote regions of human habitation.

With the Internet at its core, the "new" order is shaped by the ever-expanding trends in cyber access, thus reinforcing the salience of cyberspace. The renegotiation of the 1988 treaty revolved, to a large extent, around the degree to which cyberspace management would be included under the suzerainty of the International Telecommunications Union (ITU).

The WCIT-2012 was hardly a success although it did lead to a formal document representing "consensus." Participating nations were generally divided into warring camps that did not engage in serious negotiations in any operational sense. When the conference came to an end, less than one-half of the UN's member states signed the resulting document; the United States, Canada, and key European countries walked away from it.

Recent advances in the game theoretic study of resilient mechanism design offer a new approach to the negotiation processes – in theory and in practice. These advances may have applicability to the design of a more effective negotiation process that could be used in framing, articulating, and coming to agreement on the next rounds of international accord in the cyber realm.

Purpose

This paper is about the foundational dynamics for a "new" governance order in the cyber domain, or for any other and contestation over the underlying design principles and processes for renegotiated arrangements, new treaties, or other forms of formal accord. Our intent here is not to resolve the contentions nor to signal a particular treaty-without-problems, rather it is to highlight critical features of resilient mechanism design that reflect with some degree of

accuracy the complexity of actors and interests, and the diversity of preferences and motivations shaping the debates.

One goal is to "set up" the mechanism design problem correctly to help frame a viable negotiation context. Resilience here is not about the nature of the outcome, but about the *process* of negotiation. In other words, the mechanism that will produce the desired outcome (treaty, accord, agreement, or other) must itself be resilient.

2. Contentions over Principles of Order

Transcending these issues are more profound undercurrents of international contestation that will invariably shape the parameters of the negotiations and most certainly the outcomes. Central among these is the tension between the *public authority* of the state system and the *private order* created by the increasing power and influence of commercial and other non-state actors. Coexisting with the nominally dominant mechanisms of public authority, the private order is becoming more and more powerful, and in some cases even eclipsing the authority of the state system.

As the WCIT-2012 demonstrated, the norms, principles, rules and regulations that currently operate in the management of cyberspace are already being contested, demonstrating considerable differences among states on the one hand and between many states and the private order that has constructed the present order and continues to dominate it.

Since its inception, the management of cyberspace rested largely with privately held authority. Seemingly incongruent with the diffusion of cyber access across territorial boundaries – transcending jurisdiction and bypassing conventional constraints of sovereignty – the cyber order worldwide continues to be managed largely by private authority. Therein lays a profound dilemma. Anchored in the concept of sovereignty, the state system dominates all formal arrangements in international relations buttressed by public international law that has evolved over the centuries. By contrast, the cyber system with its global scale and scope is constructed in the pragmatically evolving domain of private international law.

The state system has recognized this seeming anomaly and is making every effort to rectify and bring the private word under the jurisdiction of the public. But the members of the international system today – the states – are not all united in support of the reassertion of public order over all international commercial transactions, and even less with respect to the management of cyberspace. This friction adds to the cleavages already inherent in a world of shifting priorities and preferences.

All of this is directly relevant to future negotiation processes, most notably with cyberspace and its management, including future efforts to renegotiate the 1988 treaty. To simplify: on one hand are supporters of business-as-usual management by private authority; on the other are proponents of return-to- normalcy in international relations with the principles of sovereignty to shape and guide the management of cyberspace. The resulting "tussle" is created by this

cleavage, exacerbated by the shifting "sands" of structure and processes in contemporary international relations.

To complicate matters further, the cyber arena lacks the traditional distinction between the domain of authority for state and non-state actors and interests. For example, the major power in world politics, the United States, continues to support a privately managed cyber system (the corporate position shaped by competitive principles) rather than a publically governed cyber system (the sovereign position of public international law). By contrast, other states, notably China, prefer a state-driven order managed by an established international institution, namely the ITU.

It is in this context that the agenda of WCIT-12 was transformed into negotiating positions shaped by diverse actors driven by different attributes and capabilities, and motivated by a wide variety of preferences characterized more by diversity than commonality. And it goes without saying that negotiators seldom, if ever, disclose their preferences or provide full information on their motivations.

3. Mechanism Design for a Global Cyber Order

To begin with, we must distinguish between the institutional and diplomatic processes that are usually undertaken in any major international conference for reaching accord on new rules of interaction, on the one hand, and the theoretical, analytically rich approaches to decision and choice, on the other. We shall begin with the latter. In this context, game theory offers a well-developed study of strategic decision-making among rational actors, with its various forms and extensions that have been applied to strategic decision-making in various contexts or fields of inquiry.

For most games, the structure of the game is given and players "operate" within known frames of reference. But much of the challenge in the international situation addressed in this paper pertain as much to the struggle over the nature of the game itself (the structure and process) as it does over the potential payoffs to the players and their choices of moves and strategies. In this situation, "reverse game theory" – known as mechanism design – provides notable insights and guidance for framing and organizing the domain of contention in order to help articulate "solutions."

"Reverse Game Theory"

Critical features of mechanism design include the following: (a) the game structure is chosen or selected, not given, assumed, or inherited; (b) the game designer is interested in the outcome of the game, and may or may not be an objective party or an arbitrating entity; (c) the identity of the players is known, but their preferences are not; and (d) the solution lies in motivating the players to reveal their undisclosed motives and preferences, by making it in their interests to do so.

Framed thus, mechanism design places limited constraints on the players (attributes, diversity, or knowledge about the strategic game). The game designer is only a facilitator or arbitrator, and the players' hidden motives and undisclosed preferences are taken into account.

This "reverse game theory" maps onto an international reality where the agents are public entities (states) with privately-held motives, and the convening international organization is the International Telecommunications Union (ITU) whose voting members are states. As is common for institutions, the ITU itself has developed its own logic and rationale, motivated by expansion rather than constraint in scale or scope of authority.

Interestingly, from a theoretical perspective, but fundamentally critical from a pragmatic perspective, the agents (states as players and the ITU as the convening entity) are not autonomous decision makers. They are all interconnected in one way or another with the operators of the Internet (managed by private sector entities) and the broader cyber context as well as the legal principles upon which they operate internationally.

Power and Diversity

The operating entities hold enormous instrumental and commercial power and are present at the negotiating table. They themselves have a range of types of relationships and influence with the respective state-players but no voting power of their own. (Other non-voting entities are noted later on.) Moreover, in this example, the game designer – with specific interest in the game's outcome – becomes the overarching agent, choosing the nature of the game structure and process.

At the onset, we referred to a "tussle" over the redesign of a major global accord, the international telecommunications treaty of 1988. Almost everyone will agree that the international political and strategic ecosystem of 2012 was far different from that of 1988, when the Internet was only beginning to come into widespread, global use. This sea change in the ecosystem was a fundamental motivating factor for the 2012 renegotiation of the ITR, and will continue to drive future efforts to come to terms on international agreements in the telecommunications and cyber arenas.

With a new international treaty at stake, actors in the game must see the potential product of negotiations as "better" than status quo *arrangements and more relevant to current conditions.* From a mechanism design perspective, the *procedure* used for constructing a new treaty must be resilient. In the "reversed" game theory language resilience means that the new mechanism must do at least most of what it is supposed to do – while accounting for other forces that are usually not modeled in traditional game theory. Resilience here is not about the nature of the outcome but about the *process* itself. It is the mechanism itself that must be resilient.

4. Innovations in Resilient Mechanism Design

Traditionally, the international system is viewed as anarchical, with no overarching authority enfranchised to maintain order over the entire ecosystem of sovereign entities and the non-state

actors. Differences of views prevail regarding the degree of anarchy or the extent of disorder. Everyone, however, agrees that the international system is highly complex – by any definition of complexity. Given this reality, an important theoretical move is to extend the descriptive features of traditional mechanism design noted above and provide some correctives to reduce basic fragility and buttress potentials for resilience.

In conventional game theory, the interactions among the parties, the negotiations and renegotiations, are all done after the basic game is set. The context is in place, the parties operate within the rules of the known pre-set context. The framework we shall explore, resilient mechanism design, is based on the reverse logic. The parties to the negotiation must first agree on the context itself, the nature of the negotiation arena, before they engage in moves and countermoves.

Design Features

In its skeletal form, the resilient design mechanism is anchored in five features: (1) the *players,* the actors or the entities; (2) the *preferences* of the players; (3) the outcomes, the *results* of the process; (4) what might be called the *desiderata*, a description of the kind of outcome desired; and (5) the *solution* concept. The solution concept is one place where the assumptions must converge with how the players behave.

Contentions over any new or revised treaty will reflect a world of complex, interdependent parts – the sovereign states with voting rights and the non-state actors with potential power and influence – are all embedded in highly intricate networks and linkages. These are features of the contemporary international ecosystem, but they will also remain overarching parameters irrespective of the final outcome of any new treaty. These are all contextual system-wide features within which to make informed theoretical choices to enhance the resilience of mechanism design.

Strategic Interactions

The recent theoretical developments, addressed here, focus on three features of strategic interaction in traditional game theoretical context that are particularly problematic in analytical as well as pragmatic terms. These are the actors' protection and control of information about preferred own moves and motivations (*privacy*), the presumed distorting effects of undisclosed or hidden coordination among actors (c*ollusion*), and impediments ineffective computation of strategy for realizing preferred outcomes (*computation*).

By definition, state security requires safeguarding of significant information bearing on national security, and by extension, protecting against any contingency that could compromise the negotiator's posture and position in international negotiation. Observational negotiation, coined here, refers to moves targeted to elicit information rather than to advance negotiation. This is a general approach for enhancing the strategically significant information that is elevated to "knowledge" and can strengthen the negotiating position.

In international negotiations, collusion is generally a rule rather than an exception. Formal collusion (notably due to treaty commitments), or informal collusion (in response to incentives or promises), or hidden collusion (undisclosed, privately held alignments of position) are all variations of the well-known complexities of negotiation. Collusion of any sort is a "normal" form of international interaction. Controlling collusion or eliminating it entirely would be impossible in a world-wide negotiation of the sort considered here.

By definition, the construction of the Internet and the character of cyberspace altered the traditional features of communication and information systems, contexts and practices in major ways. In addition, the growing practice of making decisions online contributed – along with other developments – to an international move to make decisions about the nature and management of the online world. The three features addressed here – privacy, collusion, and computation – while individually and collectively daunting, are also important obstacles in mechanism design. Resilience in mechanism design means finding ways of transforming these situational liabilities into intuitions with potentially significant contributions for theory development.

Innovations and Implications

The first notable innovation in theory development is *to reduce impediments to privacy* significantly by framing the context for the strategic arena of the players (actors, entities, states) in an interactive context, an "extensive-game," where only incremental and elementary moves are made and only the most essential information about preferences or motivations is revealed. From a negotiation perspective, this strategic context does not depart too dramatically from known practices. Seldom do negotiators proceed in other than incremental ways, and even more rarely do they reveal zero information in the process or share more than needed.

In the treaty negotiations considered in this paper, national government participants are likely to care deeply about the privacy of some information related to their governments' positions. No negotiator will want to tip his government's hand too early by divulging more information than is absolutely necessary to meet his own negotiating aims. Negotiators may even be motivated to dissemble, rather than reveal too much about the operations, vulnerabilities, plans, or costs of the states they represent or the private-sector firms that those states depend on. The reality of privacy concerns and the potential for false information must be assumed in any treaty negotiation that will ultimately shuffle the deck in the allocation of risk and responsibility in the cyber arena.

Using emerging models of resilient mechanisms, for example, the ITU, as mechanism designer, could ultimately devise negotiating frameworks in which each state could keep most of its information to itself—including information about its goals for the negotiation and risks about which it is most deeply concerned. Instead of releasing such private information, each negotiator can act on information already acquired about the other states' concerns and intentions, and the ITU can make use of public and ITU-held information about individual states' goals and risks. Negotiators would not be required to divulge private information. Instead, they would be rewarded for negotiating in the collective interest. They would be sanctioned within the context of the negotiation for acquiescing to a stipulation, but later renouncing it. The negotiation, which would occur in steps, would allow the negotiators' privately held information to inform

outcomes gradually, without having to be aired publicly among the negotiators or between the negotiators and the ITU. Even without the airing of private information on risks and goals, the negotiation would ultimately settle on outcomes that lowered the collective risk and improved collective progress toward goals.

The second innovation in mechanism design consists of a *counter-intuitive approach to the problem of collusion.* While common sense suggests that everything should be done to prevent collusion, an alternative logic argues instead for finding ways to protect the mechanism design from the damages of collusive behavior. In other words, rather than focus on altering seemingly generic and intractable forms of behaviors, as noted above, the alternative is to insulate the mechanism designer and the design from the corrosive properties of collusion. For the tasks at hand, the normative feature of collusion is not at issue; what is important is that the mechanism design not be contaminated by such high-probability behavior. Recently labeled as "leveraging collusion," this intuition leads to a fundamentally new logic of interaction. With this logic, collusion is not, by definition, a disruptive force; rather it is an exploitable asset as much as the players' preferences. Better still, the mechanism can be helped rather than injured by knowledge of collusive players, not just the independent players. Traditionally, game theory seeks to insulate the mechanism from collusion; in the "reverse form" the aim is to ensure that the path to the desired outcome is *resilient* against collusion. Because of the way in which the mechanism works, this aim is achieved as long as collusion does not materially affect the outcome

This "reverse" form relaxes somewhat the traditional assumption that states will not collude amongst themselves. In fact, the sorts of frameworks emerging from recent research assume outright that some states will ally to achieve common aims and lower shared risks, potentially at the expense of states outside their coalitions. Experiments with such frameworks in a theoretical context are laying important groundwork that can ultimately help the ITU establish the frameworks for step-by-step, round-by-round negotiations that can lead to treaty outcomes that will improve collective cyber security, by aggregating information in the face of collaboration by multiple subgroups in the negotiations.

The third innovative feature has special resonance with prevailing understandings of international relations, namely the importance of *internal and external information*. In game theoretical parlance, the challenge is to find new ways to align the states' choice of strategies to the strategies that they "should" use, taking into account both the usual *internal* information of every player about himself/herself as well as each player's *external* information about the preferences of other players. Thus, the aim is to design a mechanism that leads actors (players) to choose a set of strategies that leads resiliently to the desired outcome, *given* their existing preferences. To simplify, again, while game theory and its reverse form recognize and focus on the player's internal information notably related to motivations, the intuition here is to consider the player's external information as well. While this might be an innovation in the game theoretical world, in the "real" international context, all states (and other entities) are highly sensitive to the preferences and position of others, especially potentially close competitors. Information and preferences pertaining to the international ecosystem described briefly above is of value to players (and to the mechanism designer), to be sure, but over and above the "normal" situation is the value of information about player preferences regarding the nature of the emergent cyber ecosystem under consideration.

This type of alignment or mapping, known as "solution strategy" by game theorists, is yet to be developed. The intuition may be robust, but proof remains daunting. Parenthetically, none of this assumes, in theory or in practice, that external information is accurately assessed; only that it represents a player's (state) own view of the preferences of other player's states – thus avoiding any dilemmas associated with misperceptions. While it is often common practice to model the actors' external information under Bayesian assumptions, an added challenge related to this theoretical intuition is to minimize the constraining assumptions.

Bounded Rationality

Finally, there is an added consideration, not an innovation *per se,* but one that strongly supports the foregoing, is the implicit assumption of *bounded rationality*, rather than optimizing rationality. The aim of resilient mechanism design is to ensure that the path to the desired outcome is resilient against common failures of rationality. The three characteristic features of bounded rationality in its original formulation – search for alternatives, satisficing, and adaptation – are complementary to the above logic, by introducing "realistic" elements in the overall logic. If the mechanism design is successful, the desired outcome is reached regardless of whether actors are perfectly rational, fully informed and so forth.

5. The Influence of Context

Early preparations and negotiations related to WCIT-12 were deeply embedded in the politics of the international system and influenced by the growing role of cyber systems. Scholars and policy makers alike recognize that the construction of cyberspace has greatly increased the complexities of the traditional international decision-making context for telecommunications.

Complexity

Neither traditional game theory nor its "reverse" form take into account the environment or context outside the parameters of the game. The broader context of the cyber system – with its dense, user-driven, bottom-up, diffuse, and distributed features – consists of myriad individual networks that are interconnected, coordinated and routed via exchange points to enable the flow of information. It is supported by private companies (agents) that make money by providing access to their networks (in the US), by public companies supported by the state (in other parts of the world), and by not-for-profit organizations and entities that design and operate the broad rules of interaction (the driving licenses, so to speak, and the operational rules of the road). The coordination is provided by a set of pragmatic standards and procedures enabled and supported by the users and their agents. All of this constitutes the context or environment within which individual entities interact and negotiate.

The context itself carries powerful influences. An important design issue for the next round of negotiations may well be to determine whether any of the cyber features noted above represent "random" elements that provide little systematic bias, or alternatively, if their salience requires

representation by some parameter, at least in theory. To ignore these features entirely may leave external information devoid of contextual foundations.

If the international system is organized around the principle of *state sovereignty* with no central authority, then the cyber ecosystem is organized around *networks of communication*, with no overarching principle other than facilitating the transmission of information. A long record of international law provides the general principles of interaction among sovereign states, and it is the sovereign states that are now the lead players in the design of mechanisms for management of the cyber ecosystem. It has to contend with non-state players that are fundamental to the operations of the cyber sphere and are also powerful in their own right.

Characteristics

While there is, as yet, no clear understanding of the independent effects of the external context, it is naïve to ignore the environment of the players, the pressures and the possibilities. At least three significant aspects of the political and strategic context of future telecommunications treaties or modes of cyber governance can be incorporated into the framing of the resilient mechanism design logic for the next round of negotiations.

The first significant aspect involves *player information.* The players' (or negotiators') external information invariably relates to position-on-these-cleavages; and the players' internal information inevitably reflects the influence of dominant constituencies. Unlike a two-level game, where the overall structure is sequential and segmented, the players with information and preferences and the game designer are all operating in creating a solution strategy for the overall design mechanism.

The second significant feature is *player preference*. Player preference is connected to undercurrents of international contestation raised earlier in this paper. Of the many contestations that characterize both the real and the cyber contexts and their intersection, much of the variation in the preferences of the players is captured by two powerful cleavages. One is over the *degree of order* to be sought. Some players prefer the hierarchical model of order for the cyber domain, and others are pushing for a more distributed system. The other cleavage is over the driving *decision principles.* Some players support state-based *security-seeking* imperatives; others prefer the profit or opportunity-seeking imperatives.

The third feature of the global negotiation context that can be incorporated into the logic of the resilient mechanism design relates to the function of the *mechanism designer* and its role as an "interested" player. Early in the preparatory phase for WCIT-2012, it became clear that the mechanism designer – a function we attribute to the ITU in this paper – held distinctive preferences that differed from (a) the actual operators of the cyber domain, and (b) the established and dominant power in the international system, namely the United States.

Co-Evolution

In many ways the influence of context mirrors the co-evolution of cyberspace and international relations since the construction of the Internet. In the early years it was customary to consider the

Internet – and the development of the cyber arena of interaction – without any notable reference to international relations or to the quest for order in cyber-based interactions. This is no longer the case. Governance issues are becoming increasingly salient, and we can expect more rather than less attention to the conflicts and contentions that surround all matters of governance.

6. End Note

As a product of the joint MIT-Harvard Project on *Explorations in Cyber International Relations (ECIR),* this paper reflects the interdisciplinary orientation of ECIR research in its diverse investigations of interconnections of cyberspace and international relations. While we have focused on the broad issue of cyber governance, it is not difficult to appreciate some of the operational features thereof.

Consider, for example, applications of resilient mechanism design to negotiations for a treaty as specific as that of managing identify and attribution on the Internet. All of the issues addressed in this paper come sharply into focus when we consider the key players (states and non-state actors) – with diverse interests and preferences – as well as the game designer (with the ITU as a reasonable example) with its own interests and influence, all embedded in a dynamic context shaped in large part by technological, political and strategic factors. It is not difficult to imagine how the United States and China, for instance, differ on the relevance of privacy to any shared *desiderata.*

References

1. Azar, Pablo, Jing Chen, and Silvio Micali. 2012. "Crowdsourced Bayesian Auctions." *Proceedings of the 3ʳᵈ Innovations in Theoretical Computer Science (ITCS 2012)*, Cambridge, MA: pp. 236-248.

2. Chen, Jing and Silvio Micali. 2013. "The Order Independence of Iterated Dominance in Extensive Games." *Theoretical Economics (TE)*, 8 (1): pp. 125-64.

3. Choucri, Nazli. 2012. *Cyberpolitics in International Relations*, Cambridge, MA: MIT Press.

4. Choucri, Nazli. 2012. "Cyberpolitics in International Relations." in Joel Krieger (ed.), *The Oxford Companion to Comparative Politics* (267-271), New York: Oxford University Press.

5. Choucri, Nazli. 2012. "The Convergence of Cyberspace and Sustainability." *e-International Relations*. http://www.e-ir.info/2012/04/20/the-convergence-of-cyberspace-and-sustainability/.

6. Choucri, Nazli and Daniel Goldsmith. 2012. "Lost in cyberspace: Harnessing the Internet, international relations, and global security." *Bulletin of Atomic Scientists*, 68 (2): pp. 70-77.

7. Madnick, Stuart, Nazli Choucri, Steven Camiña and Wei Lee Woon. 2012. "Towards better understanding Cybersecurity: or are 'Cyberspace' and 'Cyber Space' the same?" *pre-ICIS workshop on Information Security and Privacy (SIGSEC),* Paper 27.

8. Madnick, Stuart, Nazli Choucri, Xitong Li and Jeremy Ferwerda. 2011. "Comparative Analysis of Cybersecurity Metrics to Develop New Hypotheses." *Proceedings of the Workshop on Information Security & Privacy (WISP)* (Jointly hosted by AIS SIGSEC and IFIP TC11.1), Shanghai, China.

9. Madnick, Stuart, Xitong Li, and Nazli Choucri. 2009. "Experiences and Challenges with Using Cert Data to Analyze International Cyber Security." *Proceedings of the AIS SIGSEC Workshop on Information Security & Privacy (WISP 2009)*, Phoenix, AZ: pp. 6-16.

10. Micali, Silvio, Chris Piekert, Madhu Sudan, and David A. Wilson. 2010. "Optimal Error Correction for Computationally Bounded Noise." *IEEE Transactions on Information Theory*, 56 (11).

11. Micali, Silvio and Jing Chen. 2011. "Mechanism Design with Set-Theoretic Beliefs." *Proceedings of the IEEE 52ⁿᵈ Annual Symposium on Foundations of Computer Science 2011 (FOCS),* pp. 87-96.

12. Sechrist, Michael, Chintan Vaishnav, Daniel Goldsmith and Nazli Choucri. 2012. "The Dynamics of Undersea Cables: Emerging Opportunities and Pitfalls." *Proceedings of the 30th International Conference of the System Dynamics Society,* Elke Husemann and David Lane (eds.), St. Gallen, Switzerland.