The Engineering of a Scalable Multi-Site Communications System Utilizing QKD

Piotr K. Tysowski, Xinhua Ling, Norbert Lütkenhaus and Michele Mosca

***Abstract*—Quantum key distribution (QKD) is a means of generating keys between a pair of computing hosts that is theoretically secure against cryptanalysis, even by a quantum computer. Although there is much active research into improving the QKD technology itself, there is still work to be done to apply engineering methodology and determine how it can be practically built to scale within an enterprise environment. Significant challenges exist in building a practical key management service for use in a metropolitan network. QKD is generally a point-to-point technique that is subject to steep constraints related to the key generation rate and distance between hosts. A multi-disciplinary team has researched the integration of QKD into enterprise-level computing and to enable quantum-safe communication. A novel method for constructing a key management service is presented that allows arbitrary computing hosts on one site to establish multiple secure communication sessions with the hosts of another site. Various key exchange protocols are proposed where symmetric private keys are granted to hosts while satisfying the scalability needs of an enterprise population of users. The key management service operates within a layered architectural style that is able to interoperate with various underlying QKD implementations. Variable levels of security for the host population are enforced through a policy engine. A network layer provides key generation across a network of nodes connected by quantum links. Optimizations are performed to match the real-time host demand for key material with the capacity afforded by the infrastructure, resulting in a flexible and scalable architecture.**

I. Introduction

SECURE NETWORK COMMUNICATION is a principal function of IT infrastructure. In particular, secure inter-domain communication is important to larger organizations that are distributed across multiple geographical sites. The key management service of an IT security system will typically provide a data encryption service on top of a standard network protocol. Although there are many commercially available cryptosystems in use today, the majority of them rely upon key exchange under public-key cryptography where the computational problem is infeasible to break using today's computing technology. However, rapid advances are occurring in the field of quantum computing. Once a quantum computer is built to solve problems of a practical scale, conventional public-key cryptography will become completely vulnerable to attack, and a new way of protecting transactions over a network must be proposed. To mitigate this risk, Quantum Key Distribution (QKD) has been devised, which is also based on the laws of quantum physics and is a theoretically secure form of generating keys that is resistant to attack by both conventional and quantum computers.

The authors are with the Institute for Quantum Computing (IQC) at the University of Waterloo in Canada. This work was supported by Quantum Canada and was performed in collaboration with the National Research Council of Canada.

There has been significant research activity over the years relating to the theory and implementation of QKD techniques. QKD is already in operation today in large-scale experiments using commercially available equipment, and promising research has been presented on how trusted nodes can be utilized to build a larger system. However, there is still work to be done on how QKD key management can be practically utilized within an enterprise environment to flexibly serve a large host population, from a practical engineering perspective.

II. Our Contributions and Impact

The overall goal of our work has been to show the feasibility of integrating QKD technology with classical communication networks, and to provide various options for migration. We viewed the problem primarily from an engineering lens. Our aim has been to apply best practices in contemporary security, network, and software engineering in the application of QKD to practical real-world systems. We engaged a multidisciplinary team of researchers, that included mathematicians, physicists, architects, and engineers. Our engineering staff not only had pursued doctoral research in security and networks, but also had decades of industry experience as technical leads and consultants at world-renowned firms; we were thus able to bridge research with leading industry practices based on first-hand experience to create a concrete design.

Our main contribution is the design of a scalable QKD-based system that enables secure multi-site communication, and is compatible with various QKD technologies. We provide a scalable service to support enterprise-level secure traffic between hosts in a large metropolitan network comprising sites that are indirectly connected. The design consists of a full protocol stack including an enterprise-level key management layer and a quantum network layer that performs quantum key generation. To maximize efficiency, the key generation system dynamically adapts to changes in demand and network infrastructure. Hosts are issued session keys to securely communicate over a conventional network, while quantum key generation occurs over quantum links. We further investigated integration with standards such as TLS, IPsec, and Kerberos. Our research results included design artifacts to lay the groundwork for the implementation of a research test bed.

Our work has numerous impacts; it informs QKD practitioners and equipment designers of the operating requirements of an enterprise system; it informs architects of key design choices and trade-offs to make when incorporating QKD technology into a communications system; and, it demonstrates to

software engineers how client applications can be built to make use of secure QKD key material in transparent fashion. These insights aim to tackle what sometimes appear to be insurmountable obstacles in standards acceptance and widespread use of QKD, despite its compelling intrinsic security benefits.

III. The Network Model

The network model, as shown in Fig. 1, comprises multiple sites that are geographically separated, such as buildings in a metropolitan area. Each site contains a Local Area Network (LAN) consisting of potentially thousands of heterogeneous hosts, including desktop and mobile device users. Sites are connected by fibre channels with multiplexed user data, which are considered conventional traffic channels; the fibre is typically shared as opposed to being dark. LANs are connected to switches, and in turn, to Wavelength Division Multiplexers (WDMs). The fibre channels can be exploited to carry both secure user traffic as well as to perform QKD.

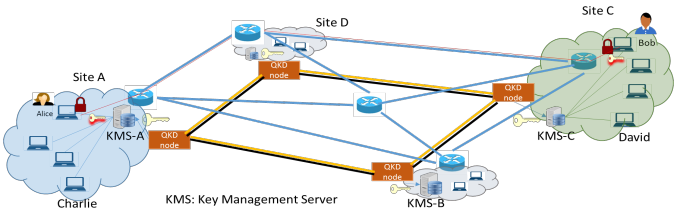


Fig. 1. The network model consisting of multiple communicating sites.

A fundamental difficulty is that the QKD protocol is designed to work for two parties only. However, in a metropolitan network, quantum-safe communication must be permitted between any arbitrary hosts on any sites; flexible addressing and scalability is therefore required. Additional challenges are the distances involved; sites may be separated by tens of kilometres, limiting the key generation rate. A metropolitan network may consist of dozens of sites, and each site may contain thousands of hosts, so that there is great contention for quantum key material. Additionally, sites may not be fully mesh-connected by quantum links, so relaying of quantum key material to remote sites is required. And finally, dedicated quantum links for QKD may not be available, so that existing fibre-optic lines for client traffic may need to be utilized for QKD, and each node in the network may have dual roles; it may be an end-point for a connection or acting as a relay.

It is important to note that exclusive reliance on QKD may not be justifiable. In fact, QKD augments well the conventional quantum-safe (or, interchangeably, post-quantum) ecosystem. Since QKD requires special hardware and has limited key rate over long distance, a possible key construction strategy is to combine the two techniques. First, quantum-safe algorithms can authenticate the QKD channel via PKI. Second, quantum-safe algorithms and QKD are used independently to create two session keys; the two keys are combined (for example, through an exclusive-or) so that an attacker must break both. In this case, if either the QKD or the quantum-safe key generation become vulnerable, the overall security is not compromised.

IV. The Design of a Scalable and Secure Enterprise Communications System

We have designed a system for enterprise-level sites to securely communicate in a large metropolitan network. Key generation occurs using QKD technology over quantum channels connecting pairs of sites. A scalable service issues session keys from a quantum key pool, containing generated quantum key material, to local hosts on each site. Hosts can then use the session keys to securely communicate over a conventional TCP/IP channel. The key generation and distribution mechanism is designed to scale to thousands of local hosts.

The system is composed in a layered architecture style, as shown in Fig. 2, which can accommodate any QKD technology with minimal changes. This software engineering model ensures that the major functions are grouped separately with well-defined interfaces across the layers. It results in a technology-independent design, as replacing fibre-link QKD with free-space QKD can occur by replacing the link layer.

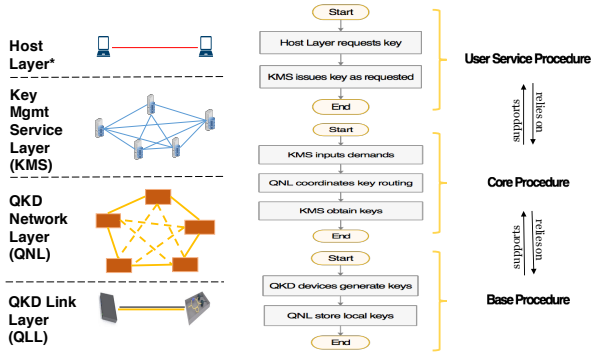
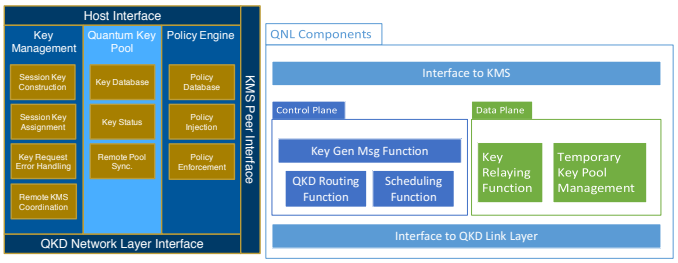


Fig. 2. The layers and procedures of the QKD communications system.

The Service Layer contains a KMS (Key Management Service) that issues quantum-generated keys to hosts to secure new communication sessions with hosts on other sites. The main functions are shown in Fig. 3. The KMS reads quantum-generated key material from the network layer below, and manages the key material in the KMS's quantum key pool. The KMS maintains key pool synchronization with other sites. It constructs session keys for hosts using an appropriate key construction strategy that is consistent with the security policy contained in its policy engine; for instance, the policy may dictate different key lengths and lifetimes. The KMS will issue session keys to local hosts upon request while the quantum key pool contains sufficient key material. As dictated by the policy, the KMS will make an appropriate response when the key pool is nearly exhausted; for example, it may wait for additional quantum key material or re-use existing material using a key expansion technique. The KMS issues keys to hosts using a generic protocol, as shown in Fig. 4, or one that is fully integrated with a standard. To ensure scalability, minimal state is maintained by the KMS. The host is responsible for negotiating a secure session after retrieval of the session key. The crucial characteristic of the protocols is that for hosts Alice and Bob communicating from separate sites, Alice's KMS issues a session key to her; Alice then transmits key

selection information to Bob so that he may retrieve the same session key from his KMS. Because the quantum key pools are always in sync during the QKD process, and the selection information is an index into the pool, the session key itself is never actually transmitted and thus cannot be eavesdropped. We have also integrated this process using the pre-shared key cipher-suites of TLS, IPsec, and Kerberos; the integration requires minimal changes to the client implementation.



(a) Key Management Service (b) Quantum Network Layer

Fig. 3. A functional view of the Service and QKD Network Layers.

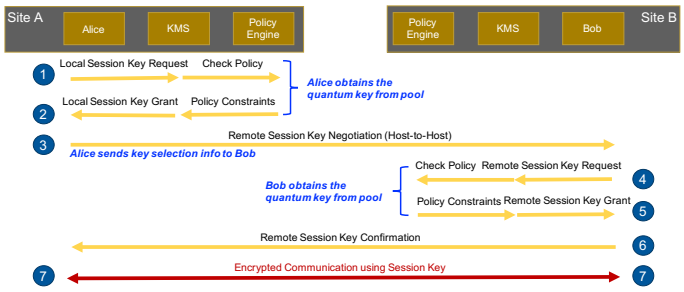


Fig. 4. The generic protocol for key negotiation.

The QKD Network Layer (QNL) provides quantum key material to the KMS. Its functions are shown in Fig. 3. It negotiates key generation through QKD between communicating sites. In case there is no direct quantum link between a pair of sites, such that they are not neighbours, then it relays key material via trusted nodes. In this way, the QNL extends QKD from point-to-point links to form a quantum network so that it can generate quantum key material for every pair of nodes. The control plane within the QNL establishes a path for the relay through key generation requests in tandem with a routing function. The routing takes into consideration the link capacities of the quantum network as well as the demand for key generation as determined by the KMS. The QNL establishes one or even multiple paths for key generation between any arbitrary nodes and it responds to demand dynamics with real-time scheduling. The data plane within the QNL temporarily stores the key material that is relayed between nodes. Robustness is built into the fabric; if the QNL sees any service interruption, it notifies the KMS.

The KMS is constantly monitoring local demand for key material through various heuristics. The quantum key pool that it maintains can serve as a cache for a surge in traffic. Once the pool is depleted, the KMS can request on-demand key generation by the QNL. If there is a constant level of demand, the KMS can request continuous key generation. The QNL will ultimately optimize the key generation across the entire network so that throughput is maximized. For users that are of higher priority, key generation demands may be weighed more heavily in the optimization calculation.

Immediately below the QNL, the QKD Link Layer (QLL) produces raw quantum key bits over each link by executing a QKD protocol. It establishes quantum key material between connected node pairs and provides it to the QNL. It can expose switching and addressing functionality to the QNL, and use existing infrastructure and shared resources. There is a plethora of QKD protocols and link technologies to choose from, with varying implementation complexities, key rates, and robustness. Metropolitan-scale distances are achievable through a combination of fibre and limited free-space links. Longer distances could be supported by quantum repeaters and satellites, and physical routing accomplished via optical switching, whether active or passive. Resource sharing is possible with the multiplexing of QKD and classical signals.

V. Lessons Learned and Recommendations

We propose that our four-layered architecture is a valuable framework for the QKD community to facilitate the advancement and adoption of QKD in practical network systems. There is a myriad of QKD technologies that are still evolving. A generic framework can accommodate any QKD technology without rework, and disruption to the entire tool chain is avoided. End-users may design their applications to obtain QKD-generated keys from a KMS without adapting to changes in the inner workings of the KMS, or the QNL algorithms and topology, or the underlying QLL technologies.

Our design is not generic in nature; it is highly customized for the challenges of QKD. It maximizes the security benefits of QKD while mitigating its limitations. Because the quantum key generation rate is still relatively lower, there needs to be flexibility in how quantum key material is consumed. The policy engine can dictate various key consumption strategies. Hosts may wish to communicate across the entire network, so it is useful to have intermediate nodes assist with key generation. Because the demand for keys may continually change, the system must adapt by deciding when to run QKD, using what paths in the network, and for how long.

The size of investment in QKD technology and infrastructure is an important consideration. To minimize the switching cost and speed up adoption, it may be best to integrate with existing security and authentication protocols and standards. The use of shared infrastructure, such as multiplexing the QKD protocol with conventional user traffic on a shared fibre-optic line, is another tenet of rationalization and fast ramp-up. Running the system as a set of micro-services will provide plug-and-play modularity within our layered architecture.

Finally, it is important to create a system that can effectively scale and support popular use cases. Optimizing the key generation activity based on real-time demand monitoring ensures maximum utilization of the security infrastructure. Defining flexible security levels and associated key issuance strategies will result in more efficient consumption of keys.